# DARWINISM, MEMES, AND CREATIVITY

## A Critique of Darwinian Analogical Reasoning From Nature To Culture

Inaugural-Dissertation zur
Erlangung der Doktorwürde
im Fach Philosophie
der Philosophischen Fakultät I
(Philosophie und Kunstwissenschaften)
der Universität Regensburg
vorgelegt von

**Maria E. Kronfeldner**

Regensburg
Dezember 2005

**Gutachter**

Erstgutachter: Prof. Dr. Hans Rott (Universität Regensburg)
Zweitgutachter: Prof. Dr. Holmer Steinfath (Universität Regensburg)
Drittgutachter: Prof. Dr. Peter McLaughlin (Universität Heidelberg)

für meine Mutter, in ewiger Liebe,
und für alle, die nicht einmal wissen, dass sie ein Recht auf ihre Träume haben

# CONTENTS IN BRIEF

# CONTENTS

# PREFACE

According to Darwinism, the biosphere constantly changes. Culture changes as well. In the biosphere as well as in the cultural realm, new characteristics arise over and over again and some novelties persist and lead to lasting changes. The types of entities that are involved in these changes are genes and organisms in the case of biological evolution, and cultural units such as ideas, values, beliefs, patterns of behavior, and artifacts in the case of culture. Today, biological evolution is believed to be explainable by Darwinian evolutionary theory.

Cultural change, however, is thought to arise through creative acts and selective choices of individuals, leading to the diffusion of novelties. *Culture* is usually defined as consisting of those characteristics of individuals that are not innate but created or learned by individuals during their life. *Creativity* in its basic sense is the human capacity to create new and valuable responses to challenges to which humans are exposed to, or to which they expose themselves. Those responses that are overtly delivered and are adopted by others become part of a certain culture. They spread. This is cultural *diffusion*. Since diffusion is a change in the frequency of certain cultural items, a culture as a whole changes as a consequence of the dual process of creativity and diffusion. If a creative act builds on past innovations, creativity accumulates through the iteration of this dual process and leads to history.

But how can we explain creativity and diffusion, the two parts that make up cultural change? How can we explain that human beings produce new answers to new challenges, and how can we explain why certain ideas spread and others do not?

Darwinian approaches to cultural change state that cultural change can be explained as an evolutionary process in the Darwinian sense. Such approaches are the subject of this investigation. They do not reduce culture to genes or other biological processes. They draw an analogy between change in culture and change in nature – an analogy between the processes of organic evolution, as explained by Darwinian theory, and the processes of cultural change. There are different versions of such Darwinian analogical reasoning from nature to culture. I will concentrate on two theories: The *Darwinian approach to creativity* and the theory of memes, usually called *memetics*. The Darwinian approach to creativity states that novelty in culture is created by a Darwinian process of blind variation and selection. In being creative, humans are – like nature – 'blind watchmakers.' Memetics goes even further. Memeticists claim that we can eliminate the human mind as the main causal force in our explanation of creativity and culture. Memes and not minds are the main causal force in cultural change, as genes and not individual organisms are regarded by some evolutionists as the main causal force in evolution. The claim implied in drawing the analogy between nature and culture shifts from the 'blind watchmaker' to 'no one watching.' Although there are many other analogical applications of Darwinian thinking, I will only consider these two, since they directly attack our traditional view of creativity and culture, namely that humans are the creators of culture.

I will differentiate between *three basic analogies* contained in these two Darwinian approaches to culture: (1) the *ontological analogy* – an analogy with respect to the entities involved in biological evolution and cultural change; (2) the *origination analogy* – an analogy with respect to the origination of novelty; (3) the *explanatory units of selection analogy* – an analogy with respect to the causal role certain entities play. I will also split my critique of these basic analogies into three questions: First, are there sufficient similarities between culture and the biosphere to justify the analogies as descriptively adequate? Second, if so, do the analogies help explain cultural change? Third, if the analogies are descriptively adequate and help explain cultural change, do they provide new descriptions and explanations of cultural change? I will show that the analogies either give wrong or unjustified descriptions and explanations, or they give trivial restatements of what we know already.

Creativity and culture are topics that are addressed by many disciplines and theories. It would be impossible to write a general theory of creativity and culture in one single book. My main goal will be a philosophical analysis of the Darwinian approach to creativity and of memetics from *within* these evolutionary theories. For this, I will introduce and combine debates from various disciplines – with all the risks involved in doing so. These disciplines include genetics, evolutionary biology, psychology, anthropology, and philosophy – to name the most important ones. I certainly will not answer all the questions that could arise with respect to what these disciplines have to say on creativity and culture, but I will answer the question I have set for this study: Whether the three basic analogies between biology and culture make sense or not.

I will introduce the whole issue in more detail in chapter 1. In chapter 2, I will explain what a Darwinian explanation of change is. The ensuing chapters 3 – 5 consider each of the three basic analogies separately. I will summarize my findings in a short epilogue.

# 1 FROM THE DARWIN INDUSTRY TO THE DARWINIAN ANALOGIES

## 1.1 THE DARWIN INDUSTRY

### Folk-Darwinism

At the end of his *On the Origin of Species*, Darwin writes: „Light will be thrown on the origin of man and his history“ (Darwin 1859: 487). Indeed, the ‘light’ of evolutionary theory shines since then. Today, roughly 150 years after the first publication of the *Origin* in 1859, evolution is almost everywhere. It would be no surprise, if Theodosius Dobzhansky’s famous statement – “Nothing in biology makes sense except in the light of evolution” (Dobzhansky 1973) – ‘mutates’ to ‘*Nothing at all* makes sense except in the light of evolution’. Almost everything in the social and cultural sphere that is able to change and does not change in a sudden and abrupt way is said to *evolve*. Political agendas, partnerships, economies, firms, or acts of any kind – they evolve. Stars, galaxies and the universe are also said to evolve. Artists and programmers refer to ‘mutants’ of artifacts or programs. Markets are said to be dominated by the ‘survival of the fittest’. Everyone has to adapt to this or that.

As Chris Buskes (1998: 1) emphasizes, this “folk-Darwinism” is usually “crude and superficial.” It has nothing to do with Darwinism. It is a mere *facon de parler,* where every kind of gradual change is referred to as evolution. The term ‘change’ is merely replaced by the term ‘evolution.’

### Evolution in philosophy, science, and politics

But the term ‘evolution’ did not only enter our language as a vague idea for any kind of change. Evolutionary thinking entered philosophy and scientific thinking in diverse and elaborated ways, either as a way of describing and explaining our innate human nature, or as a way to export the Darwinian paradigm to other domains of research. Social Darwinism, eugenics and racism were hotly debated evolutionary endeavors of the late 19th and early 20th century. Social Darwinism supported the restriction of policy programs: Policy should not help the physically, mentally, or economically weak. It is best to let the law of nature rule, the principle of the survival of the fittest. Eugenics converted this negative laissez-faire program to an active policy program, in order to intensify nature's principle of the survival of the fittest. Racism drew sharp essentialist boundaries between races and interpreted some races as evolutionarily 'superior'. In part, this stems from classical evolutionism in anthropology, which regarded the differences between cultures as an effect of an ongoing evolution from the 'primitive' to 'civilized' cultures, an evolution from the simple to the complex, relying thereby on a pre-Darwinian concept of evolution.[1] American pragmatists like William James or Charles S. Peirce were deeply influenced by Darwin.[2] Evolutionary epistemology, beginning with Georg Simmel (1895) and Ernst Mach (1905), became popular, at least in Europe, through the work of Konrad Lorenz (1941). It has been flourishing up to the present day.[3] Evolutionary ethics, evolutionary aesthetics, evolutionary economy, evolutionary psychology, evolutionary linguistics… they are all part of this 'Darwin industry.'[4]

There are also applications of evolutionary ideas in medical biology and computer science. The immune system has been shown to be a system that changes in a Darwinian evolutionary manner.[5] Computer science has developed evolutionary computing, where programs 'evolve' through mutation, recombination, breeding and selection.[6] These applications and the above mentioned approaches have different phenomena as research subject.

---

[1] See Sanderson (1990), Carneiro (2003).

[2] See Wiener (1949); Hausman (1993).

[3] For review see Bradie (1994). David Hull (1988a, 2001) developed a special variety of an evolutionary analysis of science.

[4] A term I borrow from Ruse (1996).

[5] Jerne (1955) and Burnet (1957) introduced the so-called „clonal selection theory of antibody formation". See also Jerne (1985). The model was further developed by Susumu Tonegawa (1983), who clarified how the immune system manages do guarantee the diversity of antibodies. See Cziko (1995: 39-48) for a summary of the developments. For more on the clonal selection theory, see Silverstein (1989) or Söderquist (1994).

[6] Evolutionary computing started in 1966, when Fogel, Owens & Walsh (1966) first programmed a simulation of natural selection in computers. It has been further developed by Holland's (1975) 'genetic algorithms' and by Koza's (1992) 'genetic programming'. See Goldberg (1989), Davis (1991), Koza et al (1999) and Fogel (1999) for the theoretical frame and important technical applications of evolutionary computing. See Cziko (1995: 237-260) and Nickles (2003) for philosophical interpretation in the light of a general selection theory.

## 1.2 LITERAL EXTENSIONS AND ANALOGICAL APPLICATIONS OF DARWINISM TO CULTURE

### Culture and Darwinism

Culture is a phenomenon that has been addressed by a whole cluster of theories, developed in the last decades. This cluster consists of sociobiology,[7] evolutionary psychology,[8] and human behavioral ecology[9] on the one hand. They literally extend the Darwinian paradigm. On the other hand, there are the analogical applications at issue here, the Darwinian approach to creativity[10] and memetics.[11] A further important analogical application of Darwinian ideas to culture has been developed in dual-inheritance-theories (also called gene-culture-co-evolution-theories).[12] All these approaches to culture are in some way 'Darwinian'. I will briefly introduce these five schools of thought to show the differences between them.[13]

### Literal extensions of Darwinism to culture

Sociobiology concentrated on social behavior as outcome of natural selection. Behavior or values, like the ones that guide altruistic behavior, are explained as an adaptive mean for enhancing biological fitness. Sociobiology developed further into two main schools: evolutionary psychology and human behavioral ecology.

Evolutionary psychology tries to explain behavior and culture as generated and maintained by innate, specialized, and informationally encapsulated mechanisms of the mind. These mechanisms are called 'modules.' Modules are innate adaptations to the 'environment of evolutionary adaptedness.' These modules are not merely learning devices that are

---

[7] Wilson (1975).

[8] Barkow et al (1992), Buss (2004).

[9] Cronk (1991), Smith & Winterhalder (1992), Krebs & Davies (1997).

[10] Most important are Campbell (1987 [1960], 1987 [1974a], 1974b) and Simonton (1999a, 1999b, 2001a, 20001b). But see also Stein & Lipton (1989), Cziko (1998), Dennett (2004).

[11] Most important are Dawkins (1976, 1982a, 1993, 1999), Dennett (1990, 1991, 1995, 2001a, 2001b, 2002), and Blackmore (1999, 2000, 2002).

[12] Cavalli-Sforza & Feldman (1981), Boyd & Richerson (1985), Durham (1991), Richerson & Boyd (2005).

[13] For a fuller account of the different schools see the extensive comparison in Laland & Brown (2002).

specialized for certain domains. They contain ‘content’: Some of the things we think and do are not acquired, but innate. According to this nativist view, individual and social learning is relatively unimportant for the explanation of why individuals have certain ideas, values or behavior. Hence, culture, which is often defined as relying on social learning and leading to traditions, is unimportant for an explanation of our thinking and behavior. Evolutionary psychology tends to reduce the influence of the beliefs of others to a mere triggering condition for innately specified contents of beliefs, ideas or patterns of behavior.[14]

Behavioral ecology also regards culture, and therefore social learning as irrelevant, but for different reasons: For behavioral ecology neither social learning nor innate mental mechanisms can explain what we think and do. For them, what explains thinking and acting is individual learning. Although they assume an innate learning ability, this learning ability is considered to be a mere precondition for the actual learning. The mind is filled with content mainly through learning. They are nonetheless a literal extensions of Darwinism to culture since they regard our thinking and behavior from the functional point of view as purely Darwinian in a literal sense: as adaptively optimal solutions, i.e., as fitness-maximizing solutions to adaptive problems. In each new environmental context, we adjust our thinking and behavior and optimize it thus in adaptive ways. As other animals, humans are mere biological fitness maximizers, able to overcome outdated traditions in the face of new challenges. For that reason, behavior can be predicted and explained by typical optimality models that allow us to deduce the behavior that would be optimal, given the knowledge about a certain problem in an ‘environment of adaptedness.’[15]

These three literal extensions of Darwinism are all oriented at biological fitness maximizing. In addition, they all state that culture in the sense of a system of social learning does not play an important role when it comes to the

---

[14] Most explicit on these things are Cosmides & Tooby (1992).

[15] For a more detailed comparison between behavioral ecology and evolutionary psychology see Smith (2000).

interpretation and explanation of our thinking and behavior. Even if these approaches concede to culture as a distinct domain with a distinct 'label,' culture is not regarded as important for the explanation of emotions, values, beliefs, and behavior, either because culture is considered to be based on innate characteristics of the human mind and therefore describable as the effect of biological evolution, or because individual learning is assumed to be most important for an explanation of our thinking and behavior.

## Analogical applications of Darwinism

The three analogical approaches go beyond biological fitness and innate characteristics of the human mind. And only the analogical approaches regard culture as a separate system that cannot be fully explained by reference to biological evolution. First, they normally do not assume that new behavior and artifacts, created by humans, all serve biological survival and reproduction. Second, implicitly, in the case of the Darwinian approach to creativity, or explicitly, in the case of dual-inheritance-theories and memetics, it is assumed that culture relies on a second system of cultural inheritance, namely social learning. Culture is a system of change in its own right, independent of biological inheritance and biological fitness values. Culture, in the sense of such a system of inheritance, is thus considered as important for explaining what we think and do.

Although analogical applications do not assume that culture can be reduced to biological evolution, they assume that culture changes *in the same way* as biological evolution. Analogical applications draw an analogy between change in culture and biological evolution. Cultural change is *cultural evolution*. Cultural change relies on a selective process, similar to the process postulated by Darwin for the explanation of the origin and change of species.

Notice that the term 'cultural evolution' means that culture evolves itself. To ask whether culture is in an analogical sense Darwinian is a question about its *dynamic*. To ask when and how culture arrived at the evolutionary stage of our phylogenetic history is a question about the biological *origin* of our capacity for creating culture, i.e., a question about the biological evolution of our *capacity for culture*. To avoid misunderstanding, I will use the term 'cultural evolution' only for the dynamic of culture. Only cultural evolution, and not the biological evolution of our ability for culture, will be at issue in this investigation. I will take it for granted that there are some innate phylogenetic differences between us and other species. However, these innate capacities alone do not explain the *process* of origination and inheritance of cultural novelty, if the analogical approaches are correct in assuming that culture, as a separate system of change, is important for explaining what we think and do. I will briefly introduce the reasons why I think this assumption is justified in section 1.6.

## 1.3 THE ANALOGICAL APPROACHES TO CULTURE

### History of Darwinian analogical reasoning

Darwinian analogical reasoning was used already back in the days of Darwin. Cziko (1995: 134) refers to Alexander Bain as the first one stressing an analogy between biological evolution and scientific discoveries as early as 1868. For Bain the key about scientific discoveries was trial-and-error, which was interpreted as analogous to the process of biological evolution as Darwin described it. Thomas H. Huxley, James M. Baldwin, Chancey Wright, William James, Paul Souriau, and Ernst Mach are others that have drawn an analogy between evolution and the development of human thought and mind.[16] William James, for instance, wrote in his famous essay on *Great Men and Their Environment* (1880): "A remarkable parallel, which I think has never been noticed, obtains between the facts of social evolution on the one hand, and of zoological evolution as expounded by Mr. Darwin on the other" (James 1979 [1880]: 163).

Despite these forbearers, the historical reference point of Darwinian analogical reasoning in the 20th century are two classical papers of Donald T. Campbell: *Blind Variation and Selective Retention in Creative Thought as in Other Knowledge Process* (1987 [org. 1960]) and *Variation and selective*

---

[16] See Campbell (1987 [1960], 1987 [1974a]), Richards (1987), Plotkin (1994: 61-72), Cziko (1995: 134-140).

*retention in socio-cultural evolution* (1965). The now dominating three schools of Darwinian analogical reasoning that are directed towards culture (Darwinian approaches to creativity, memetics, and dual-inheritance-theories) developed then in different directions.

## The three main analogical approaches to culture

*The Darwinian approach to creativity* treats creativity as based on a Darwinian evolutionary process – a process of 'blind' trial and selection. Donald T. Campbell (1987 [1960], 1987 [1974a], 1974b) and Dean K. Simonton (1999a, 1999b, 2001a, 20001b) will be introduced as the main defenders of this view. In creating culture, we are as blind as nature is, the 'blind watchmaker' of biological evolution. We create culture, but we are blind creators, blind minds. This point has caused considerable criticism. It will be of utmost importance what 'blindness' means in Darwinian thinking and in which sense creativity can really be said to be 'blind' in a Darwinian sense.

Memetics and dual-inheritance-theories concentrate less on what goes on in one individual. They look more at the overall process of cultural change in a group of individuals. According to them, this inter-individual process is a Darwinian process, an idea that is only immanent in the Darwinian approach to creativity.

*Memetics* relies on so-called 'memes,' basic building blocks of culture, which are considered as having analogous properties and causal roles as genes have in biological evolution. Richard Dawkins introduced this idea in his book *The Selfish Gene* (1976). It was mainly Daniel C. Dennett[17] and David Hull,[18] who backed up memetics with philosophical details. Others followed the idea with varying sophistication and emphasis.[19] For memetics, cultural items are, like genes, replicators and it is the *fitness of the meme itself* that accounts for the diffusion of cultural items. As evolutionary biology is reducible to the replication of genes, cultural diffusion is reducible to the replication of 'memes' – a process that is guided by the fitness of genes or memes alone.

---

[17] Dennett (1990, 1991, 1995, 2001a, 2001b, 2002).

[18] Hull (1982, 2000).

[19] E. g. Brodie (1995), Lynch (1996), Balkin (1998), Aunger (2002).

Organisms, in the case of genes, and minds, in the case of memes, are mere hosts that are built by these replicators. They are mere consequences of the replicative power of memes. The thesis implied in the analogy between biological evolution and cultural change thus shifts from the postulation of ‘blind watchmakers’ to the thesis that ‘no one is watching’: We can eliminate mind in our account of cultural change – if not ontologically, then as an explanatory important unit. Susan Blackmore is, besides Dennett, most famous for defending this radical thesis. At the end of her book, *The Meme Machine (1999)* she writes:

> “This is the power and beauty of memetics: it allows us to see how human lives, language, and creativity all come about through the same kind of replicator power as did design in the biological world. The replicators are different, but the process is the same. We once thought that biological design needed a creator, but we now know that natural selection can do all the designing on its own. Similarly, we once thought that human design required a conscious designer inside us, but we now know that memetic selection can do it on its own. […] If we take memetics seriously there is no room for anyone or anything to jump into the evolutionary process and stop it, direct it, or do anything to it. There is just the evolutionary process of genes and memes playing itself endlessly out – and no one watching” (Blackmore 1999: 242).

In a nutshell, according to memeticists, the unit that plays the main *causal role* in cultural change, and hence an important explanatory role, is not the human person, it is the meme itself, which is thought to be a ‘selfish replicator’ as the gene.

In parallel to memetics, Luigi L. Cavalli-Sforza & Marc W. Feldman (1981), Robert Boyd & Peter Richerson (1985), and William H. Durham (1991) developed the philosophical frame of D.T. Campbell into *dual-inheritance-theories*, quantitative theories of cultural change. They use Darwinism in the sense that they use the complex mathematical methods developed by population genetics to compute diffusion processes and the consequent higher frequency of the cultural items in a given population. They calculate the effects of different transmission systems in culture and how cultural inheritance thereby coevolves with biological inheritance. They thus apply the formalized models of population genetics and show in a statistical way how biologically maladaptive behavior can evolve on the basis of certain cultural transmission settings. Preferences for reduced family size, for instance, are maladaptive in the biological sense, since they reduce the reproductive output. These preferences can nonetheless spread in a population, if the transmission of these preferences is not vertically, between parents and children, but horizontally, between peers and unrelated people.

However, dual-inheritance-theorists do not derive any philosophical consequences from this for the traditional point of view that we are the agents of culture, consciously and intentionally creating and selecting cultural items. Although relying on the idea that culture is a diffusion process that is analogous to a selection process in nature, they deny that there is a strong analogy between cultural change and biological evolution. According to them, cultural items are not like genes, the origination is not 'blind', and the selection is driven by rational decisions of individuals. Against the Darwinian approach to creativity, they insist on 'guided' instead of 'blind' generation of novelty. Against memeticists, they state that memes do *not* replicate and that they do not have any explanatory priority over individuals. They insist that the fate of cultural items is determined by a set of multiple factors, including the human person and the structure of the social system, which are not themselves memes. These factors have an explanatory significance for the task of explaining why people exhibit this and not other cultural items. Because of these claims, I will not count them as defenders of a strong Darwinian analogy, even though they incorporate the general Darwinian analogy that culture is a selection process. I will rather take them as critics of the Darwinian approach to creativity and as critics of memetics, since the latter two definitely rely on a strong analogy and want to draw philosophical consequences from the analogy: That human minds with their goals and beliefs have less explanatory importance than thought so far, since, first, they create ideas 'blindly' and since, second, they are mere consequences of the replicative power of memes.

## 1.4 COURSE, PROCESS, AND CREATORS OF EVOLUTION

### Fact, course and process of evolution

To understand the general scope and meaning of the Darwinian approach to creativity and of memetics, I want to point to a further contrast not mentioned so far. Darwinian theories of cultural change are not merely about the *fact* that culture changes or about the *course* it thereby takes. They are about the *process* or *mechanism* of change. This is one of the differences between *Darwinian* theories of cultural evolution and pre-Darwinian classical evolutionism in anthropology, as Campbell (1965) has pointed out. The main concern of Darwinian theories of cultural evolution is not the *macro-evolutionary course* of cultural change, from simple to complex, from 'primitive' to 'civilized,' but the *micro-level process* that leads to cultural evolution.[20] The second main difference between Darwinian theories of cultural change and classical, pre-Darwinian cultural evolutionism in anthropology concerns the pattern of change that is assumed. Classical evolutionism was pre-Darwinian, since it assumed a progressive, Lamarckian pattern of change, mainly applied to whole cultures as the basic unit.[21] Darwinism assumes a totally different pattern of change. The difference between the two patterns assumed in Lamarckian and Darwinian evolution will be discussed in the following chapter 2.

### Existence of a creator

It is also very important to see that Darwinian analogical reasoning from nature to culture does not ontologically imply that there is no creator of culture. Before Darwinism came to the fore, a different but similar analogy between nature and culture was predominant. Our creativity was thought to be analogous to the creativity of a perfect creator of the world. This analogy is almost as old as philosophy. It started with Plato's demiurgos in the *Timaios*: God is manufacturing the world like a human craftsman, according to eternal forms. The famous argument from design also rests on this analogy. That is

---

[20] See also Carneiro (2003: 175), Boyd & Richerson (1985: 296).

[21] See Carneiro (2003), Sanderson (1990).

why I would like to call it the ‘design analogy’. Natural theology used it to stress that the origination and the maintenance of the order of the world (abiotic and biotic) can only be explained by reference to a designer. He created the order of the universe, the order of species, and the order inside the organisms. He did this through conscious design: *from the scratch, consciously and intentionally, with a definite purpose and a foresightful plan for its realization in mind.* The designer was thought to have done this work of design in multiple creative acts or in one great creative act at the beginning of the world. In the latter case, the creative act was imagined either as a direct creation of all existing biotic and abiotic forms, or as an even greater single act: an ingenious, foresightful fixing of the initial conditions and natural laws so that each consequent event was pre-designed. In case of the latter, the rise of new species would have been the mere unfolding of the plan of the creator, who created a perfect *creation machine*, namely nature. This understanding of novelty in nature through the idea of creation or design was paralleled or maybe even derived from our understanding of our *own* creative acts.[22]

Darwin has shown that novelty in nature can be explained by evolution. If this is right, God is not needed anymore to explain the order in nature. That is why Darwinism has been taken to refute one of the proofs of the existence of God, namely the argument from design. But neither the Darwinian approach to creativity nor memetics state that the origin and persistence of a cultural item is considered to be explainable *without the existence* of a creator of culture. These theories do not exclude that there is a creator of culture. They do not say that humans are not involved in culture, even though Darwinism in its original meaning is often understood to exclude that there is a creator of the biological world. The Darwinian approaches to culture considered in this investigation only state that the *process* how one individual or a group of individuals bring

---

[22] It is an intricate historical question to find out which idea was the basis for the other: Did we derive our image of god from our self-understanding as creative agents, or did we conversely derive our self-understanding from our image of god? – This question is a chicken-egg question that I will not try to address and that does not have to be addressed for the goal of this investigation. See Milton C. Nahm (1956: 63-83) on the history of the analogy between artists as genius and God’s creativity.

about change *is similar to biological evolution* (in the case of the Darwinian approach to creativity), or that individuals are merely epiphenomenal, reducible to the causal power of 'memes' (in the case of memetics).

## 1.5 A CRITIQUE OF DARWINIAN ANALOGIES

### Culture is important

One aim of philosophy is to show hidden patterns. This is an aim of great importance, especially with respect to scientific theories that refer to humans as thinking and acting agents. The aim of this research is thus to analyze and criticize, when necessary, the hidden patterns of the Darwinian approach to creativity and memetics: assumptions, concepts, and their consequences. However, there is one assumption that I will not analyze in detail: I will take for granted that the analogical approaches are correct in assuming that culture consists of a separate system of inheritance that is not reducible to the effects of biological evolution. I do this for the following reasons.

First, the question of culture versus innate characteristics or individual learning and creativity, when explaining modes of thought and behavior, is not an either-or question. It is a question of relative importance. The contrast between nativist extensions of Darwinism and analogical approaches should not be taken as an update of the outdated nature-nurture contrast in the sense that mind is either a 'blank slate' or totally determined by its biological heritage. Nobody believes that the mind is a total 'blank slate'. At the same time, nobody believes that there is no individual learning, and nobody believes that there is no social learning. For this reason, I suggest that the contrast between analogical approaches and the two descendents of sociobiology, evolutionary psychology and behavioral ecology, should be taken as a contrast with respect to the relative importance of different factors: (i) genes and innate mental mechanism, (ii) individual learning, (iii) culture. These factors are all relevant for an explanation of human thought and modes of behavior. Furthermore, the different explanations and methods of the approaches arise in part from different explananda: (i) psychological mechanism in the case of evolutionary psychology, (ii) behavioral fitness maximizing strategies in the case of behavioral ecology, and (iii) cultural change in the case of Darwinian theories of creativity, memetics and dual-inheritance-theory.[23]

Second, it is pretty obvious that there is at least some human thought and behavior that cannot be explained by reference to changes in gene frequencies or by reference to innate content. Culture did change independently of changes in the genetic inventory of humans. The invention of script, for instance, cannot be explained by changes in our innate and gene-based human nature, since no such genetic changes have been shown and I doubt that such genetic changes could be shown. Not every mental content is innately specified in ‘modules.’ As Kim Sterelny and Paul E. Griffiths say in their book on philosophy of biology, there is no innate “weather-prediction-module”, and beliefs about weather vary significantly across human cultures (Sterelny & Griffiths 1999: 327). At least the changes in culture that occurred over the last thousands of years cannot be explained by innate content.

Third, culture is not reducible to individual learning or individual creativity. Culture is not just a ‘cultural’ or ‘social environment’, evoking or triggering innate or learned content. Culture is created and maintained in and between humans, and only the social process of diffusion guarantees that cumulative cultural change can occur. Individual learning and creativity on the one hand and social learning on the other are two sides of cultural change. They are not opposed to each other. Culture is created by individuals and maintained through social transmission. In addition, social transmission happens in a society that has a certain communication structure that causally influences diffusion processes. Culture is thus not only filling the flexible leftovers: Although culture is in part dependent on the products of biological evolution, it exists as a second system of origination and inheritance. It is therefore in its own right important when explaining our thinking and behavior.

Furthermore, if we grant that culture is important in this sense, analogical applications of Darwinian thinking, which address the question how culture changes over time, leads to a much more radical and therefore more

---

[23] See Smith (2000) and Laland & Brown (2002) on a similar point about compatibility of the different approaches.

interesting thesis, than literal extensions of Darwinian thinking. For Daniel Dennett, for instance, the claim that biological evolution has led to this or that innate mental characteristics is just "minimal Darwinism" (Dennett 2000: ix), since it has to leave at least some space for culture. If there is some space for culture that cannot be reduced to our biological heritage, Darwinism cannot explain everything. But if culture itself is regarded as a Darwinian evolutionary process, everything falls under the umbrella of Darwinian theory. This can indeed be called "strong" Darwinism, as Dennett does (*ibid.:* ix).

The basic question of this investigation is therefore not whether culture exists as something in its own right, but whether it is, as such a system, best regarded as a second Darwinian process of change. I will assess the merits of the analogical reasoning from nature to culture, in the specific sense stated by the Darwinian approach to creativity and by memetics.

### Darwinism, creativity, and culture

In order to assess the analogies, it is important to be very clear about Darwinism as such. What are the characteristics of a system that changes in a Darwinian manner? Unless this can be answered clearly, one cannot see in which sense the Darwinian theory of creativity and memetics are *Darwinian* theories and what that entails. First of all, Darwinian evolution has to be distinguished from change through God-like creation. But it also has to be distinguished from Lamarckian evolution, which is as gradual as Darwinian evolution, but different in other respects. Furthermore, Darwinism itself has changed since 1859, the year of the first publication of Darwin's *Origin*. Today, there are different versions of Darwinism with different claims. Thus some claims of the analogies might rest on a very specific version of Darwinism. These issues about Darwinism itself will be addressed in chapter 2 – a detailed, systematic reconstruction of Darwinian thinking for the purpose of this inquiry.

### Three basic analogies

After this foundational chapter, I will introduce and discuss what I call the three basic analogies. They are, first of all, distributed unequally across the two analogical applications at issue here, and, second, they can *in principle* be at the foundation of any version of analogical reasoning from biological evolution to cultural change. These three basic analogies are the *ontological analogy,* dealt with in chapter 3, the *origination analogy,* at issue in chapter 4*,* and the *explanatory units of selection analogy,* at issue in chapter 5. In the following, I introduce them briefly and justify this partition.

The *ontological analogy* includes the claim that those kinds of entities that are the building blocks of culture share basic features with those entities that are the material basis of biological evolution. The entities that are discussed today are not complex artifacts as analogous to organisms, but ideas, values, or instances of behavior as analogous to genes. An important feature of genes is the way they secure their persistence through time. Genes replicate. Memes are claimed to do the same. Therefore, both count as replicators, a special kind of entity that is claimed to be essential for any evolutionary change. The ontological analogy is therefore dependent on the kind of process that secures persistence. The ontological analogy includes that the process of transmission (and therefore duplication) of cultural units, from one person to the next, is similar to the transmission of genes. It includes that cultural units are replicators.

The *origination analogy* includes the claim that the *origin* of certain features of ideas (and of the behavior or artifacts based on them) is analogous to the origin of certain features of organisms. The origination analogy therefore refers to the *general pattern how certain changes arise*. Is the pattern more analogous to a God-like creation, to a Lamarckian kind of evolution, or to a Darwinian kind of evolution? Even if the origination of cultural novelty can be shown to follow a Darwinian pattern in principle, because it involves a selection process (the differential spread of cultural items), it still stands to question whether this selection process is Darwinian in a more narrow sense, namely based on *blind* variation and selection.

The *explanatory units of selection analogy* is that analogy for which memetics has become very prominent. It refers to the *causal and explanatory roles* certain entities play, not to the basic properties of these entities, nor to the basic pattern of change. The analogy is a direct transfer of the *units of selection debate* in biology to the cultural sphere. Gene selectionism (also called the doctrine of the ‘selfish gene’) states that genes are the units of selection: those units that have a causal and explanatory priority compared to others, such as organisms and groups. Through the explanatory units of selection analogy, gene selectionism becomes meme selectionism: memes have an analogous special status in culture. According to this analogy, human individuals are merely the effects and the ‘survival machines’ of memes, as they are mere survival machines of genes, serving the ‘selfish interests’ of these genes or memes. This idea does not only comprise that there are memes, it includes that memes and not minds explain culture – as it is genes that explain biological evolution, and not organisms and groups, which are secondary for the purpose of explanation. In both cases, it is important to ask whether these theories distort the causal picture of change, by giving replicators a special causal and explanatory status.

To summarize, (1) the ontological analogy is an analogy with respect to the entities involved in the respective changes; (2) the origination analogy is an analogy with respect to the kind of pattern accounting for the origination of cultural items; (3) the explanatory units of selection analogy is an analogy with respect to the causal and hence explanatory roles certain entities play. The difference between these three basic analogies also becomes evident, if we look at the different questions that can be associated with them: (1’) What is the x whose selection and/or evolution is at issue? (2’) How does x come about and persist? (3’) Why does x spread and persist?

The three basic analogies are present to a different degree in the Darwinian approach to creativity and in memetics. Darwinian theories of creativity are mainly interested in the origination question. Memetics mainly addresses the units of selection analogy and is not interested in the origination question. All memeticists and some defenders of the Darwinian approach to creativity assume that there is an ontological analogy.

Although the matrix of the three underlying basic analogies adds complexity in the sense that it explicitly adds a new layer of analysis, it nonetheless simplifies the issue. First, the three basic analogies make the structure of the overall issue explicit. Furthermore, through this procedure one should also be able to assess the merits of other actual or potential analogical applications of Darwinism to mind and culture, if they draw on one or more of the three basic analogies. Last but not least, it has the advantage that one can assess each of the three basic analogies separately, in order to show the problems of each of them. This also allows to analyze in which way they are dependent or independent of each other.

It might well be, for instance, that the origination analogy and the explanatory units of selection analogy are not dependent on the ontological analogy. In case of such an independence, the failure of one basic analogy does not imply the failure of the other, and the success of one basic analogy does not imply the success of the other. One analogy can then not be used to argue for or criticize the other one. It might be objected that through treating the three analogies separately I might already assume that they are independent. But that something can be distinguished from something else does not mean that the two things at question are independent. So far, I only stated that the three basic analogies can be distinguished and should be assessed separately. Whether they are dependent will be addressed in the respective chapters.

### Evaluating an analogy

Since the analytical goal of this investigation is to evaluate analogies, the following methodological question arises naturally: How can one judge an analogy? Analogies have long ceased to be excluded as a valuable tool from science. One cannot condemn them outright as pure ‘metaphoric’ or ‘unscientific’ devices. Darwin also used an analogy in his construction of his evolutionary theory: He compared natural selection to artificial selection of animal breeders. Generally, analogies are judged valuable, if they help increase the knowledge of a yet to develop domain by using the knowledge from another well-developed domain.[24] Darwin's analogy is a perfect example for this.

Analogies can have such a positive role thanks to similarities between the base and the target of the analogy. But analogies never state similarities in all respects, i.e., a total equivalence of the base and the target of the analogy. In Aristotle's words, an analogy states "similarity in dissimilars."[25] We can therefore not condemn an analogy simply because there are differences between the base and the target. As Paul Thagard has put it, "there is much more to evaluating an analogy between a base and a target than just counting their similarities and differences" (Thagard 1988: 101). The evaluation of analogies has to be with respect to their "role in problem solving" (*ibid.*: 101), as he puts it. This role must be a heuristic one: The similarities must *add* something that is new and appropriate to our knowledge of the target domain.

Applied to the issue of this inquiry, the important question is whether the Darwinian theory is really a theoretical tool that offers a viable and new description or explanation of creativity and culture. Therefore, I will base my evaluation of the Darwinian analogies on the following three questions:

*(i). Descriptive adequacy.* First, are the analogies *descriptively adequate*? In order to be descriptively adequate, the analogy must lead to a correct description of the phenomenon at issue. If the application of the analogy fails to give a correct description, then this is because it states similarities where only dissimilarities exist. But which dissimilarities between a base and a target count as a refutation of the analogy? As just mentioned, not every dissimilarity diminishes the value of the analogy. I will solve this problem by demanding the following: A correct description of the phenomenon must include *relevant* similarities between the two systems of comparison, in our case, biological evolution and cultural change. Relevant similarities are central for the phenomena at issue and cited in justification for the analogy. For

---

[24] For more details on the scientific role of analogies, further references and examples see Bradie (1998), Gentner & Jeziorski (1993), or the seminal accounts of Black (1962) or Hesse (1963).

[25] The Works of Aristotle, vol. XI: 1459 a5ff (Oxford: Clarendon Press, 1952); quoted after Bradie (1998: 316).

instance, memeticists state that memes and genes are both replicators. In case this leads to a correct description of culture, then the analogy is descriptively adequate, even if memes are different than genes in other respects.

*(ii). Explanatory force*. Since descriptive adequacy is not the only standard that an analogy can be asked to meet, the second question is: Do the analogies contain more than a correct description of the phenomenon, i.e., do they contain an *explanation* of the phenomenon? In order to answer this question, some account of explanation is necessary. I will not develop or assume a *general* account of explanation. What it means to explain something is itself a problem in philosophy that has attracted considerable attention since a long time. It continuous to be one of the unsolved problems in philosophy. However, I will confront Darwinian analogical reasoning with two standards, that can be demanded from it, even if it is unclear what explanations *in general* requires. First, explanations should not be tautological. Some correct descriptions of a phenomenon might appear to be an explanation but in the end fail to really be one. A famous example is the tautological explanation of why opium causes sleep in Molière's play *The Imaginary Invalid*: A physician says in this play that sleep is caused by a *virtus dormitiva*, a virtue that causes sleep, a virtue supposed to be in the opium. This explanation does not add anything to the question asked and is therefore tautological. Second, apart from tautological explanations, I will demand the following standard, which the analogies have to fulfill in order to have an explanatory value: A thesis contains an *adequate* explanation, only if it reaches the level at which an explanation is usually sought for in a certain discipline. If the analogies cannot give such an adequate explanation, I will call them *trivial in terms of their explanatory force*. As just mentioned, philosophy has always had and still has difficulties with a *general* concept of explanation – a concept of explanation that covers different kinds of explanations *at the same time*: everyday explanations, psychological explanations, historical explanations, explanations that involve deduction, explanations that involve inductive statistical

inferences, or biological explanations that involve functions.[26] For the standard, with which I want to confront the three basic analogies, we do not need such a general account. We only need a comparative account of what kinds of explanations are considered as appropriate in the respective disciplines that have creativity or cultural change as one of their basic subjects of investigation. Psychologists try to explain creativity by pointing towards psychological processes or 'mechanisms' involved in creative thinking, such as perception, conceptual combination and the like. I will show in which sense the origination analogy fails to give an explanation comparable to this standard in psychology. Likewise, psychology also explains social learning at the level of such cognitive processes or 'mechanisms,' i.e., in terms of those basic cognitive processes or 'mechanisms' involved in social learning. I will show that the ontological analogy fails, since it rests on an account of learning that is either false for most cases of social learning, or does not offer an explanation at the level of cognitive processes at all. Note that such explanations, in terms of basic cognitive processes, do not entail that these processes or 'mechanisms' have yet been described and explained at the level of the brain, or that these mechanisms can be modeled in a computational way. To point to certain cognitive processes is certainly not a full psychological explanation in terms of anatomical structures and the causal processes going on at that material level. Usually, it does also not allow for prediction, deriving an output from an input-history according to certain 'principles,' 'laws,' or 'rules.' This is not the standard reached so far in typical psychological explanation of how people are able to be creative or learn from others. Last but not least, with respect to the explanatory units of selection analogy, the standard, with which I will compare the analogy, is the standard that is used in social sciences, such as economy, sociology or anthropology. When these disciplines explain diffusion of cultural items, they refer to choices and judgments of individuals, which are assumed to be made by these individuals according to certain beliefs and values. It is a standard that does not even reach the level of basic cognitive process, but is not necessarily incompatible with any further explanation at a deeper level, be it at

---

[26] See Kitcher (1998) for a summary of debates about explanation.

the level of psychological processes or at the level of a truly naturalistic explanation in terms of causal laws and initial conditions.[27]

*(iii). Heuristic value*. Last but not least, the third question on which I will base my evaluation of the Darwinian analogies is: If correct descriptions or explanations are given at all, are these descriptions and explanations heuristically valuable, i.e., 'fruitful'? In my terminology, they are heuristically valuable or fruitful, if they add something to our current knowledge about the phenomenon at issue, either through adding a description or an explanation of a phenomenon that highlights something *that is invisible from another already available explanation or description*. If they are not heuristically valuable in this sense, if they merely reinvent the wheel, I will call them *heuristically trivial*. In such a case, the respective analogical application of evolutionary theory would be done by a mere superimposing of a new language on old insights. The application would come down to a superficial 'perspective' that can be chosen, if one wants to, motivated by various reasons, but not by facts, since the facts can be described in a Darwinian terms or not.

To summarize, my evaluation of the three basic analogies will be guided by three questions: (i) whether the analogies are descriptively adequate, (ii) whether they have explanatory force, and (iii) whether these descriptions or explanations have a heuristic value. In case the analogies fail to be descriptively adequate, I will consider them *wrong* with respect to the similarity or dissimilarity at issue. In case they fail to have much explanatory force or fail to be fruitful, I call them *trivial*, either in terms of explanation or heuristics.

### What is at stake for us

My critical evaluation of the three basic analogies is guided by a meta-question about the general implications of the analogies for our self-understanding: What is at stake *for us* in the description of creativity and culture in terms of the resources provided by modern Darwinian evolutionary theory? What *role* does the individual human being play in Darwinian theories of cultural change?

---

[27] See Rosenberg (1998) on this latter issue.

The traditional roles of individuals in cultural change are the following: Human individuals create cultural items, choose between created cultural items, and build on the ones they adopt thereby. In short, humans are the creators, the *authors* of cultural change. This is the traditional vision of human beings as agents of culture. We are consciously designing and transmitting culture with our creative mind. This vision of ourselves as agents of culture is often taken to be at stake when evolutionary theories are applied to culture. It is the aspect that makes these applications suspicious for most laymen and scholars alike. Many are afraid of these applications. But maybe there is no reason to be afraid. It may well turn out that analogical applications of evolutionary thinking are not in conflict with this traditional conception, since they are trivial in the sense specified above. Maybe "Darwin's dangerous idea" (Dennett 1995) is less dangerous as it appears in the first place. Maybe Darwin's idea is, in the form it is at issue here, a mere restating of what we always knew. The usual error in debates about applications of Darwinism to culture is to take for granted that the vision of us as the agents of culture is at stake. Therefore, to find out in which sense the analogies are indeed in conflict is an important philosophical issue. As mentioned above, the conflict is not so severe that a Darwinian account of creativity and cultural change denies that individuals are involved as creators. But a Darwinian account may have implications for the way we conceptualize our creativity and our capacity to choose between different cultural items, since it states that we create and choose novelty as 'blind minds' similar to blind natural selection, and that memes and not we direct the diffusion in culture.

Freud once wrote about the three narcissistic insults of humanity by science, honoring his own 'discovery' of the unconscious as the third insult. The first insult was the Copernican revolution. Darwin's revolutionary theory was the second one, since it means that humans are not the necessary destiny of the world, that the world was not created *for them* by God, that they are just animals. Accordingly, the meta-question of this inquiry is whether we have to face a fourth insult, a 'second Darwinian revolution', where not only our human nature, but also culture is explained by the *evolution of memes* and not by our God-like creativity and ability to choose between cultural items – just as nature has been explained as evolution rather than God’s creation in the first Darwinian revolution.

# 2 THE STRUCTURE OF DARWINIAN EXPLANATIONS OF CHANGE

## 2.1 GENERAL SELECTION THEORY

### Fact, mechanism and formal principles of evolution

Charles Darwin's main contribution to our understanding of evolution was twofold. First, he managed to gather enough evidence to show the *fact* of evolution. He could convincingly show that species were not independently created, that later ones descended from earlier ones and evolved in a gradual way. Even more important was, secondly, that he put forward a special *mechanism* for evolution, namely *natural selection*. This mechanism is supposed not only to describe evolution as a gradual change, but also to explain the respective change. A Darwinian explanation of change differs significantly from its two great rival explanations of the origin and characteristics of species, the creationist and the Lamarckian explanation. I will present these alternative explanations in more detail in section 2.2 – to show by contrast what a Darwinian explanation of change comes down to. Before that I want to point at an abstract account of Darwinism, a general selection theory, and at some issues that will be addressed later in this chapter.[1]

A general selection theory provides an abstract, general account of evolutionary thinking in the Darwinian legacy. It asks for formal or abstract principles that have to be met in order to call the dynamics of a system Darwinian. The generality of such a formal account has important merits. First, a formal account is indifferent to the ontology of the respective system. The account does not distinguish between a biological system, which is based on genes, and a computer program, which is based on digits. Hence it does not rule out *from the start* that there are other systems than the biological system of

---

[1] The term 'general selection theory', used for instance by Campbell (1997), has gained considerable dominance in the last years. The similar terms 'universal Darwinism', introduced by Dawkins (1983), or 'universal selection theory', used for instance by Cziko (1995), can also be found frequently.

genes, organisms and species that change in a Darwinian manner. If a system meets the formal principles, then it can be regarded as changing in a Darwinian manner – despite ontological differences. Second, such a general perspective serves the task of evaluating analogies, since it avoids disputes about differences between culture and organic evolution that are irrelevant for the analogy. An example for such a difference would be that culture is not based on sexual reproduction, while organic evolution is. Despite this difference, both can be Darwinian systems from a general point of view. Although the DNA-based biological evolution was the first Darwinian process discovered, not all of its characteristics are essential for considering it as Darwinian. According to "Campbell's Rule,"[2] when drawing analogies between nature and culture, one should not directly compare special features of one system with the features of another one. Rather one should start with a *general* theory: with general principles that are essential for a Darwinian system and therefore relevant for an evaluation of analogical applications of Darwinian thinking.

What are these general principles? Darwin himself (1859: 343 and 459) is often quoted with his formula of the *theory of descent with modification through natural selection*. Today, evolution of a system through natural selection is usually described by three principles. If a system contains *variation, differential fitness* and *heritability,* natural selection and adaptation follow. The classic reference for this description of evolution by natural selection is Lewontin (1970). According to Lewontin (1970: 1),

> "Darwin's scheme embodies three principles:
> 1. Different individuals in a population have different morphologies, physiologies, and behaviors (phenotypic variation).
> 2. Different phenotypes have different rates of survival and reproduction in different environments (differential fitness).
> 3. There is correlation between parents and offspring in the contribution of each to future generations (fitness is heritable).
> These three principles embody the principle of evolution by natural selection. While they hold, a population will undergo evolutionary change."

This is the "logical skeleton" (Lewontin 1970: 1) of Darwin's mechanism of natural selection. Note that it leaves out 'overpopulation' and the resulting

---

[2] See Durham (1991: 187 or 425f), who introduced the term. Campbell stated his 'rule' in Campbell (1987 [1960]). For a recent formulation see Campbell (1997: 6) or Hull (2000: 45f).

competition as criteria for natural selection. Competition in this sense is often cited as necessary for selection to occur. Contrary to this, Lewontin states that "[n]atural selection [of bacteria, MK] occurs even when two bacterial strains are growing logarithmically in an excess of nutrient broth if they have different division time" (Lewontin 1970: 1). As Lewontin's example shows, competition is not important for Darwinian evolution through natural selection. Even in good times the 'good' ones will have a higher fitness.

### The essential blindness of Darwinian evolution

What is, however, of utmost importance, even though it is totally left implicit in the three principles, is the so-called 'blindness' of Darwinian evolution. If the three principles are fulfilled, selection can occur, leading to adaptations as products of multiple steps of selection. However, the Darwinian paradigm describes, explains and predicts adaptation *without a designer*. There is no 'watchmaker' or 'designer' of organic evolution. This is why the natural process of evolution is called *'blind.'* Even if we describe nature – metaphorically – as a 'watchmaker,' nature is a 'blind watchmaker.' I will show in section 2.3 that the common, but metaphorical expression of 'blindness' amounts to different aspects of Darwinian evolution: Darwinism explains adaptation with repeated cycles of *undirected variation,* followed by a selective process, which can only proceed *ex post facto*, i.e., after the occurrence of variants, and only in an *opportunistic* manner.

Through these aspects of nature's blindness, Darwin initiated two major changes in the understanding of evolution. In Darwin's theory, the inclusion of blind, i.e., undirected variation led to a new pattern of change, namely *variational* evolution. Through this and through the opportunistic manner of selection, he completely replaced not only the idea of divine creation but also a certain idea of progress. The latter sharply distinguishes the Darwinian concept of evolution from Lamarck's theory of evolution, the second grand evolutionary theory that arose in the 19th century. Neither in divine creation, nor in Lamarckian evolution does blindness play any explanatory role. Therefore, blindness, in its different aspects, can be regarded as the distinguishing feature of a Darwinian evolutionary process. *Gradualism* can be part of all three models. Therefore, it cannot count as distinctively Darwinian.

### Explanations of origination and fitness differences

There are further aspects of Darwinian thinking that are only implicit in the skeleton but important for the evaluation of the three basic analogies. The first is about explanations of origination and fitness differences. Evolutionary theory as a theory about the mechanism of change does not have to address the origin of life. Apart from the very first moment of the appearance of biological substances on earth, evolution can treat every question of origin as a question about the change from one species to another one. Evolutionary theory has, however, to answer other questions of origin in order to explain the respective change. At each step in cumulative cycles of *variation, differential fitness, heritability*, and consequent *selection*, the explanation of change must mean an explanation of the origin of the novel features that lead to the variation. Otherwise it would, first, not really explain the *origination* of new characteristics, but only their maintenance in each cycle of the cumulative evolutionary process. Second, it would not even explain the maintenance of new characteristics in each cycle, since it would not be able to explain the fitness differences leading to a differential maintenance of traits, since fitness differences are an effect of novel features.

The three principles only describe the dynamics of a sorting process. In order to *explain* evolution, evolutionary theory must explain the origin of variation that leads to fitness differences and hence fuels the sorting process. The first issue about origination is relevant for the origination analogy. It shows that the main question about origination is one about the origin of variants and not so much about maintenance. The issue about the Darwinian account of the origin of variation will dominate section 2.3. However, the second issue concerns the Darwinian explanation of the maintenance of characteristics, which has been criticized for being tautological. I will explain the latter issue in more detail in section 2.4. The issue is important for the explanatory units of selection analogy.

## Ontological generality

Ontological generality is a further aspect of the three principles that has to be taken seriously. As Lewontin states, the three principles can be applied to different entities and mechanism of inheritance:

> "No particular mechanism of inheritance is specified, but only a correlation in fitness between parent and offspring. The population would evolve whether the correlation between parent and offspring arose from Mendelian, cytoplasmic, or cultural inheritance. […] The generality of the principles of natural selection means that any entities in nature that have variation, reproduction, and heritability may evolve" (Lewontin 1970: 1).

In principle, we just have to replace 'individuals' in the above cited threefold recipe with another entity, check whether the principles are fulfilled, and we get other "units of selection" (*ibid.*). Groups of organisms as well as cultural entities can be such units of selection. Gene selectionism, however, argues that neither groups nor individuals can be units of selection in biological evolution. This thesis is the main issue of the famous *units of selection debate*, which has been a major concern of evolutionary biology for decades.

For gene selectionism, there is only *one* fundamental unit of selection, namely the gene. One of the main arguments is that the individual is not preserved through evolution, since it dies and since the respective offspring are never a copy of one of the parents, even if there is heritability of traits. For evolution through natural selection to occur, gene selectionism demands that there must be a so-called replicator that persists over time. The replicator and only the replicator is the real unit of selection. Evolution is *differential replication*. Thus, heritability, the third of the three principles of evolution by natural selection, is regarded as not enough for evolution to occur. Heritability between offspring and parents means that 'like begets like', that parents and offspring are *more* similar than other pairs of individuals, and that the similarity between parent and offspring is due to shared genes and not due to environmental influence. In the case of sexual reproduction, heritability does not mean, and cannot mean that offspring have all the genes of each of the parents.

I will say more on the argument that organisms cannot be units of selection in section 2.5. At this moment, it is merely important to realize why this issue is important. From the point of view of gene selectionism, evolution in general becomes *replicator selectionism*. Whatever change is at issue, the unit of selection must be a replicator, be it genes in nature or so-called 'memes' in culture. Therefore, the units of selection debate is important not only for the theory of organic evolution, but also for the analogy to culture, mainly for the ontological analogy and the units of selection analogy. Furthermore, it shows that there are different versions of Darwinian theory, gene selectionism and traditional individual selectionism.

## 2.2 PATTERNS OF CHANGE

### Creationism, Lamarck, and Darwin

In the following two sections, I want to give a detailed picture what Darwinian evolution is by showing the difference between three patterns of change: creation, Lamarckian transformational evolution and Darwinian variational evolution. In this section, I will first introduce the pattern of change that was postulated by 'arguments from design': divine *creation*. I will then introduce Lamarck's theory of evolution as an example of a second pattern of change, namely *transformational evolution*. Although transformational evolution does not postulate a designer, it still does *not* imply the aspects of blindness involved in a Darwinian *variational pattern of evolution*, which will be presented at the end of this section. The different aspects of blindness will then be described in detail in section 2.3.

### Creation

Ernst Mayr writes that "[n]o consequence of Darwin's theory of natural selection was a source of greater dismay to his opponents than the elimination of design from nature" (Mayr 1976 [1962]: 30). In the context of evolutionary debates, the term 'design' can refer to two different things: (i) to *properties of an object* that is regarded as the designed thing, such as order, adaptedness, function, complexity, etc.; (ii) to the *process of designing* the order, adaptedness, function, complexity etc. by a conscious, foresightful plan.

What Darwinism eliminated was not (i) design in the sense of properties of organisms, but (ii) design as a process that explains the existence of (i). The debate about Darwinism and creationism is a debate about the question whether we really need (ii), a design process, in order to explain (i). In order to prevent confusion about the term ‘design,’ I will henceforth use the term ‘creation’ for the idea of a design process.

The idea of creation is as old as philosophy. In the *Timaios* (27a ff), Plato introduced the demiurgos as the creator of the world of becoming, which he regarded as derivative from the eternal forms, templates after which the world is created. The creator did not only bring about the adaptive characteristics of organisms, but also the harmony of the whole universe, which contains the “great chain of being” (Lovejoy 1936), i.e., an order of the species that can be depicted as an ascending line.

In order to elicit what characterizes creation and how it differs from Darwinian evolution, I will start with the famous ‘argument from design.’ The *argument from design* goes back to Plato and Aristotle and was advocated by Thomas Aquinas in his *Summa Theologiae* as the fifth way to prove the existence of God. It has received a devastating critique by David Hume in his *Dialogues Concerning Natural Religion.* It nonetheless survived and had its “heyday” in 19th century pre-Darwinian times, when evolutionary thinking was already in the air (Sober 2000: 30). William Paley became its spokesman. His form of the argument, which was laid down in his *Natural Theology* (1802), became the template for creationist arguments against Darwinism. That is why I will concentrate on his version of the argument from design.

At the center of Paley’s argument from design stands the adaptivity of organisms, not the order of the whole universe or the great chain of being, although they are important as well. Paley used the analogy between artifacts and organisms, as others did before him. Since the opening paragraph of his *Natural Theology* has become the “conceptual center of gravity for the entire discourse” (LeMahieu 1976: 5), it should not be withhold from the reader:

> “In crossing a heath, suppose I pitched my foot against a *stone*, and were asked how the stone came to be there: I might possibly answer, that, for anything I knew to the contrary, it had lain there for ever […] But suppose I

> had found a *watch* upon the ground, and it should be inquired how the watch happened to be in that place; I should hardly think of the answer which I had before given, - that, for anything I knew, the watch might have always been there. […] when we come to inspect the watch, we perceive (what we could not discover in the stone) that its several parts are framed and put together for a purpose, *e.g.* that they are so formed and adjusted as to produce motion, and that motion so regulated as to point out the hour of the day […] it requires indeed an examination of the instrument, and perhaps some previous knowledge of the subject, to perceive and understand it; but being once, as we have said, observed and understood, the inference we think is inevitable, that the watch must have had a maker: that there must have existed, at some time, and at some place or other, an artificer or artificers who formed it for the purpose which we find it actually to answer: who comprehend its construction and designed its use" (Paley 1825 [1802]: 1-2).

In a nutshell, the "argument from design" then is:

> "There cannot be design without a designer, contrivance, without a contriver; order, without choice, arrangement, without anything capable of arranging; subserviency and relation to a purpose, without that which could intend a purpose; means suitable to an end, and executing their office in accomplishing that end, without the end ever having been contemplated, or the means accommodated to it. Arrangement, disposition of parts, subserviency of means to an end, relation of instruments to a use, imply the presence of intelligence and mind" (*ibid.*: 8).

This holds for works of human creativity, such as a telescope, as well as for "the works of nature", such as an eye: "[T]he eye was made for vision, as […] the telescope was made for assisting it" (*ibid.*: 12-13).

Debates about the argument from design are abundant. Some address the logical form of the argument.[3] However, one of the most complicated issues is the precise nature of (i), design in the sense of properties of objects. What is design? Paley used terms like "purpose and design," "contrivance," "arrangement, disposition of parts, subserviency of means to an end, relation of instruments to a use," "organized," "complicated," or "complex." Dawkins uses "adaptive complexity" (Dawkins 1983, 1986). Contemporary creationists stress "complex specified information" (Dembski 1998) or "irreducible complexity" (Behe 1996). Despite differences, they all more or less stress two aspects: complexity and function. In a nutshell, this means that the parts of an organism, as well as the parts of a watch, are arranged in a complex way and serve a purpose. They have a function. However, it is still unresolved how to *understand* complexity and functions precisely.

---

[3] For instance: LeMahieu (1976), Sober (2000), Ruse (2003).

The latter question would lead us into issues that deserve an extended treatment on their own, which cannot be given here. Furthermore, what is important in this inquiry is not (i), but the assumptions about (ii), the idea of creation as a design activity itself. However, for analyzing (ii), a few words on functions are needed, since the idea of creation explains the 'internal teleology' of organisms (i.e., the functions of organs, which serve the well-being, including survival and reproduction, of organisms), as effect of an 'external teleology.'[4]

Darwinists as well as creationists assume *internal teleology*. Even if Darwinists do not assume a creator, they nonetheless believe that there is something special about organisms that needs a special explanation. The existence of an organ like the eye has to be explained differently than the existence of a stone, as Paley stressed. The explanation uses functional language, which enters biology because at least some organs can be explained as adaptations.[5] An eye, for instance, is considered to be an adaptation since it has been selected because it enables organisms to see. Because of this selective history of the organ, seeing is the function of the eye. Let me explain this. If we want to explain why a bird has an eye, we could cite facts about embryological development or genetic dispositions for such eyes. But this does not answer why the eye came into existence at all, why there are birds with genetic dispositions for eyes. Hence, the answer must go beyond this developmental point of view. The evolutionary point of view suggests that the eye exists, because eyes are good for seeing. In other words, eyes exist because to have them was beneficial for the well-being of the bird. It therefore enhanced its survival and reproduction. This is meant when it is said that the bird has an eye because the eye fulfills a specific function. Thus, when I say 'function,' I mean what has been called 'etiological functions' (Wright 1973). Current debates about functional talk distinguish this notion of function, which is linked to a selective history, from a notion of function that wants to define function

---

[4] See Ayala (1970, 1977, 2001), Lennox (1992), or McLaughlin (2001) on the distinction between internal and external teleology and the history of the distinction in the different concepts of teleology in Plato and Aristotle.

[5] Dawkins (1983, 1986), Ruse (2003).

irrespective of selective history, often called Cummins-function, after Cummins (1975). Cummins and his followers argue that the notion of function should not be linked to adaptation and selective history and that etiological accounts of function are mere vestiges of the old idea of creation. I do not want to take a stand on the question how function should be properly defined.[6] What I want to say is the following: First, although talk about adaptations might do without the *term* function, they cannot do without talking of beneficial effects. Second, the explanation of an adaptive organ rests on the assumption that organ has been selected because of its beneficial effects *for* the survival and reproduction of the organism. Both creationism and Darwinism rely on such reasoning. The former say that the organs have been *created* because of these effects, the latter that it has been *selected* because of these effects. The difference between 'created' and 'selected' is not one about internal teleology. It is a difference about external teleology, to which I turn now.

*External teleology*: The term 'external teleology' refers, first, to an intentional plan, which is external to the product having a certain function, and, second, to an external value of the object – a value for a creator or perceiver.[7] That is the point where Darwinians depart from creationists. Creationists assume that the functionality, the internal teleology of organisms, did come about because a creator designed them to have this internal teleology. God created the eye because it enables animals to see. Such explanations refer to an anticipation of future outcomes and the planning and desire of these outcomes in the mind of a creator. And it is precisely *because* of the pre-existing *external* anticipation and intention that the eye exhibits a certain function. An external

---

[6] See on the issue about functions in biology the papers in Allen et al (1998), Buller (1999), or Ariew et al (2002).

[7] The second aspect of the concept of 'external teleology' is not important here, but see McLaughlin (2001), who shows that this second aspect is important to understand the difference between function attributions to artifacts and function attributions to organisms: artifacts do not have an internal teleology, although we can attribute functions to them, since they do not have a good for themselves, whereas organisms do have such a good for themselves. "The reason why adaptations have functions is not because they contribute to the performance of the function of the organism for some external or containing system but rather because they are useful to the organism. They are good not for external beneficiaries of the functions of organisms but for the organisms themselves. This is where the analogy between artifacts and organisms breaks down" (Mc Laughlin 2001: 145).

teleology leads to, and consequently explains, the internal teleology of the eye. In Darwinian functional explanations, the internal teleology of the eye is not understood as superimposed by an external agent. The eye evolved gradually through the selective interaction between a lineage of organisms and their selective environment. The effects of selection then lead to adaptation or function, but it does not involve an intention of a designer. Selection thus leads to *internal teleology without external teleology*.

Before I explain in a more positive way what such a Darwinian selectionist explanation amounts to, I want to say a little bit more about the idea of creation and the pattern of change assumed in that idea. *Divine creation* – the designing activity of a deity – is usually understood as unobservable as the deity itself. We can only infer the creative process from the products of creation. But Paley tried to say something about how God is doing his job and about which attributes such a creator must have in order to do his job. According to Paley, organisms show – as do artifacts – complexity and functions. Complexity shows the designer's wisdom, and functions show the designer's purposes and benevolence. Let us imagine that God saw, for instance, that it would be good for this or that type of organism to have vision. He recognized then what is needed for a device that enables a creature to see. He thought up a structure of an organism and its parts so that the respective organism can have vision. Finally, he created this type of organism. That is the idea of God as creator of adaptations. But God's purpose and benevolence is not only visible in the diverse adaptations of the diverse species. It is also visible in the *order* of the species. God created the universe in order and harmony. For the biological sphere, the "great chain of being" (Lovejoy 1936), a chain from the simplest to the most complex organisms, is the order that gives every species its place and secures harmony. This means that the deity created the world according to a "rational plan of creation that human mind might hope to understand" (Bowler 2003: 28). This "*master plan*" was not only thought to make the biological realm rational, but also to give meaning to life and the struggle involved in it (LeMahieu 1976: 71). Whatever happens, it is all part of the master plan and makes sense as a consequence of the deity's benevolence and perfection. Furthermore, since the whole creation is perfect, every link in the chain was created or at least planned from the start. Since the order that springs from this divine master plan was thought to be perfect, there was no place for extinction and evolution of species. Although there might be individual differences, species have unchangeable essential characteristics, which are perfect and created at the beginning. Individual differences were thought to exist, but the variation based on these differences was thought to vanish as soon as it comes into existence. Variation was considered as negligible noise in the system. From the point of view of the idea of divine creation, the biological realm is thus static, perfect and conceived in essentialist terms. The pattern of change assumed in such a creationism actually is no pattern of change. It is the negation of change; it is a static pattern. Below I will show how such a static creationism can nonetheless turn into a deistic evolutionism.

Now, how is the creator doing his job? In the chapter on the personality of the creator deity, Paley (1825 [1802]: 284) writes that the creating agency, which brings about the adaptations and the chain of being, is “that which can contrive, which can design.” It “must be a person,” since contrivance implies “consciousness and thought.” This means that there must be a creator who “can perceive an end or purpose,” and who has the “power of providing the means, and of directing them to their end.” According to Paley, creation is thus characterized by at least five elements: *personal unity, purpose, power, knowledge and foresight*.[8]

In addition, the following points have to be taken into account: Such a concept of creation also means (a) that everything is decided by an agency that is *external* to the products, as shown above, and (b) that everything is decided already *before* the actual occurrence of the organisms, and (c) that the deity can always start *afresh*. The latter two points are important for our comparison between the idea of creation and the blindness of Darwinian evolution. Because

---

[8] Paley (1825 [1802]) also mentions various further attributes of the deity: omnipotence, i.e. the power to create the world and the adaptations; omniscience, which is the base of knowledge and foresight; omnipresence; eternity; self-existence; necessary existence; spirituality; unity; goodness.

of perfect knowledge and foresight, the creator – if as perfect as God – does not have to try or test. God does not make errors. He plans and thinks before he acts. Since he knows what is going to happen, he can avoid errors. Everything is perfectly designed *before* the occurrence of the products, thanks to the deity's perfect foresight. As Dawkins writes, "[a] true watchmaker has foresight: he designs his cogs and springs, and plans their interconnections, with a future purpose in his mind's eye" (Dawkins 1986: 5). This also means that God creates *from scratch.* Since he can do this all the time, he can start a new blueprint for each species whenever he wants. He does therefore *not have to build on old stuff.* As perfect and omnipotent as he is, he does not have to tinker. I will show in section 2.3 that nature's blindness, its lack of foresight, condemns nature to errors and to tinkering. At this point, it is important to realize that a creator deity can create each species independently of others.

However, he could as well design a process of development, where one species leads to another one, instead of creating all the species independently at one moment of time. And indeed such a deistic conception of creation was used against evolutionary ideas, after evidence of massive change in the biological realm could not be ignored anymore during the late 18th and early 19th century (Bowler 2003). Independent creation is thus not a necessary part of a creationist worldview. And gradualism is compatible with the idea of creation through a benevolent creator. It is compatible with his perfection, external purposiveness, knowledge and foresight, as long as the outcome is the same perfect system of adaptive species. The only thing that happens to adaptation and the great chain of being, if gradualism becomes part of the idea of creation, is that adaptation and the order of species are temporalized. Deistic creation (or evolution, if you prefer the term) assumes a single creative act: an ingenious, foresightful fixing of the initial conditions and the natural laws, so that each consequent event in the biotic and abiotic world was pre-designed by the creator. The rise of new species would have been the mere unfolding of the plan of the creator, who created a perfect *creation machine*, namely nature. Temporalized perfection is progress through an unfolding process: evolution by creation. Lamarck, to whom I will turn to next, also assumed such a kind of automatic unfolding progress, but he omitted the external creator. For the idea of creation it is important to realize the following: Gradualism is compatible with the idea of creation. But a concept of change that explains change without external teleology, without foresight, and without perfection or progress is in direct opposition to it. Deistic evolutionism still includes these three parameters.

To summarize, the idea of creation implies that a deity creates a system that is perfect and either static or progressively developing, according to a pre-determined plan. The deity does this through a conscious design process, with a definite purpose and with a perfect, foresightful plan for the realization of his plans in mind. The resulting internal teleology of organisms rests on an external teleology. Everything is decided beforehand and excludes errors. Each species can be created independently of others.

### Transformational evolution

Lamarck's theory of evolution, which was the main 19th century alternative to a Darwinian theory of evolution, still includes progress. Lamarck omits external teleology, but keeps progress – perfection in its gradualist guise – by postulating a pattern of change that has been called *transformational*. One of the reasons is that he was still deeply influenced by a kind of essentialism.

In Lamarck's theory of evolution, laid down in his *Philosophie zoologique* (1809), the cause of change can be split into two factors: (i) an *internal drive towards complexity*, securing progress in terms of increasing complexity, and (ii) a *mechanism of adaptation* to local environmental circumstances through use and misuse of organs. The first factor determines the broad phylogenetic 'destination' of evolution, leading to a temporalized version of the traditional *scalae naturae,* the great chain of being, with man as the most developed and most complex of the various biological species. Lamarck explains this drive towards complexity with a normal physical mechanism that automatically causes an increase in complexity. The second factor is held responsible for the deviations from this destination towards complexity. Today Lamarck is mainly remembered for this second factor, since it is connected with the inheritance of acquired characteristics.[9] It is a mechanism that leads to a "strange irregularity in the increasing complexity of animal organisation" (Lamarck 1984 [1809]: 107). What Lamarck called irregularities would today be called adaptations to the local environment. Lamarck formulated the mechanism of this second factor in the form of two laws. The first one secures that changes are acquired, the second one that they are inherited:

> "First Law. In every animal which has not passed the limit of its development, a more frequent and continuous use of any organ gradually strengthens, develops and enlarges that organ, and gives it a power proportional to the length of time it has been so used; while the permanent disuse of any organ imperceptibly weakens and deteriorates it, and progressively diminishes its functional capacity, until it finally disappears."
>
> "Second Law. All the acquisitions or losses *wrought by nature on individuals*, through the influence of the environment in which their race has long been placed, and hence through the influence of the predominant use or permanent disuse of any organ; all these are preserved by reproduction to the new individuals which arise, provided that the acquired modifications are common to both sexes, or at least to the individuals which produce the young." (Lamarck 1984 [1809]: 113; Emph. added)

To use an organ is an activity of an individual. Consequently the organism is not totally passive in this process of adaptation. How can Lamarck then say, that the gains and losses are "wrought by nature on individuals" without contradicting himself? He saw the point and wrote:

> „I must now explain what I mean by this statement: *the environment affects the shape and organisation of animals*, that is to say that when the environment becomes very different, it produces in course of time corresponding modifications in the shape and organisation of animals. It is true if this statement were to be taken literally, I should be convicted of an error; for, whatever the environment may do, it does not work any direct modification whatever in the shape and organisation of animals. But great alterations in the environment of animals lead to great alterations in their needs, and these alterations in their needs necessarily lead to others in their activities. Now if the new needs become permanent, the animals then adopt new habits, which last as long as the needs that evoked them." (Lamarck *ibid.*: 107).

---

[9] I am taking Lamarckian evolution as a historical suggestion of how to understand change. Lamarck's theory has proved to be wrong. One of the reasons is that the inheritance of acquired characteristics seems to be impossible. This was first experimentally tested and claimed by Weismann (1892) and still counts as the "central dogma" of Neo-Darwinism (Crick 1958: 153). A second reason is that Lamarck's two factors have been replaced by undirected 'mutation' as the sole true factor that introduces novelty into evolution.

That Lamarck talks about needs has led to a debate about which roles 'striving' and will play in the design of Lamarck's theory. The problem begins historically with the mistranslation of French *besoin* (needs) as 'wants' or 'desires' (Cannon 1957) and goes on in current discussions.[10] However, this question is irrelevant here, since there is no important difference between Lamarck's ideas about needs, use and disuse, and other theories of so-called direct induction by the environment, to which Lamarck is implicitly referring to in the quotation above. According to Lamarck, even if the environment cannot have a *direct* effect on the organism without any intermediary reactions of the organisms as a medium, the changes of the organisms are *directed by* the environment. They are 'wrought by nature on individuals.' Nature causes changes in needs, and these changes '*necessarily* lead to [changes] in their activities,' which lead to new habits and heritable changes in the shape and organization of the individuals. Therefore, according to Lamarck, all *adaptive* organic changes are *induced* or '*instructed*,' as others have called it, by the environment in a *directing* way.[11] The response of the organism is determined by its environment. Through this mechanism, it is guaranteed that organisms adapt to their environment. Given a certain environment, the modifications or new characteristics that arise are necessarily adaptive.

Together, the two factors – (i), the internal drive towards complexity, and (ii), the mechanism of adaptation to local environmental circumstances – impose a definite direction on evolution by *always* causing changes in each individual in one direction: increase in complexity and adaptedness. Change is brought about through *directed generation of novelty*. Because of the directed generation of novelty, increase in complexity and adaptivity, the two outcomes

---

[10] Considering plants, Lamarck says: "There are no activities and consequently no habits, properly so-called, great changes of environment none the less lead to great differences in the development of their parts…" (Lamarck 1984 [1809]: 108). In his well-known study of Lamarck's theory, Burkhardt (1995: 175) stresses that only for the most perfect animals did Lamarck postulate initiative or will in the narrow sense. See also Mayr (1976 [1972]: 225) and Gillispie (1956: 332).

[11] Jerne (1967), Medawar (1982 [1977]: 173), Dawkins (1982a: 167ff, 1983: 410ff), Avital & Jablonka (2000: 19-21), to name but a few.

of evolution that Lamarck postulated, are guaranteed to arise. They are law-like outcomes of evolution. Through this, evolution is secured to be progressive.

Lamarck also assumed that the two factors act on *all* individuals of a population in the same directing way. Therefore not only some individuals get more complex and adapt automatically to the respective environment, but all of them. This is important since it leads to a totally different pattern of change: Lamarckian evolution is *transformational evolution*. In transformational evolution, variation and selection are superfluous. They are not necessary for evolution to occur. Let me explain this step by step.

As Mayr has put it, between Lamarckian and Darwinian evolution, there is a

> „radical difference in the mechanisms by which the environment is effective. Lamarck's conceptualization provided him with no opportunity to utilize natural selection. In spite of his nominalistic emphasis on the existence only of individuals, not of species or genera, Lamarck unconsciously treated these individuals as identical, hence typologically, just as an essentialist would. All of his statements on the impact of the environment are phrased in typological language: […] If a given environment induces very specific needs, Lamarck postulates that different organisms entering this environment will respond with the same activities and efforts and thus acquire similar structures and adaptations." (Mayr 1976 [1972]: 241)

Lamarck thought that all individuals of a given species change in a given environment in the same way. In parallel to the idea of creation, even if individuals are never identical, individual differences are treated as a mere noise that does not influence evolution. Therefore, in Lamarckian evolution there is – at a certain point in time, in a given environment – no *evolutionarily significant variation* between individuals of one species. All individuals of one *type* are treated as essentially the same. Through this 'typological' view, as Mayr calls it, it is not only guaranteed that complexity increases and adaptive changes arise somewhere in a population. It is also guaranteed that *all* of the individuals will acquire the same changes.

Since the two factors, which direct the emergence of novelty, act on every individual in the same way, the two factors change the species *as a whole*, which is thereby *transformed*. Such a pattern of change is what Lewontin called "transformational evolution". For transformational evolution, it is the case that, if a system has changed in time, this happened "because each element in the system underwent an individual transformation during its life history" (Lewontin 1985 [1983]: 85), as shown in *fig_1. Lamarckian transformational evolution.*

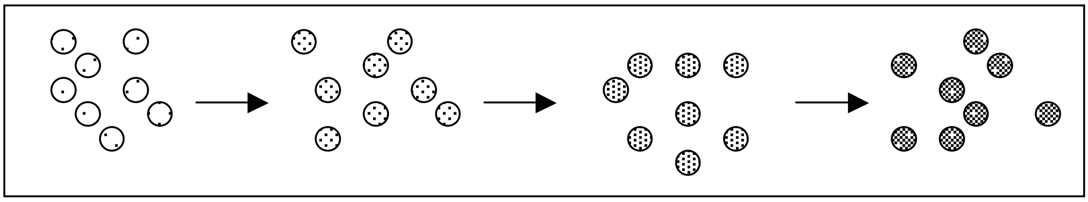

*Fig_1. Lamarckian transformational evolution: The population changes because each individual changes in the direction of evolution*[12]

From this it follows that variation and selection are not necessary for evolution to occur: First, variation is evolutionarily unimportant. Because of the transformational pattern, which rests on the typological view of individuals, evolutionarily important intra-specific variation only exists diachronically. There is no evolutionarily intra-specific variation in space at a given point in time. Second, there is no need for selection for evolution to occur. Since there is no evolutionarily variation, i.e., since all individuals are guaranteed to change in adaptive ways, there is also no *need* for selection of individuals in such a model of directed generation of novelty. Every new feature is brought about by a combination of Lamarck's two factors (i.e., the directional unfolding of the internal trend towards complexity and the induction by the environment). Since the generation of novelty was supposed to be directed and essentialist, Lamarck's theory shares an important aspect with the idea of creation: There is no waste, no selection of the unfit. Selection is unnecessary because the novelties that arise are guaranteed to arise in the direction of evolution anyway.

The idea of creation guaranteed directed generation of novelty through the deity's providence. Although Lamarck did not assume any wise and intentional creator, variation is not considered as significant, errors that are weeded out are no important factor, and progress is still secured in the

---

[12] Figure is loosely based on Medawar (1981 [1953]: 64f).

transformational pattern. As Ruse (2003: 54) has stressed, the idea of progress, which first came up with respect to social life and was mirrored then in progressivistic evolutionary ideas, was a kind of *Ersatz* for the idea of divine creation. It still made sure that life and effort made sense as part of an automatic development of the human species towards perfection.[13]

To summarize, through directed generation of novelty and the typological view of individuals, Lamarck could develop a model of evolution that is gradualist, that does not refer to an external teleology – neither for the great chain of being, nor for adaptation. There is no foresight involved. There is no personal creator that designs organisms so that they are adapted and arranged along the great chain of being. There is only an internal drive towards complexity, and the lawful instructive influence of the environment on organisms. The environment has no foresight, only causal power. This distinguishes Lamarckian evolution from the idea of divine creation. But the transformational pattern is still progressive, since it relies on directed generation of novelty and a typological view of individuals. As a consequence, variation and selection are unnecessary in such a transformational pattern of change. Denial of external teleology is only half-way to a Darwinian kind of change, which *can* – as a matter of fact – lead to progress, but does not have progress built in it – as a guaranteed result of the occurring change – since it relies on variation and selection, which are 'blind' in ways that will be illustrated below.

### Variational evolution

The core of Darwinian evolution, however, consists of the three principles Selection is an essential part of the process: variation, fitness differences, and heritability lead to selection. Variation means that individuals of a species differ from one another and that the differences can add up in the course of

---

[13] Even after Darwin had developed his evolutionary model, in which selection was a necessary part to explain adaptation, Victorian style so-called 'social Darwinism', which was in fact inspired by Spencer's Lamarckian evolutionism, rejected selection. According to Bowler (1988: 13f), this was because of the demotivating implications of natural selection. See, for instance, Peel (1971: 147ff), Bowler (1988: 38-40), Richards (1987: 291-294), or Rindos (1985: 66-68) on the common misrepresentation of Spencer as Darwinian.

evolution. This means that individual differences are an essential part of a Darwinian explanation of evolutionary change. Variation has evolutionary significance. Because of variation there also has to be selection, in order to weed out the less fit variants. This is the main difference between a Darwinian pattern and a Lamarckian transformational pattern of evolutionary change.

Darwin can thus be said to be the real ‘inventor’ of the individual. He made synchronic intra-specific variation a necessary building block of his evolutionary model. Thereby he laid the ground for what Mayr called ‘population thinking’ (Mayr 1959: 2-4). For the first time in evolutionary thinking, he thereby overcame the old philosophical essentialist concept of the relation between individuals and a species. Individuals are unique. They vary. Their differences are not only real in the sense that there is no real essence of a species; individual differences are even necessary for evolution to occur. Since Darwin was a gradualist, individuals were not ‘hopeful monsters’ for him. They nonetheless were the hopeful reformers – because of their individuality. In Lamarck’s theory, however, individuals were hopeful reformers because of their *lack* in individuality.

On the basis of variation, the change of the whole system comes about through a sorting process. This is the basic pattern of what Lewontin called *variational evolution*: The change in the system comes about through a sorting process that changes the statistical distribution of different types of variants of individuals. The whole “system evolves by changes in the proportions of the different types” (Lewontin 1985 [1983]: 86), as depicted in *Fig_2. Darwinian variational evolution*.

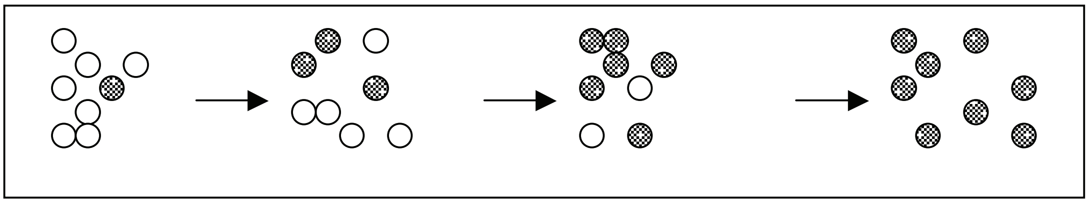

*fig_2. Darwinian variational evolution: The population changes because the statistical distribution changes.*[14]

---

[14] Figure is based on Medawar (1981 [1953]: 64f).

Hence, for variational evolution, a sorting process – be it natural selection, drift, or artificial selection – is *necessary* for evolution.[15] Furthermore, it is the sorting process that imposes the direction of evolution – *ex post facto*, after the occurrence of new characteristics.

### Conclusion

Darwin did not only introduce a naturalist and gradualist account of adaptation. This is also part of Lamarckian evolution and is not impossible for a creationist account. Darwin went further. He introduced variational evolution, a totally new model of how change can take place, making variation and selection a necessary part of any evolutionary change. The model rests on certain implicit assumptions about how the essential variation comes about, and it leads to some changes in thinking that have not been addressed so far. The assumptions and consequences are all connected with what I above called 'blindness' of Darwinian evolution. Only through these further aspects, do we understand why Darwin was able to omit all the principles contained in the idea of creation: not only external teleology and foresight, but also automatic progress through directed generation of novelty, which were still part of Lamarck's transformational pattern of change.

## 2.3 DARWINIAN EVOLUTION AS 'BLIND'

### Blind variation

As I have just illustrated already, it is common to say that Darwinism rests on 'blind,' 'random,' or 'undirected variation,' as opposed to Lamarckism, which rests on directed generation of novelty. It is yet often unclear what 'blindness' as undirectedness, the opposite of directedness, means. I will distinguish between two slightly different meanings of undirectedness: (i) undirected variation as denial of a Lamarckian instructive mechanism of adaptation,

---

[15] Drift is a sorting process that is neutral with respect to fitness differences and the underlying adaptedness of the organisms. It is survival of the 'luckiest,' not of the 'fittest,' as Kimura (1983) puts it. Drift is nonetheless considered as compatible with Darwinism, since even neutral evolution can be understood in Darwinian variational terms. For more on 'drift' see section 2.4.

through which the environment causes adaptive changes in organisms; (ii) undirected variation as the absence of a statistical correlation, called ‘coupling’, between the factors that produce novelty and the factors that select novelty in biological evolution.

*(i). Undirected variation as denial of a Lamarckian instructive mechanism of adaptation*. Often, the claim that variation in biological evolution is undirected is thought to exclude the above-mentioned *Lamarckian mechanism* of environmentally induced adaptedness. From a philosophical point of view, one could also say that in a pattern that excludes a Lamarckian mechanism, *producing* as well as *selecting* can be instances of ordinary causation, but in any case they are part of different contexts of causation. In directed generation of novelty the individual variant arises *because* of its adaptive value. If such a mechanism is absent, the cause of the generation of new properties is a non-adaptive process, while the cause of preservation lies in the interaction with the environment, leading to the selection of the adaptive ones. New properties do then not occur *because of* an adaptive value, even if these properties *have* an adaptive value. In the end, the variants that arise and are adaptive will survive. They are adaptations and maybe more complex (as *products*), if they are selected at all, but they are not produced in the first place *by* adaptation as a *process*. That is why variational evolution has been described as a discontinuous “two-step process” (Mayr 1976 [1962]: 33), or as consisting of “two separate processes, rather than a single force” (Gould 1979: 24). Evolution can thus be interpreted as relying on coincidences: a co-occurrence of events having independent causal histories.[16] William James described this coincidental structure in the following way:

> “There are, in short, *different cycles of operation* in nature; different departments, so to speak, relatively independent of one another, […] The causes which operate in these incommensurable cycles are connected with one another only if we take the whole universe into account. For all lesser points of view it is lawful – nay, more, it is for human wisdom necessary – to regard them as disconnected and irrelevant to one another. […] If we look at an animal or a human being, distinguished from the rest of his kind by the possession of some extraordinary peculiarity, good or bad, we shall be able to

---

[16] See also Dobzhansky (1974: 313-318) on randomness of mutation as coincidental, leading to unpredictability of evolution.

> discriminate between the causes which originally *produced* the peculiarity in him and the causes that *maintain* it after it is produced." (James 1979 [1880]: 166f).

Neo-Darwinism totally excludes a Lamarckian mechanism of adaptation and rests its paradigm exclusively on mutations as the ultimate source of novelty. Mutations are undirected in the just mentioned sense. But the total exclusion of Lamarckian mechanisms is not a necessary part of a Darwinian system. Directed generation of novelty, a Lamarckian mechanism of generating adaptive novelty, is compatible with a Darwinian variational pattern of change. There is no principled reason why it should be impossible for a variational model to allow for a directing influence of the environment in the generation of novelty, as long as the directing force is not the *only* source of novelty, or as long as it does not act systematically on each individual in the same way.[17] What is distinctive for Lamarck's theory of evolution is not Lamarck's second factor as such, the mechanism of directed adaptation, but that this mechanism is thought to act on each individual in the same way, leading to a transformational pattern. Lamarck's assumed pattern of change is transformational, because it *exclusively* relies on directed generation of novelty. Consider, for example, that a Lamarckian mechanism of directed generation of novelty only acts on one or a few individuals. The new characteristics will be adaptive. They will be in the direction of adaptation. If these changes can be inherited, these features will spread in the respective population, despite the presence of other individuals that do not exhibit these new adaptive features. All three central Darwinian principles are fulfilled: There is variation, there are fitness differences exhibited in this variation, and the differences can be inherited. As long as the Lamarckian mechanism does not make selection superfluous, a Lamarckian mechanism of adaptation is compatible with Darwinism. In this sense, a Darwinian can be a Lamarckian without contradicting himself. Darwin himself is the best example for illustrating this compatibility. Darwin's Darwinism did in part rely on directed generation of novelty. Although the importance of directed generation of

---

[17] Thanks to Richard Lewontin (personal communication), who drew my attention to this point.

novelty shifted back and forth over the years, he never gave up his variational theory of evolution. At the end of the 1876 edition of the *Origins,* Darwin summarizes his position:

> "I have now recapitulated the facts and considerations which have thoroughly convinced me that species have been modified, during a long course of descent. This has been effected chiefly through the Natural Selection of numerous successive, slight, favourable variations; aided in an important manner by the inherited effects of the use and disuse of parts; and in an unimportant manner, that is in relation to adaptive structures, whether past or present, by the direct action of external conditions, and by variations which seem to us in our ignorance to arise spontaneously" (Darwin 1876: 420).

Darwin considered natural selection, the cause of preservation, to be "aided" by three causes of variation: the "effects of the use and disuse of parts," the "direct action of external conditions," and the "spontaneous" generation of new characteristics. According to Darwin, all these novelties can be inherited through the mechanism of "pangenesis" (Darwin 1868). Pangenesis allowed the inheritance of all changes that occur during the lifetime of an individual, irrespective where these changes occur, be it in the somatic tissue or in the germ cells. Darwin thought that changes in the somatic tissue, especially the ones that result from the use and disuse of parts, are directed, i.e., necessarily adaptive. But as Winther (2001: 429ff) has shown, some changes in the somatic tissue and most of the changes in the germ cells were thought to be not necessarily adaptive, i.e., undirected. Even with respect to individuals of the same species, the 'changed conditions of life' sometimes "caused one individual to vary in one way and another individual in another way" (Darwin 1875, vol. 2: 260). This cause of variation was designated by Darwin as 'indefinite' or 'spontaneous.' Today it would be called 'undirected.'

To recapitulate the issue about compatibility: As long as there is some indefinite generation of novelty, selection is necessary. What Darwin called 'spontaneous' origin of novelty excludes that a Lamarckian mechanism of adaptation operates in these individuals. Darwinian variational evolution does, however, not require the *exclusion* of a Lamarckian mechanism of generating novelty for all individuals. It only requires the *inclusion* of some cases of undirected generation of novelty in the sense just specified, since that is necessary for variation to occur.

*(ii) Undirected variation as the absence of a statistical correlation.* The Neo-Darwinian paradigm, however, excludes any Lamarckian mechanism of adaptation. As already mentioned, it has mutation, which is held to arise in an undirected way, as the only ultimate source of novelty. In order to give 'undirectedness' a precise meaning that could be tested in experiments, the concept of undirectedness was formulated in statistical terms. This statistical concept of undirectedness excludes *any* influence of the selective environment on the *probability* of the occurrence of mutations. On the basis of such a statistical concept of undirectedness, Lamarck's mechanism would amount to a probability of 1 that mutations are adaptive. But the statistical concept does not only exclude a probability of 1, it excludes *any* statistical bias of mutations towards adaptivity. It is thus a much stronger requirement. If the selective environment has absolutely no influence on the occurrence of adaptive features, not even a statistical one, the occurring variation is statistically not even biased towards adaptivity. The Neo-Darwinian statistical concept of undirectedness thus refers to the absence of such a bias. If such a bias is absent, variation is said to occur at *random* and to be in this sense undirected.

In principle, mutations are said to be random in two senses:

> "First, although we may be able to predict the probability that a certain mutation will occur, we cannot predict which of a large number of gene copies will undergo the mutation. The spontaneous process of mutation is stochastic rather than deterministic. Second, and more importantly, mutation is random in the sense that *the chance that a particular mutation will occur is not influenced by whether or not the organism is in an environment in which that mutation would be advantageous*" (Futuyma 1998: 282).

The second kind of randomness is the important one for our purposes here. Mutation (and also recombination) is not influenced by the adaptive challenges the environment provides for the organism. The kind of randomness and independence involved here is what Margaret Boden calls '*relative randomness*':

> "'Relative' randomness (R-randomness) is the lack of any order or structure *relevant to some specific consideration*. Poker-dice, for example, fall and tumble R-randomly *with respect to both the knowledge and the wishes of the poker-players* – as you may know only too well. They also fall R-randomly with respect to the pattern on the wallpaper, but nobody would bother to say so" (Boden 2004: 239; Emph. in the orig.).

Something is random relative to a certain perspective or state of affairs. Biological mutations are r-random: *relative to the environment to which they must adapt and relative to the functional needs of the organism itself.* Mutations are therefore not influenced by the fitness that arises for the organism from the respective mutation. In formal terms, this comes down to the following: Let $u$ be the probability of mutating from A to a and $v$ be the probability of mutating from a to A. Mutation is undirected if it is *not* the case that (i) $u>v$ *and* (ii) $u>v$ because $w(a) > w(A)$, where $w(X)$ is the fitness of X.[18]

As Stephen Toulmin said, if new characteristics arise in an undirected way, evolution is "decoupled":

> "... we should notice […] that the twin sub-processes of variation and selection can be related in either of two quite different ways. They may take place quite independently, so that the factors responsible for the selective perpetuation of variants are entirely unrelated to those responsible for the original generation of those same variants. Or, alternatively, they may involve related sets of factors, so that the novel variants entering the relevant pool are already pre-selected for characteristics bearing directly on the requirements for selective perpetuation. To mark this difference, we may say that in the latter case variation and selection are 'coupled'; in the former case "decoupled'" (Toulmin 1972: 337f).

That is, variation is undirected, random, or decoupled, only if the producing factors are not correlated with the factors that determine the selection of the variant. If, instead, variation and selection are correlated, the two processes are "coupled." Coupling leads to *directed variation*: Adaptive new variants are more likely to occur *because* of the influence of the selective environment. In such a case, the chance that a particular mutation occurs *is* influenced by whether or not it would be adaptive in the respective environment.

It is important not to misunderstand this statistical concept of undirected variation. It means that there is no statistical correlation between the selective and the productive factors. This does *not* mean that adaptive and non-adaptive features are *equiprobable*.[19] Although mutation is stochastic, there is no equiprobability of specific mutations: It is not the case that every theoretically

---

[18] This definition is based on Sober's definition of undirectedness in Sober (1992: 39, footnote 30).

[19] See Amundson (1989) on equiprobability as a mere ideal condition for Darwinian evolution that is not fulfilled by nature.

possible mutation has the same probability. Recall the formal definition from above: Let u be the probability of mutating from A to a, and v be the probability of mutating from a to A. Mutation is undirected if it is *not* the case that (i) u>v *and* (ii) u>v because w(a) > w(A), where w(X) is the fitness of X. This definition simply says that if there is a higher likelihood of u, it must be excluded that this is *because* of a higher fitness of the mutation from A to a. It does *not* follow that we have directed variation if u>v. Therefore, it does not follow that undirected variation requires that specific adaptive trials are no more likely to occur than specific maladaptive ones. Undirected generation of novelty only entails that the *factor that determines selection does not increase the likelihood of an advantageous mutation*. That is why the following statement of Donald T. Campbell must be judged as misleading. Campbell wrote that blindness of variation means:

> "[T]hat the occurrence of trials individually be uncorrelated with the solution, in that specific correct trials are no more likely to occur at any one point in a series of trials than another, nor than specific incorrect trials" (Campbell 1987 [1960]: 93).

The statement is misleading, because it suggests that in decoupled evolution specific adaptive trials are as likely as specific maladaptive ones. But, as shown, in principle there is no kind of equiprobability implied in decoupled evolution. In fact, there is no such equiprobability of mutations. First of all, different loci or regions of loci in a genome have different mutation rates. Second, environmental factors, for instance ultraviolet and other radiation, chemical mutagens, or nutrition can increase mutation rate (Futuyma 1998: 282). Third, as Dawkins (1986: 306) mentions, there is the phenomenon of mutation pressure, where the forward mutation rate at a locus is different from the backward mutation rate. His example of such a mutation pressure in hemoglobin molecules can illustrate the phenomenon: "Form 1 and Form 2, are selectively neutral in the sense that both are equally good at carrying oxygen in the blood, [but] it could still be that mutations from 1 to 2 are commoner than reverse mutations from 2 to 1. In this case, mutation pressure will tend to make Form 2 commoner than Form 1" (Dawkins 1986: 307). Last but not least, developmental constraints limit the *range of variation*, i.e., the novelties that can occur, given the effects of previous selection: At a certain point of time in evolution, certain variants are not possible or at least very unlikely, since evolution is cumulative. Even if a mutation occurred, it would not lead to a viable variant since the respective developmental pathways are missing. As Futuyma writes,

> “Mutations with phenotypic effects alter developmental processes, but they cannot alter developmental foundations that do not exist. We may conceive of winged horses and angels, but no mutant horses or humans will ever sprout wings from their shoulders, for the developmental foundations are lacking. […] Thus, some morphologies are highly unlikely, or even impossible, for reasons that we usually do not understand because of our ignorance of developmental pathways. For instance, the numerous ankle bones of a salamander can be organized (by fusions and fissions) in many imaginable ways, but some conceivable patterns have never been found, either as intra-specific variants or as species-typical characters” (Futuyma 1998: 276).

Developmental constraints are compatible with Darwinism. According to Dawkins (1986: 307-312), it is a caricature of Darwinism, if Darwinism is described as stating that all changes are equally possible and equally likely, i.e. that at each point in time, every conceivable variation can occur for selection to operate on. Some variations are never available for selection. As Dawkins writes in his characteristic prose, "[g]enes can mutate till they are blue in the face, but no mammal will ever sprout wings like an angel unless mammalian embryological processes are susceptible to this kind of change" (Dawkins 1986: 311). The species’ past, its history, can act as a constraint on variation. History thus *restricts* the range of variation, i.e. the pathways that are open or accessible from the respective starting-point. This limit on the range of variation, which also leads to a statistical bias in variation, is compatible with Darwinism and has to be distinguished from undirected variation. Developmental constraints are an influence of the effects of past selective environments. Directed variation is the effect of a current selective environment. Darwinism is compatible with the former, but it is – in its Neo-Darwinian version – not compatible with directed variation in the sense of a statistical coupling between factors that cause novelty and factors that select novelty. The difference between developmental constraints and directed variation is very important and will reoccur in chapter 4 on the origination analogy, where I will analyze whether creativity is based on undirected variation or not.

In addition to the just mentioned possible misunderstandings about equiprobability, it is very important to clearly see the difference between (i), undirectedness as the absence of a Lamarckian mechanism that guarantees adaptive features, and (ii), the absence of a statistical bias in the occurrence of adaptive and maladaptive mutations. Directed variation, i.e., the existence of coupling, is often called 'Lamarckian.' But, as indicated already, Lamarck's concept of directedness was stronger. With respect to adaptive evolution, Lamarck assumed a factor that *necessarily* produces adaptive features in organisms, i.e., that induces adaptive features. In terms of the concept of decoupling, leading to undirected variation, and coupling, leading to directed variation, Lamarck's concept of directedness would mean that the coupling is 100 per cent. It means that the selective factors *are* the productive factors. In terms of the formal definition of undirectedness offered above, Lamarck's concept of directedness contains that $u = 1$, and $u = 1$ because of $w(a) > w(A)$. While Lamarck's concept of directedness amounts to 100-per-cent-coupling, the Neo-Darwinian concept of decoupling is stronger than the mere exclusion of a Lamarckian mechanism of adaptation. It excludes *any* influence of the environment on the occurrence of new characteristics. That is why (i) can be subsumed under (ii) in the following way:

> *Undirected variation* excludes coupling, a 'Lamarckian' correlation between the factors bringing about new characteristics and the factors selecting these new characteristics. It excludes a partial correlation as well as a total correlation, as in a Lamarckian instructive mechanism that guarantees that adaptive novelty arises.

This inclusive definition of the concept of undirected variation will be used to analyze the origination analogy.

The question whether a system is based on undirected variation must, however, be distinguished from the question whether a system is a variational one. First, a system is only transformational if variation is prevented by a mechanism that guarantees a unity of response given a certain environment, e.g. through a Lamarckian mechanism of adaptation. Second, a system can be variational, even if variation and selection are coupled (i.e., the origination of novelty directed), as long as the coupling is not 100 per cent. Therefore, if one wants to ask whether a system is Darwinian the question whether a system is variational or not and the question whether it exhibits decoupling or not have to be distinguished and addressed separately. This is what I will do in chapter 4. I will indicate that culture can be understood as a variational system. However, the important question with respect to the origination of variants is whether novelty arises blindly in the sense assumed by Darwinian theory, i.e. in an *undirected* way.

### Selection as blind, natural force

The non-adaptive novelties that occur in a variational change must be negatively selected, so that adaptations can gradually evolve. Non-adaptive features are 'trials' that turn out to be 'errors.' Undirectedness of variation thus means excess, wastefulness, and death. Yet, wastefulness is only the blindness of the productive side of the overall evolutionary process. As illustrated in the last two sections, blindness of variation leads to selection after the occurrence of variation as a necessary part of the evolutionary process. This shows already that selection is also 'blind' in a certain way. It cannot be otherwise, since Darwinian evolution is a natural process. 'Natural' in this case means that 'nature' is not the sort of thing like a deity or another person-like entity with a mind, plans and foresight. The blindness of Darwinian evolution in non-metaphorical terms thus means the absence of a mind with plans, values, knowledge and foresight. Undirected variation is one consequence of this. However, it has other consequences as well. The absence of plan and foresight implies that nature, although discriminating between adaptive and non-adaptive features of organisms, has not only no foresight in producing variants, it has also no foresight in selecting the variants. If we describe nature in a metaphorical way, i.e., as having a mind, its discriminating power could be described as a mere *shortsighted selection by hindsight.* In this sense, nature 'sees,' i.e., discriminates between adaptive and non-adaptive features, although it does not discriminate with the help of foresight and planning and is in this sense *blind*. In non-metaphorical terms, nature does not see anything. The only thing nature does is to execute its power to 'let live or die.' This has consequences that make clearer in which sense natural selection differs from a process of creation, or even artificial selection, although both can be conceived as forms of selection.

*(i). Selection is not blind in the sense of 'random'; it is a law-like cumulative process*. The first kind of blindness of selection, which I want to discuss, is one that does not exist. Creationists often mention it. Paley, for instance, rejected chance and necessity as an alternative to the creationist explanation of design. For Paley (1825 [1802]: 44ff), the idea that something comes about by chance, and is maintained or not, is ruled out as a good explanation for the origination of complex functional organs, such as an eye. It is ruled out, since it is too *unlikely* that a complex organ arises by a single coincidence. The probability that the many parts that have to fall in place in such and such a way to build a viable eye, fall in place by chance is very low. It is as low as the probability that a tornado blowing through a junkyard produces an airliner, as the standard contemporary example goes. In a nutshell, the creationist would argue that, since selection is 'random' in this sense, a deity must have created these entities. If natural selection were operating by such a haphazardous single coincidence, it could indeed hardly be used to explain the existence of complex functional organisms.

But natural selection is not random in such a way. This is an outright misunderstanding of Darwinian natural selection. The Darwinian answer thus contains two parts: (i) selection is not random, it is law-like; (ii) selection is not a *single* 'coincidence', but a cumulative process. In sum, natural selection is a fitness-driven natural process that makes the improbable possible through gradual cumulativity.

*(a). Selection is law-like.* Selection itself is in no sense random. It is the opposite of randomness. "Natural selection is the antichance factor in evolution. […] On the contrary, selection is, as a rule, directed towards maintenance or enhancement of the Darwinian fitness," as Dobzhansky (1974: 318) puts it. Selection is determined by fitness and fitness is not given by

chance, but determined by certain characteristics of the organism (adaptedness) and the corresponding environment. Hence, natural selection is *determined* by the relationship between organisms and their contingent environment. Nonetheless, natural selection is *contingent,* but not random, since environments vary and "which organisms reproduce more effectively depends on what variations they possess that are useful in the environment where the organisms live" (Ayala 2001: 236). To summarize this point in the words of Mayr: "Nothing succeeds like success […] where survival and differential reproduction are concerned, anything but blindness prevails" (Mayr 1976 [1962]: 34). Nature discriminates in a law-like way between adaptive and non-adaptive features of organisms.

*(b). Selection is not a single coincidence.* Creationists might even admit that selection is fitness-driven and in this sense law-like. But they would still insist that it is a process of chance that cannot explain adaptive complexity. The reason for this is that they regard selection as a single-step coincidence. On this basis, they claim that natural selection cannot produce things like eyes, as a tornado blowing through a junkyard cannot assemble an airliner.

They ignore the effect of cumulativity involved in Darwinian evolution. Selection is a cumulative law-like process, thereby making the improbable probable. Dawkins (1986: 43ff) showed the effect of cumulativity with the old example of the monkey, bashing away at random on a typewriter and producing finally the works of Shakespeare.[20] Dawkins discusses the issue with respect to the production of only one sentence from Shakespeare: Hamlet's saying "METHINKS IT IS LIKE A WEASEL." Creationists are right: As it is very unlikely that the tornado, blowing through a junkyard, produces an airliner, it is equally unlikely that the random typing of a monkey on a typewriter produces this sentence. The working of the tornado and the randomly typing monkey is what Dawkins calls 'simple sieving' (*ibid.*: 44) or 'single step selection of random variation' (*ibid.*: 47). He compares it to an opening of a combination lock with many dials. The typing of the sentence

---

[20] This thought example of the typewriting monkeys (originally in the base of the British Museum) is quite widespread. It was already used by Newell, Shaw & Simon (1957).

'METHINKS IT IS LIKE A WEASEL' is the opening. Each bash on the keyboard is comparable to a dial. Now imagine the monkey is a computer-monkey. Let's say that there are 26 letters and a spacebar on they keyboard, nothing else, hence 27 possibilities for each hit. The sentence has 28 characters. If we let the monkey start, there are $27^{28}$ different possible sequences. If the process is truly random (i.e., if the probability for each letter being hit by the monkey is the same), then the probability that the monkey types exactly 'METHINKS IT IS LIKE A WEASEL' is $(1/27)^{28}$ – 1 in 27 multiplied by itself 28 times. As Dawkins writes, "[t]hese are very small odds, about 1 in 10000 million million million million million million. To put it mildly, the phrase we seek would be a long time coming, to say nothing of the complete works of Shakespeare" (*ibid.*: 47).

But cumulative selection is much more effective. Dawkins tries to illustrate this through a change in the computer-monkey program. The computer-monkey

> "again begins by choosing a random sequence of 28 letters, just as before: WDLMNLT DTJBKWIRZREZLMQCO P. It now 'breeds from' this random phrase. It duplicates it repeatedly, but with a certain chance of random error – 'mutation' in the copying. The computer examines the mutant nonsense phrases, the 'progeny' of the original phrase, and chooses the one which, however slightly, most resembles the target phrase, METHINKS IT IS LIKE A WEASEL" (*ibid.*: 47).

This procedure is repeated over and over again. A test of this program showed that the computer-monkey reached the target sentence in 40 generation of sentences.[21] A second run has led to 64 generations. The chances that the target sentence will appear after 40 trials or generations of sentences are almost zero in single-step selection. But the chances that the target sentence appears on the screen in the multiple selection program within 40 trials or generations of sentences are quite high.

Natural selection is such a cumulative selection process. It is repeated sieving with *memory* that operates over many generations: "[T]he results of

---

[21] Dawkins also developed a more elaborate program that mirrors more accurately what happens in biological evolution. This program, the 'blind watchmaker' program, also integrated the complexities of development. For a description see ch. 2 of Dawkins (1986).

one sieving process are fed into a subsequent sieving, which is fed into…, and so on” (*ibid.*: 45). Dawkins concludes:

> “There is a big difference, then, between cumulative selection (in which each improvement, however slight, is used as a basis for future building), and single-step selection (in which each new ‘try’ is a fresh one). If evolutionary progress had had to rely on single-step selection, it would never have got anywhere. If, however, there was any way in which the necessary conditions for *cumulative selection* could have been set up by the blind forces of nature, strange and wonderful might have been the consequences. As a matter of fact that is exactly what happened on this planet” (*ibid.*: 49; Emph. in the orig.).

The cumulative selection monkey-program has still to be translated into *natural* cumulative selection. In our example the computer has been given the ideal target of evolution. In a more realistic simulation of natural selection, the computer would have to realize the fittest trial in an analogous way to nature. Such a computer simulation of cumulative *natural* selection would require a “very sophisticated pattern-recognition program” (*ibid.*: 61), a thing hard to come by. However, this is no problem for real world natural selection. Nature evidently has such a recognition program. Nature discriminates according to fitness. Death and differential reproduction, as a consequence of fitness differences, is all that is needed.

To conclude, cumulative variational evolution is not ‘random’ or ‘blind’ in the sense of an undetermined single step coincidence. It can account for adaptive complexity through a multiple-step process, consisting of multiple cycles of variation and selection. It therefore provides a ‘third’ alternative to ‘chance’ and ‘necessity’ as conceived as single-step selection on the one hand, and creation on the other hand. This alternative combines chance (in the sense of undirected generation of novelty) and necessity (in the sense of fitness differences leading to differential survival and reproduction), *iterated* over many generations, with memory of the intermediate results. That is how natural selection makes the improbable probable: by working in cumulative small steps with memory.

All this happens without the help of foresight in the generation of novelty (which would guarantee directed generation of novelty) and without the help of ‘mind,’ ‘purpose,’ or ‘foresight’ in the selecting part of the cumulative process. Nature does not see anything or have any purposes, when she selects between the different variants available for selection. Natural selection is blind in this sense. Differential death because of fitness differences just occurs. As Dawkins writes:

> "Natural selection, the blind, unconscious, automatic process which Darwin discovered, and which we now know is the explanation for the existence and apparently purposeful form of all life, has no purpose in mind. It has no mind and no mind's eye. It does not plan for the future. It has no vision, no foresight, no sight at all. If it can be said to play the role of watchmaker in nature, it is the *blind* watchmaker" (Dawkins 1986: 5).

The fact that nature does not have a mind to govern evolution by foresight, leads to two further aspects. Without foresight in *producing* new variants selection can only happen *ex post facto* – after the occurrence of undirected variation: If nature had a mind, its 'method' would be comparable to the hindsighted artificial selection of animal breeders. This directly follows from undirected variation. In addition, having also no foresight of the long-term consequences of undirected variation, nature selects in an absolutely *opportunistic* way: In metaphorical terms, if nature had a mind, she would be comparable to a quite myopic breeder, who has no overall plan, does not select for future pay-offs, and who tinkers around with what he has. We saw that (i) selection is not blind in the sense of a random simple sieving. Let me elaborate on blind selection in the sense of (ii), selection ex post facto, and (iii), opportunistic selection.

*(ii) Selection ex post facto.* In variational evolution, it is selection that imposes direction and it does so *ex post facto*, by 'hindsight' after certain variants, which are not directed, have come about. Selection ex post facto is a consequence of the variational pattern: Since the overall process consists of an iteration of two steps – variation and selection – the directing force can only exert its directing power by selection of consequences, in metaphorical terms by 'hindsight.' In Darwinian variational evolution,

> "[p]rogress […] is by trial and error. […] Whatever purposiveness the organic world seems to have is thus not a finalistic one but, if I may say so, an *a posteriori* one, or, in other words, the result of past natural selection" (Mayr 1976 [1962]: 42).

Lacking foresight, nature has to wait until some adaptive variants occur, and can only impose direction through weeding out the maladaptive ones.

Therefore, there is no guaranteed adaptive novelty involved. As I said, Lamarckian evolution is transformational because it *only* consists of instructive processes, directing, or 'instructing,' and hence securing not only novelty but also the course of evolution by acting on every individual. Transformational evolution and creation is *directed from the start* (either by the internal drive towards complexity and the directing influence of the environment, or by a purposeful creator).

The difference between direction through a transformational pattern or foresightful creation and the direction resulting from blind natural selection is important. Ignoring the difference is distorting the actual historical picture. If we look from the present to the past and if selection works in one direction for a certain amount of time, the surviving species can be visualized as a directional lineage. But the direction visible from our hindsight can be misleading: It only shows the winning ones and consequently distorts the actual historical paths. To ignore the 'false starts' would be similar to what modern historians call 'whig-history,' the technique of interpreting the past in terms of the winners and of presenting the change as a necessary development, ignoring the actual historical path. The term whig-history derives from the Whigs, the political ancestors of the Liberal party in Britain, who were engaged in 'whig-history,' when they rewrote the country's history in such a way that it vindicated their own values (Bowler 1988: 16). The difference between direction by foresight and direction by hindsight can therefore be depicted in the way shown in *fig_3. Darwinian direction versus Lamarckian direction.*

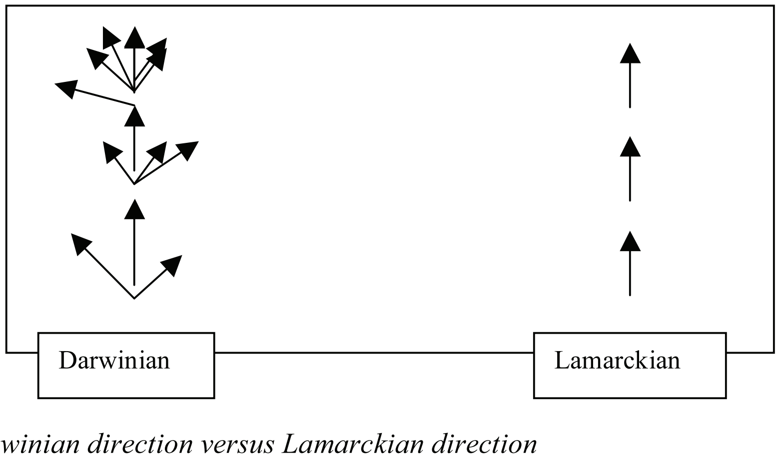

*fig_3. Darwinian direction versus Lamarckian direction*

In Darwinian evolution, direction is the effect of the overall process of Darwinian variational evolution. But to say that Darwinian evolution is progressive, that it has a purpose, a predetermined final end, would be – as Mayr clarifies – "ludicrous," since it would be similar to saying "that it is the purpose of every individual to die because this is the end of every individual, or that it is the goal of every evolutionary line to become extinct because this is what has happened to 99,9 percent of all evolutionary lines that have ever existed. Indeed, one would be forced to consider as teleological even the second law of thermodynamics" (Mayr 1976 [1974]: 388).

*(iii) The opportunistic watchmaker.* This leads to the second characteristic of natural selection that follows from the fact that nature has no mind that can direct by foresight the ongoing evolution. Even if we personalized the process of natural selection, by summarizing all the causes that make up a selective environment as 'Mother Nature', the 'watchmaker,' the selective activity of this 'natural watchmaker' would be necessarily opportunistic. The opportunism of natural selection sets restrictions on the process that do not hold for creation. Some of the restrictions do not even hold for artificial selection.

In principle undirected generation of novelty as such is compatible with creation and artificial selection, as long as we do not assume a perfect creator or selector. Yet, if blind generation of novelty is paired with natural selection, then the overall process of Darwinian evolution is different from what one can expect from the creative or selective activity of persons with certain cognitive abilities. The process is different in three ways: (a) since natural selection knows no higher goal or progress, only 'haphazardous,' *local* adaptation and trends are possible; (b) natural selection does not respect *future* positive effects; (c) natural selection is bound to *tinkering*.

*(a). 'Haphazard,' local adaptation.* Nature does not follow a higher external and overall progressive plan. The master plan that was part of the idea of creation and temporalized in transformational evolution through Lamarck's first factor is missing. Natural selection is opportunistic, or 'shortsighted' in

metaphorical terms. It always reacts to contingent, local conditions. There is no higher or future goal, no master plan, neither diachronically nor synchronically.[22] Natural selection can lead to adaptations, but excludes foresight and a goal beyond the variety of adaptations. In transformational evolution the overall plan was excluded as an external teleological cause, but included as a necessary natural transformational process. In artificial selection, the overall plan can be superimposed by selection on the variety at hand. Certainly, artificial selection can proceed whimsically as well, but it also can exhibit an overall plan. Natural selection is the only mode of producing novelty, which totally lacks such an overall plan. As I mentioned already in chapter 1, Darwin's theory was an insult for man (in Freud's sense) because of two aspects of his theory: The first one is the animal heritage; the second one is the thesis that evolution is based on a process that knows no higher goal of the overall path of evolution. Natural selection was thus interpreted as "casting men adrift in a meaningless world" (LeMahieu 1976: 71).[23]

*(b). Natural selection does not respect future positive effects*. Natural selection is opportunistic in a further sense. If a characteristic of an organism does not immediately constitute an advantage for survival, natural selection cannot favor it. Even if a change pays in future generations, in case other mutations occur that make the trait beneficial, or in case the environment and the selection pressure change, nature will ignore these future effects and will select against the characteristic. It cannot respect future positive effects. "Evolution has no foresight, and a genetic element cannot be selected because it might someday be of some help," Jacob writes (1982: 40). This is because "natural selection does not anticipate the environments of the future" (Ayala 2001: 236). A breeder and a creator can look ahead and select, breed, preserve organisms, even if they are not fitter than other ones at the moment of decision.

---

[22] See Dobzhansky (1974: mainly 326ff) or Ayala (2001: 237).

[23] Furthermore, as indicated already, even the local adaptations are not guaranteed to arise because of undirected generation of novelty. This also has a consequence for human self-understanding. Humans are not the necessary destiny of evolution. If the respective characteristics that make man peculiar had not arisen or if the local selective criteria had been different, it could well be that humans do not exist.

In fact, humans can respect future outcomes, for instance, in our personal projects as well as in political reforms. Natural selection cannot have this kind of ‘patience’ and this has an effect on the outcome. Natural selection always punishes ‘arriving at the wrong time or place.’ If a personal decision maker selects with foresight, as in creation or artificial selection, this ‘punishment’ can be prevented.[24]

*(c). Natural selection is like a tinkerer.* In explaining the meaning of undirected variation, I mentioned developmental constraints. A creator and a breeder, artificially selecting between organisms, could ignore these constraints through starting a new blueprint. A creator does not have to build on old material. He can ignore cumulativity. In order to design something new, natural selection must *necessarily* build on old material. It cannot go a step back or start totally anew “from scratch,” as Francois Jacob once said. Natural selection “works on what already exists, either transforming a system to give it a new function or combining several systems to produce a more complex one” (Jacob 1982: 34). According to Jacob, this process does not resemble engineering, a frequent metaphor used to describe natural selection. It resembles “tinkering,” the English word for “bricolage” (Jacob 1982: 33-46), a term Jacob takes from Claude Levi-Strauss (1962). Jacob describes the process, by which the tinkerer brings about novelty, in the following way: “[T]he tinkerer picks up an object which happens to be in his stock and gives it an unexpected function. Out of an old car wheel, he will make a fan; from a broken table a parasol. This process is not very different from what evolution performs when it turns a leg into a wing, or a part of a jaw into a piece of ear” (Jacob 1982: 35). Furthermore, “in contrast with the engineer, the tinkerer who wants to refine his work will often add new structures to the old ones rather than replace them” (*ibid.*: 36). The development of the brain in mammals, through which the neocortex got added to the old rhinencephalon of lower vertebrates, serves him as an example: It “strongly recalls the way the tinkerer

---

[24] See also Hull (1982: 318f) on the claim that natural selection differs from artificial selection and rational choice of humans in general because it only leads to local adaptations and does not respect future effects. Hull refers to Nagel (1977) and Elster (1979) for stressing the difference.

works. It is somewhat like adding a jet engine to an old horse cart" (*ibid.*: 37). As Gould writes, with respect to Darwin's book *on Orchids* (1862):

> "If God had designed a beautiful machine to reflect his wisdom and power, surely he would not have used a collection of parts generally fashioned for other purposes. Orchids were not made by an ideal engineer; they are jury-rigged from a limited set of available components. Thus, they must have evolved from ordinary flowers […] Odd arrangements and funny solutions are proof of evolution— paths that a sensible God would never tread but that a natural process, constrained by history, follows perforce" (Gould 1982: 20f).

As a hypothesis, natural selection leads to different predictions then creation or artificial selection, as Sober (2000: 39) stresses in his critique of creationism.[25] Natural Selection does not lead to a prediction of 'perfection' in nature. Hence it can explain why we find differences in functional organs although the differences do not contribute to functional optimality. Why do the wings of birds and bats differ? Because they are descendent from different ancestors. The differences are a consequence of tinkering. Since a tinkerer is bound to what he has in hand and cannot ignore already established features, there will be – given different evolutionary histories – different solutions for wings. From a Darwinian perspective, these differences are evidence of ancestry and can be explained by the Darwinian paradigm. From a creationist perspective, it is hard to imagine what reason the hypothesized perfect God would have had to include the properties that are unnecessary for building an organ for flight. Only a trickster God, who consciously and purposively behaves as a tinkerer, who builds in the unnecessary differences 'from scratch,' could be the basis of a creationist prediction of such unnecessary differences.

Artificial selection is a borderline case. The breeder of organisms can also not take parts of an organism and breed the parts, albeit he can select organisms for certain traits, i.e., because of certain traits. But the breeder of organisms can sometimes go back some steps in the evolution of the respective species (if the species is not yet extinct) and begin a new series of breeding from an individual that does not already exhibit the traits he wants to leave out.

---

[25] See also Jacob (1982: 33f).

Furthermore, the geneticists of our days can eliminate parts of an organism and reproduce the organism without that part.

Note that the restriction that is imposed on evolution through developmental constraints and tinkering stems not only from nature's mindlessness, but also from the fact that nature can only select wholes and never separate parts, even if it selects these whole organisms for a certain part, a point stressed by McLaughlin (2001: 153-160).

To summarize the point about nature's opportunism: Nature is bound to shifting goals (bound to the contingences of local circumstances), bound to opportunistic exploitation (it can only react to immediate pay-off), and bound to opportunistic tinkering (it cannot ignore the past). Natural selection lacks, so to speak, reflective distance. Therefore, it often does not lead to perfect solutions. What natural selection leads to is relative optimality. The solutions of nature are not the best of all possible solutions, but the best solutions of the variants that are available at a certain point in time and place.

## Conclusion

The different aspects of the blindness of variational evolution (undirected variation, selection ex post facto, opportunism) are all direct consequences of the absence of a mind with plans and foresight. Now, transformational evolution does not imply a mind either. Hence, Darwinian as well as Lamarckian evolution are both blind in that sense. What is then specific for Darwinism, that allows to distinguish it from creationism as well as from Lamarckian evolution? Let me summarize this section on Darwinian evolution as 'blind' with the answer to this question: The difference between Darwinian and Lamarckian evolution is that a Darwinian evolutionary pattern excludes progress, because of undirected variation and consequent selection ex post facto, and because of opportunistic selection – the aspects of blindness following from the absence of mind. In Lamarck's theory of evolution, there is no intelligent creator that – by foresight – imposes structure, form, order, and adaptivity. There is only a natural (and in this sense blind) process of development. Although transformational evolution eliminates a static picture of the world through temporalization of the 'great chain of being,' transformational evolution still differs in important ways from a Darwinian variational evolution. Progress is guaranteed and already immanent in the directed *generation* of novelty. Darwinism instead did not even secure this, since it is based on undirected variation and 'hindsighted' natural selection of the unfit, with opportunistic adaptation – based on local adaptation, immediate pay-off, and tinkering – as the only kind of optimality. Blind variation, selection after the occurrence of diverse variants, and opportunistic adaptation are therefore the distinguishing features of a Darwinian variational pattern of change.

## 2.4 THE TAUTOLOGY PROBLEM AND THE CONCEPT OF FITNESS

### The tautology problem

In the three-part recipe of Darwinian evolution, introduced in section 2.1, differential fitness appears as a necessary criterion for selection to occur. The higher fitness of a certain type of individuals, compared to other types of individuals, explains the selective part of the overall process of evolution, which comprises iterated cycles of variation and consequent selection. The result of this evolutionary process is a relative increase of the fitter type of individuals, as long as heritability holds. Fitness therefore is essential in explaining the change in frequency of types of individuals, and therefore in explaining the existence of, or survival of, individuals of this type in a given population. This is often summarized under the principle 'survival of the fittest,' "natural selection's *alter ego*" (Ruse 1998a: 11).

The problem is that explaining the evolution of adaptive organisms with the principle 'survival of the fittest' has often been criticized as tautological. The proposition corresponding such an explanation, based on the principle of 'survival of the fittest,' could be formulated in the following way:

> *(1)* The existing entities survived *because* they were those that were fittest
>
> or
>
> *(2)* Those who survive are those that are fittest.

If we now ask 'who are the fittest?' and answer 'those that survive,' then these propositions come out to be:

> *(1')* The existing entities survived because they were those that survive.
> or
> *(2')* Those who survive are those who survive.

These statements are tautological and are therefore empty. If we want to explain why the existing entities are here and not others, these propositions do not give an instructive answer. The charge that Darwinian explanations of this kind are tautological is similar to the charge that sentences like 'every bachelor is unmarried' are analytical and 'tautological' in the sense that the sentence can be translated into 'every unmarried man is unmarried.' As the analytical character of 'every bachelor is unmarried' depends on the definition of 'bachelor' as unmarried man, the tautological character of the principle 'survival of fitness' is the effect of defining fitness in a way that makes the propositions tautological. Hence, at the center of the debate about the tautological character of Darwinian explanations is the question what fitness means.

Karl R. Popper is famous for making the tautology charge against Darwinism. In the section *Darwinism as a Metaphysical Research Program,* which was part of his *Intellectual Biography,* he wrote:

> "To say that a species now living is adapted to its environment is, in fact, almost tautological. Indeed we use the terms 'adaptation' and 'selection' in such a way that we can say that, if the species were not adapted, it would have been eliminated by natural selection. Similarly, if a species has been eliminated it must have been ill adapted to the conditions. Adaptation or fitness is *defined* by modern evolutionists as survival value, and can be measured by actual success in survival: there is hardly any possibility of testing a theory as feeble as this" (Popper 1974a: 137).

Although Popper himself (1978: 344-346) recanted from this opinion, the debates about the tautological character have persisted since then.[26]

---

[26] See Sober (2000), Hull (1999), Paul (1992), Beatty (1992), and Ruse (1977), for discussion and review. Smart (1963), Manser (1965), and Bethell (1976) are further frequently cited for criticising evolutionary theory for being tautological. Already Campbell (1987 [1960]) admitted that the "basic insight [of Darwinian theory], so useful and so thrilling when first encountered, is close to being an analytic tautology rather than a synthetic description of process: if indeed variations occur which are differentially selected and propagated, then an evolutionary process toward better fit to any set of consistent selective criteria is inevitable"

### Relevance for analogical applications

Since the tautology charge is relevant for all explanations that refer to the 'survival of the fittest x,' it is important to look at it closely. If culture is considered to be a process of differential spread of ideas and the spread is explained through differential fitness of these ideas, as it is done by memetics, culture is not only *described* as a Darwinian process, but also *explained* in a Darwinian way. If this alleged explanation is, however, by itself 'tautological,' then it does not add anything to our understanding of culture.

In order to handle the question about culture adequately, we must first understand how the tautology problem is handled for the case of biological evolution. For this reason I will provide a short review of the debates about the concept of fitness in light of the tautology question. In chapter 5, I will then analyze whether memetics is subject to an analogous tautology problem.

### The solution for biological evolution

As I said, the core of the tautology problem lies in the meaning of 'fitness.' What is fitness and what role does it play in evolutionary explanations? Popper and also my formulation of (2') assume that fitness means survival value or actual survival. Evolutionary theorists have indeed often defined fitness as the product of *actual* survival (*viability*) and *actual* number of offspring (*fertility*).[27] But must fitness be defined this way? No, and it is indeed not wise to define it that way, irrespective of the tautology charge, as Susan K. Mills and John Beatty (1994 [1979]) have shown in an important paper in 1979.

The problem with fitness as actual survival and reproductive output is best shown by a well-known thought-experiment about two twins: There are two *identical* animals in a forest, just before their reproductive age. They have identical physical characteristics and live in the same selective environment. Suddenly, one is struck by lightning and consequently leaves no offspring. The

---

(Campbell 1987 [1960]: 109). According to Campbell, this is one of the problems Darwinian thinking has as a scientific theory.

[27] For examples see Mills & Beatty (1994 [1979]: 4f).

other twin survives and proliferates her genes happily.[28] If we take the above mentioned traditional definition of fitness as actual survival, then we have to admit that the "lucky twin is *far* fitter" (Mills & Beatty 1994 [1979]: 7), even if there is no physical difference between the two twins, except that one was unlucky and the other not. Remember, that this thought-experiment not only assumes that the twins have physically identical characteristics. It also assumes that they live in the same selective environment. One answer to the thought-experiment could thus be that actually, looking at the example closely, the two twins did *not* live in the same environment: The environment of the dead twin contained the lightning, whereas the environment of the other did not. As John Beatty (1992: 116) replies: "Yet, in an important sense the twins do share the same environment, one in which the chance of an encounter with lightning is rare." Hence, in this *selective* environment the two identical twins would exhibit different fitness, as long as fitness is defined as actual survival and reproductive output. The thought-experiment shows how counterintuitive it is to define fitness in terms of actual survival and reproductive output.

At the same time, the thought-experiments points to the endpoint of a development of the concept of fitness that started with what I would like to call the *qualitative notion of fitness*. In Darwin's time, fitness was closely connected to physical properties of individuals that make the organism 'fitter' in the sense of 'better adapted'. Darwin himself used the term fitness only once in his first publication of the *Origin*. It meant adaptedness. Only from the 5th ed. of the *Origin* (1869) onwards and in his *Variation of Plants and Animals under Domestication* (1868), he used Spencer's (1864) phrase 'survival of the fittest' as synonymous to his term 'natural selection.' Adaptedness is a property of an organism that arises through the relationship between an organism and its selective environment. Given an environment with high trees that carry the leaves, which a giraffe is able to use for nourishment, a giraffe with a long neck is better adapted to this environment then a giraffe with a short neck. Adapted organisms *fit* into the selective environment. To have 'fitness' means

---

[28] The thought-experiment was introduced by Scriven (1959).

to ‘match’ the selective environment. The property of ‘being adapted to x’ is a quality of the organism that is independent of actual reproductive success, even if being adapted is the main factor for survival and reproduction. The unlucky twin might well have the property of being well adapted, even if her life ended the way it did. Adaptedness can be attributed independent of actual survival and reproduction.

After the rise of population genetics, starting with J.B.S. Haldane, S. Wright and R. A. Fisher, fitness was defined as mere “success in producing offspring, irrespective of the causes of that success,” as Diana Paul (1992: 113) summarizes this change in meaning. The development of the concept of fitness ended with a mere *quantitative notion of fitness*: fitness as sheer reproductive output. This quantitative notion is still the conventional one. Nonetheless, it is well recognized that a proper concept of fitness must somehow lie in-between the qualitative and quantitative concept of fitness: On the one hand, fitness has something to do with physical properties. On the other hand, fitness evidently is connected to the explanation of evolutionary success, i.e., differential reproduction. Hence, fitness must be connected to *both* – adaptedness and reproductive output – and somehow lie in-between.[29]

Although quantitative fitness is sometimes still defined as actual number of offspring, philosophical debates about the concept of fitness have reached a kind of mid-level consensus that comprises four aspects: (i) propensity interpretation of fitness; (ii) supervenience of fitness; (iii) optimality models as providing an independent criterion for fitness; (iv) drift as an alternative explanation to natural selection.[30] These additions to the quantitative notion connect it with the physical properties that build the causal basis for quantitative fitness. To prevent misunderstanding and in order to stick to the nowadays conventional use, I will use from now on the term ‘fitness’ for

---

[29] See Mills & Beatty (1994 [1979]), Beatty (1992).

[30] Beyond this consensus there are many interesting debates about how to measure fitness, for instance whether fitness is to be measured with respect to one generation or with respect to more than one generation, whether we should abandon the concept because of the high context-dependency of the physical causes of fitness. A further issue is how to include the effects of fluctuating selection pressures and statistical variance in the measure of fitness. For discussion, see Beatty (1992), Sober (2001), Charlesworth (2002), Ariew & Lewontin (2004).

this quantitative notion and adaptedness for the qualitative notion of fitness in order to explain the four just mentioned aspects of the revised concept of quantitative fitness.

*(i). Propensity interpretation of fitness.* According to the propensity interpretation of fitness, fitness is not actual survival and reproductive output. Fitness is instead the *ability* for survival and reproductive output. This thesis has been put forward by Mills and Beatty, who write:

> "[W]e suggest that fitness be regarded as a complex *dispositional* property of organisms. Roughly speaking, the fitness of an organism is its *propensity* to survive and reproduce in a particularly specified environment and population. [...] When we say that an entity has a propensity (disposition, tendency, capability) to behave in a particular way, we mean that certain physical properties of the entity determine, or are causally relevant to, the particular behavior whenever the entity is subjected to appropriate 'triggering conditions'" (Mills & Beatty 1994 [1979]: 8f).

The example they use to clarify this propensity notion of fitness is the solubility of salt. Solubility is physically based on the salt's ionic crystalline structure. The triggering condition of solubility is the immersion into water. We can transfer this to fitness in the following way: Fitness is based on certain physical properties of the organism. The triggering conditions of fitness as a propensity consist of certain characteristics of the environment, including facts about the population the organism is part of (Mills & Beatty 1994 [1979]: 20).

*(ii). Supervenience of fitness.* Although fitness is based on physical properties, it is not identical with these physical properties. On the contrary, it supervenes on them. This claim was introduced by Alexander Rosenberg (1978). Sober (2000) summarizes the assumed idea of supervenience, with reference to Kim (1978) in the following way: "One set of properties P supervenes on another set of properties Q precisely when the Q properties of an object determine what its P properties are – but not conversely. If P supervenes on Q, then there is a one-to-many mapping from P to Q" (Sober 2000: 75). For the fitness of organisms this means: If our two twins are identical in their physical properties and live in physically identical environments, then they must have the same fitness. The physical properties and the relation to the environment can be summarized as an organism's Q properties. They determine its fitness. But if we have the same fitness of two organisms (the

organisms' P properties), then it does not follow that these two organisms have the same physical properties (the organisms' Q properties). Two types of organisms can have the same fitness (probability of survival and number of expected offspring), but for different reasons. If a giraffe has a propensity to have two offspring and a bacterium has the same propensity, this does not allow to conclude that there are similar causes in the mammals and the bacterium. They can exhibit different adaptations to survive in their respective selective environments, although they have the same fitness. The same quantitative fitness value can be realized totally differently in terms of adaptedness.[31]

Going back to the tautology problem, the following is important: Even if (i) and (ii) mark central steps in the explication of the fitness concept, they do not yet provide a solution of the tautology problem, as John Beatty admits:

> "[T]he propensity interpretation does *not* resolve […] the supposed problem of the circularity of the principle of natural selection. To be sure, the claim that 'the fittest are most likely to leave the most offspring' is a tautology when 'fittest' is defined in terms of actual offspring contribution. But the claim is no more empirical when 'fittest' is defined as 'best able to leave the most offspring'" (Beatty 1992: 118).

The reason is that we can only claim to have evidence for the propensity, and not only for the actual survival and reproduction, if we have evidence for the propensity that is independent of actual survival and reproduction. Such an evidence, or the falsification of it, is provided by (iii) and (iv).

*(iii). Optimality models as providing an independent evidence for fitness*. Stephen J. Gould replied to Tom Bethell's (1976) reaffirmation of the tautology charge: In order to get out of the circularity there must be an "independent criterion for fitness," so that we can "identify the fittest beforehand" (Gould 1998 [1976]: 94). Only then can we really explain evolution as a consequence of the identified fitness differences. What makes a given organism fitter? – Adaptedness. In a given environment, certain physical properties (morphological, physiological, and behavioral traits) are better adapted or superior than others. Gould writes: "These traits confer fitness by an engineer's criterion of good design, not by the empirical fact of their survival

---

[31] See Sober (2000: 74-78).

and spread [...] superior design in changed environments is an independent criterion of fitness". According to Gould, fitness gets "*expressed* as differential survival, but it is not *defined* by it" (Gould 1998 [1976]: 95).

In terms of fitness as supervening propensity, this means that only the physical basis of fitness – adaptedness – can provide a criterion for fitness or evidence for fitness that is independent of actual survival. (Note that Gould's terminology is ambivalent: On the one hand he wants to define fitness in qualitative terms, as adaptedness. On the other hand, he merely requires adaptedness as an independent criterion for quantitative fitness. I will come back to these two options in a minute, after I have said a little bit more on optimality models as providing insights on adaptedness.)

Adaptedness means that an organism exhibits traits that are an optimal solution for the task that arises for an organism given its relationship to an environment. Optimality models try to find out which design that is. To find out whether a given trait is an optimal design may be a hard task. In the case of the giraffe in an environment of high trees, it is easy to see that the long neck is optimal or at least advantageous. In other cases, adaptedness is less easy to recognize. For instance, Theodosius Dobzhansky investigated fitness differences in *Drosophila*. They were caused by chromosome inversions. He observed that the chromosome inversions led to differential replication, but the "phenotypic consequences of these inversions were difficult to identify, and so it often was quite unclear *why* one inversion was fitter than another" as Sober (2000: 70) reports. He concludes: "Traits do not always wear their adaptive significance on their sleeves" (*ibid.*: 70). In principle, there are two ways to find evidence for the fitness values, as Sober (2000: 68-70) stated: The physical basis of fitness, i.e. adaptedness, or the actual change in frequency, i.e. the consequences of fitness differences. Adaptedness is the independent evidence for fitness differences. Bernard Kettlewell (1973) used both lines of investigation when he made the famous case for a selectionist explanation of industrial melanism in the peppered moth in England. He observed that dark moths increased in frequency. But why? The selectionist hypothesis was: because they were fitter in the given environment of industrial pollution, which darkened the trees on which the moths live. To find evidence for such a differential fitness, Kettlewell looked for lower mortality rates of the dark moths. But he also looked for the reason for the lower mortality rates of dark moths and could show that the higher fitness resulted from the fact that dark moths on darker trees were less visible to predators. The reason could well have been different. So he found the adaptive feature that accounts for the differential survival of the different kinds of moths. However, as the example of Dobzhansky showed, it is sometimes not so easy to find the respective causes of certain observed fitness differences. But it is not impossible either.

So far we have encountered two slightly different solutions to the tautology problem. They are mixed together in Gould's above-quoted position, since on the one hand he asks (a) for independent evidence of quantitative fitness, and on the other hand he wants to define (b) fitness "by the engineer's criterion of good design." Both requirements point into the same direction: An evolutionary explanation, based on the principle 'survival of the fittest,' must be linked to the physical basis of differential survival and reproduction. This is done in (a) via keeping the quantitative concept of fitness in its propensity interpretation and connecting it to the qualitative fitness, the physical basis of quantitative fitness. It is done in (b) via a total move towards adaptedness, the qualitative notion of fitness itself. The two options (a) and (b) are different solutions to the tautology problem. I see it as a matter of choice which solution one takes. It does not make any difference, at least not for the tautology problem. However, since contemporary biologists use the quantitative notion of fitness, I suggest to choose (a). In short, Darwinian explanations of change are in danger of being tautological, unless we can point to the physical basis of fitness differences.

*(iv). Drift as an alternative explanation to natural selection.* Drift is an explanation of frequency change that is not connected to fitness differences. Although an alternative to natural selection, the inclusion of drift into the Darwinian paradigm provides testability of explanations that are based on the principle 'survival of the fittest.' Drift thereby complements (iii). It shows that

‘survival-of-the-fittest’-theses are empirical claims that can be tested and falsified and are therefore not tautological.

As the twin thought-experiment shows, actual differential replication of genes (higher viability and fertility of organisms) is not always evidence for fitness differences. Differential replication can be due to causes that are not correlated with differences in replication potential. Processes of random genetic drift are instances of frequency changes that are not connected to fitness differences, since they are due to “accidents of sampling caused by random variation in rates of survival or reproduction”, as Futuyma (1998: 304) writes.

The so-called *founder effect* can serve as an illustration. It is not only that an individual might be struck by lightning, as in the thought-experiment of the twins. A whole population might be struck by a chance event. Such an event might change the distribution of types so radically, so that a gene and a corresponding phenotypic effect, which had no chance of getting predominance before this random event, suddenly is predominant. Thus, if we want to know, for instance, why there are only giraffes with long necks on an island, it might as well be that this is not because they were better adapted to their environment in the past, but merely because the following happened: When this island was cut off another one through an earthquake, all the giraffes with long necks happened to be on this island and the others on the other one. The giraffes with long necks became the founders of a population with a totally new statistical distribution of certain traits. In such a case, the dominance of the trait, i.e., its changed frequency, does not arise because the individuals with the trait were in any way better adapted. The frequency changed for no reason that was connected to fitness differences. Drift can also arise through random sampling effects in mating, if population size is finite or other conditions that are the base of the Hardy-Weinberg-law are not fulfilled.[32] Random drift counts today as an alternative to natural selection. However, the importance of drift is still

---

[32] The Hardy-Weinberg-law says: if a population is infinite and randomly mating, if no genes are added from outside the population, if no mutation occurs, and if there are no fitness differences, then the frequencies of the genotypes and alleles stay constant. For a more detailed textbook description of the Hardy-Weinberg-law see Futuyma (1998: 235-239).

debated. Kimura is the most well-known defender of random genetic drift, which is the cornerstone of his *Neutral Theory of Molecular Evolution* (1983, 1989).[33]

Drift can explain changes in gene frequencies, even if there are no selective forces acting on the population. But it is beyond doubt that drift cannot explain the evolution of adaptedness, precisely because the frequency change has nothing to do with fitness, which is based on adaptedness. Consequently, if drift is responsible for a frequency change, the resulting evolution is *neutral* with respect to adaptive significance. As Kimura says, the neutral theory adds to Darwinism the "survival of the luckiest," as an alternative to 'survival of the fittest' (Kimura 1992: 230). Therefore, the inclusion of drift into the Darwinian paradigm means that not all evolution is adaptive. A frequency change is not due to natural selection and therefore not adaptive, if "the influence of fitness upon offspring contribution is disturbed by any factors which separate successful from unsuccessful reproducers without regard to physical differences between them", as Mills & Beatty (1994 [1979]: 15) put it.

It follows that fitness determines – and therefore explains – the fate of the population only if nothing interferes (e.g., a catastrophe like lightning), but not otherwise.[34] This is analogous to the fact that the solubility of salt determines the solution of salt only if nothing interferes (e.g., that the salt is coated with plastic). If nothing interferes, the event of solution of salt in water can be explained by the solubility of salt, as the change of frequency can be explained by fitness (Mills & Beatty 1994 [1979]: 9). Therefore, in the same way as the solubility of salt can be tested, whether change in frequency occurs due to natural selection or not can in principle be tested as well. As long as evolutionary theory can specify which instances of frequency changes do count as drift and not as natural selection, the claim that evolution is due to the 'survival of the fittest' is in principle testable and can be falsified, as also Popper (1978: 345-6) concluded from the inclusion of drift for evolutionary

---

[33] For a detailed description of drift see Futuyma (1998: 297-307).

[34] See also Sober (2000: 67) on this position.

theory. This was why he dropped his original claim that the principle of survival of the fittest is tautological.

To recapitulate this last point: Variation, heritability and differential fitness lead to natural selection and hence to a frequency change of genes, if nothing else interferes. But sometimes exactly this happens. Not every frequency change can thus be explained by natural selection, in the Darwinian adaptationist sense. Genetic drift is a falsification of natural selection. The principle of 'survival of the fittest' is thus not tautological. However, it is then also not universally applicable.

### Conclusion

Together, (iii), the existence of independent evidence for fitness differences (adaptedness), and (iv), drift as an alternative explanation of frequency changes, led to the consensus that the principle 'survival of the fittest' is not tautological. If we want to revise the two propositions cited at the beginning of this section,

> *(1)* The existing entities survived *because* they were those that were fittest.
> *(2)* Those who survive are those that are fittest.

then a correct revised description of the explanation of the existence of certain entities, based on the principle 'survival of the fittest,' would thus be:

> *(1'')* If no random drift processes interfered, the existing entities survived *because* they were those that had a higher propensity for survival and reproduction since they were better adapted to their respective environment.
>
> *(2'')* Those who survive are those that have a higher propensity for survival and reproduction since they are better adapted to their respective environment, if drift does not interfere.

## 2.5 THE UNITS OF SELECTION DEBATE

### The centrality of the individual and the attack from below

In Darwin's own theory of evolution individual organisms were central. As Stephen J. Gould writes, in Darwin's theory, "the 'struggle for existence' is a matter among individuals" (Gould 1982: 85). Individuals try to pursue their self-interest through survival and reproduction or, to put it less intentionally: Those types of individuals, which have a higher fitness and can realize it, will increase in frequency, if nothing else interferes. In such a case, their fitness explains the differential spread. From a systematic point of view, natural selection is then, first, a selection *of* individual organisms and, second, a selection *for* adaptive traits that serve their benefit – their benefit in terms of their adaptivity and consequently in terms of their survival and reproduction. First, selection is selection *of* individuals, since they are the entities that vary in fitness by exhibiting adaptive characteristics, that interact with the selective environment, and that are selected. Second, selection is selection *for* the adaptive traits that evolve through the differential survival and reproduction of organisms. Since these are traits that enhance the propensity for reproduction and survival of these organisms, organisms can be said to benefit from the evolutionary process. Natural selection selects the organisms because of these adaptive traits: The organisms are selected *for* these adaptive traits.

Because of selection of individuals, for traits of individuals, and for their benefit, individual organisms have been regarded as the units of selection. Natural selection is individual selection. It is their adaptivity that explains their fitness, and it is their fitness (and therefore their 'benefit') that explains their spread. "[T]he reason that selection […] is effective is that *what reproduces differentially* are individuals with traits which are differentially adapted to the environment," as Brandon & Burian (1984: xi) summarize the position that individuals are the units of selection.

From the mid 20th century this "central theorem," as Dawkins (1982a: 5) calls it, has been jeopardized from above (group selection) and below (gene selection). Researchers stated that groups can also be units of selection: There can be selection *of* groups and *for* their traits. Partly as reaction to this, others, most famously Richard Dawkins, stated that neither groups nor organisms can be units of selection. Genes are the only possible units of selection in the biological realm.[35] This is the doctrine of gene selectionism, also called the gene-centric view of evolution. As Dawkins writes, gene selectionism means "that the fundamental unit of selection, and therefore of self-interest, is not the

---

[35] See Dawkins (1976, 1978, 1982a, 1982b). The view has first been suggested by Williams (1966).

species, nor the group, nor even, strictly, the individual. It is the gene, the unit of heredity" (Dawkins 1989 [1976]: 11).

The statement refers to a general, philosophical argument for genes as the only possible units of selection. According to Dawkins, in all cases of biological evolution, genes are the units of selection. First, (1) genes *survive*, while organisms die: Only genes – as abstract entities – persist. Gene tokens make copies of themselves. They are replicators. Organisms do not make copies of themselves. Second, (2) genes are '*selfish'*: It is the 'fitness' of genes, i.e., their propensity for survival, that is favored in evolution, since they are the ultimate causal agents that determine phenotypic adaptations. Consequently, they explain the spread of these phenotypic adaptations. Ultimately, it is therefore always the fitness of the surviving gene that makes the spread of adaptations differential. In a nutshell, genes build organisms and have thus a fitness of their own, including the corresponding 'self-interest' in replication. In non-metaphorical terms – since genes do not have 'interests' in the literal sense – this means: Those genes that have a higher fitness, that lead to organisms that increase their survival, will be those that spread. This is the background of Dawkins' talk of 'selfish genes' and 'the good of genes.' Within this doctrine, organisms are regarded as mere 'vehicles' or interactors – merely interacting with the environment as a coherent whole, and making through this the spread of genes differential. Before I explain Dawkins's argument in more detail and in a more systematic way, I have to add a note about the ambiguity of the term 'selfish gene,' in order to prevent possible misunderstandings.

The general philosophical argument just mentioned has to be distinguished from the more restricted issue about 'selfish genes' in the narrow, technical sense, often called 'outlaw genes.' Selfish genes in the narrow sense are genes that can spread in a population *despite* the fact that these genes do not lead to positive effects on the fitness of the organism incorporating these genes, or that do even lead to negative effects on the fitness of the organism. These outlaw genes violate the Mendelian fair lottery, in which each allele of a genome has a fair chance of 50 per cent to make it to the next generation. 'Selfish genes' in the narrow sense circumvent this impartiality of heredity through complex and diverse mechanisms. In a metaphorical language, they can be said to ‘compete’ with the other alleles for their representation in the next generation, while heredity is normally organized as a ‘cooperative’ endeavor of all the cells and genes of an organism. Normal genes have to ‘cooperate’ and enhance the fitness of the organism. Outlaw genes work only for their own survival. Not all genes are selfish in this more narrow sense (Burt 2002: 1020). Nonetheless outlaw genes are used as an argument for the general claim of gene selectionism, which assumes that *all* genes are selfish, in the wide sense, since they are the ultimate units of selection: (1) those units that survive across generations, and (2) the ‘selfish,’ ultimate causal agents that determine phenotypic adaptations.

The philosophical debate about units of selection concentrates on the two thesis (1) and (2). I will do the same. Until today, no consensus has been reached in the debate. I will not try to solve the issue and will also put aside the group selection debate, since it is not relevant for this study on analogical reasoning from nature to culture. I will first of all show which conceptual issues are involved: What kinds of entities are central in evolutionary theory? What must a unit of selection do, i.e., what *role* does it play in the evolutionary process? This leads to the central concept of a replicator. After explaining the concept of a replicator, as it is used in Dawkins radical gene selectionism for (1) and (2), and after pointing to some critical aspects, I will briefly review the critique against (2), the causal issue about the ‘power’ and ‘self-interest’ of genes. What is at issue with respect to (2) is mainly whether one distorts the causal picture when one describes all processes that occur at the organismic level as phenotypic effect of replicators. If to do so distorts the causal picture, then genes merely ‘keep the book’ of evolutionary change.

### Relevance for analogical applications

Before I continue, I want to indicate briefly why the debate about units of selection is relevant for this investigation. The issue about replicators as the only possible unit of selection is most relevant for the ontological analogy: Memeticists state that memes and genes are both replicators. A precise understanding of the replicator concept as developed in evolutionary biology is therefore mandatory for any discussion of the transfer of the replicator concept to the cultural sphere. The issue about causation is most relevant for the explanatory units of selection analogy: Gene selectionism states that genes – and not organisms – are the ultimate beneficiaries of selection: ultimately, it is their fitness that is optimized through evolution; organisms are mere consequences of their replicative and organism-building power. Thus, they are explanatorily prior to organisms. In an analogous way, Dawkins, Dennett and others state that memes are the units of selection of culture: Memes and their differential survival and replication explain cultural evolution; the causal influence of human individuals is secondary.

### Replicators, vehicles and interactors

The standard ontology of 20th century evolutionary thinking assumes an organizational hierarchy of entities. At the most basic level, there are bits of DNA, lined up on chromosomes and identified with genes. Genes are parts of cells. Cells are lumped together to organs and these organs build organisms. Families, kinship groups, population and species, comprised of organisms, form the next level of the hierarchy. Local populations of different species form communities, which are parts of ecosystems, building the highest level of the hierarchy. From the lowest to the highest level, there is a part-whole relationship. At the center of evolutionary debates are genes, organisms and groups of organisms.

At the beginning of the units of selection debate, there was just the question whether genes *or* organisms (or groups) are the unit of selection. Now, evolutionary theorists mainly distinguish in a formal way between two different causal roles genes or organisms can play in the evolutionary process: the role of a replicator and the role of a vehicle (or interactor).[36] The concepts of ‘replicator,’ ‘vehicle,’ and ‘interactor’ were introduced by Dawkins (1976, 1982a, b) and Hull (1980). According to Hull, a replicator is “an entity that passes on its structure directly in replication.” An interactor is “an entity that

---

[36] Lloyd (1992, 2001) even distinguishes between four roles involved in the units of selection debate: the replicator, the interactor (or vehicle), the beneficiary and the manifestor of adaptations.

directly interacts as a cohesive whole with its environment in such a way that replication is differential" (Hull 1980: 318). Dawkins uses similar definitions. A replicator is the unit of heredity that survives as an identifiable unit over long time spans. Selection, which works over many generations, must be based on such long-lasting units. But replication is not enough for evolution to occur. At each generation selection directly works *on* interactors, which make the replication of replicators differential. Genes are the paradigmatic replicators, whereas organisms are the paradigmatic interactors, although genes and groups can be interactors as well. On this basis, Hull (1980: 318) and also Dawkins (1982a: 82; 1982b: 162) argued that there is not one unit of selection but two. Each has its role to play in the evolutionary game. The question whether genes or organisms are the units of selection has thus to be divided into two questions: Who is the replicator and who is the interactor?[37]

However this differentiation between two important causal roles, between two kinds of units of selection, did not end the debate. One of the reasons is that some gene selectionists, like Dawkins or Sterelny & Kitcher (1988), Waters 1991, Sterelny et al (1996) still claim that the causal power of interactors can be *reduced* to the causal power of replicators or (if formulated in more pluralistic terms) at least be *represented* at the level of replicators.[38] In other words, even if we grant that the unit of heredity is the gene, that the unit of interaction is in most cases the organism, we leave out the following question: What is the unit that accounts for the traits that selection favors? Does evolution select *for* traits of organisms or for single genes themselves? If the traits of organisms can simply be reduced to or represented by genes, then the gene is necessarily the ultimate unit of selection. Before I can turn to this issue, the concept of replication has to become specified in a precise way.

---

[37] The conceptual distinction between replicators and interactors has led to a second distinction: Brandon (1982) introduced the distinction between units of selection (replicator question) and the levels of selection (interactor question). In general, it is now taken for granted that the interactor question is an empirical question. Selection normally acts on many levels (Keller 1999). Yet, many still disagree whether genes deserve a special status as the sole replicator of evolution.

[38] Hull is less radical with respect to what counts as replicator and opposes the reduction of the causal roles of interactors. See Hull (1980, 1988b) on his position. Therefore, I do not count him as a gene selectionist.

## The replicator

In *The Selfish Gene* (1976), Dawkins introduced his concept of a replicator with the following words:

> "At some point a particularly remarkable molecule was formed by accident. We will call it the *Replicator* […] it had the extraordinary property of being able to create copies of itself" (Dawkins 1989 [1976]: 15; Emph. in the orig.).

A replicator is a copy-making entity. Genes are such replicators. Furthermore, a good replicator is characterized by three "kinds of stability": "high longlevity/ fecundity/ copying-fidelity" (*ibid.*: 18). Since replicators produce copies of themselves, they are almost, or at least potentially almost, "immortal." The "potential near-immortality" of a replicator, such as the gene, is not only the "defining property" of genes, but of replicators as such (*ibid.*: 35).

*(i). Similarity requirement.* In this definition of a replicator, we find the first general requirement an entity has to fulfill to count as a replicator. Something is only then a replicator, if there is a high similarity between the original and the copy. This is what I call the *similarity requirement*: A replicator is an entity that persists over time in form of similar copies. Implicit in this similarity requirement is a certain definition of genes. According to Dawkins, the gene-as-replicator is that string of DNA, out of a gene complex, that "survives through a large number of successive individual bodies" (Dawkins 1989 [1976]: 25). This gene concept is "not a rigid all-or-nothing definition, but a kind of fading-out definition" (*ibid.*: 32). The question, how big a portion of the genome a single gene is, cannot be answered in a general way. Dawkins writes: "The more likely a length of chromosome is to be split by crossing-over, or altered by mutations of various kinds, the less it qualifies to be called a gene in the sense in which I am using the term" (*ibid.*: 32).

This gene concept is an evolutionary one. It differs from two kinds of functional gene concepts. The evolutionary gene, the gene-as-replicator is not identical with that sequence of DNA that codes for a single protein and is transcribed into a functional RNA, which would be the molecular gene concept, since this segment of the genome can indeed be split up during reproduction. For the same reason, a gene is not necessarily identical with those bits of DNA, distributed over the whole genome, that can be said to be involved in 'programming' one phenotypic trait (functional or developmental gene concept).[39] Dawkins gene concept is an evolutionary gene concept, pointing to the role DNA plays in an evolutionary perspective.

Furthermore, the gene-as-replicator is a concrete, particulate, relatively stable, atom-like string of DNA-bits that persists over time. At the same time, however, the gene-as-replicator must be an abstract entity, a *type*, or a lineage of tokens, since no particular replicator-token (DNA-string), which, indeed, makes copies of itself, survives. What survives is the copy, the replica. But the replica will die as well. No gene-token, no replicator-token, survives, but the gene or replicator *as a type does,* or – if one does not want to involve types – the lineage of concrete DNA-bits survives.

*(ii). Lineage requirement*. This directly leads to the second replicator requirement. Not any type-token-relationship that secures similar tokens (similarity requirement) counts as replication. A replicator involves a special ancestor-descendant relationship of the tokens. In replication, the tokens must form a *lineage*, as Dawkins adds in a note to the second edition of the *Selfish Gene*:

> "All printed copies of this book will be the same as one another. They will be replicas but not replicators. They will be replicas not because they have copied one another, but because all have copied the same printing plates. They do not form a lineage of copies, with some books being ancestral to others. A lineage of copies would exist if we xeroxed a page of a book, then xeroxed the xerox, then xeroxed the xerox of the xerox, and so on" (Dawkins 1989 [1976]: 274).

A lineage in the narrow sense is only present if the replicas "have copied one another," i.e., if one ancestor has only one descendant and each descendant has exactly one ancestor. This is the case for gene-tokens. If, however, many copies are made from one original, as in Dawkins book example, we do not have a replicator, since one ancestor will have many copies. Likewise, if a

---

[39] See Kitcher (1992) for a short review of these different gene concepts. See Beurton, Falk & Rheinberger (2000) for the history of the concept of the gene in development and evolution.

book is produced by copying many superimposed books, then the copy is a blend or average with many ancestors. In such a case, we would also not have replication in the narrow sense. In both cases there would be no lineage and hence no replicator in the narrow sense. Therefore, the second requirement an entity has to fulfill in order to count as a replicator is what I call the *lineage requirement*.[40] As Dawkins says, whether the lineage requirement is fulfilled, is important for evolution: Only if such an ancestor/descendant relationship holds, does the series of copies have the "potential to evolve," since only then can a "new blemish that showed up anywhere along the series […] be shared by descendants but not by ancestors" (Dawkins 1989 [1976]: 274). If all copies were made from one original, evolution would not be able to be cumulative.[41]

*(iii). Non-triggering requirement*. There is a further way how two items can be similar without being replicated in the narrow sense, namely if the first item merely triggers the second item. I will call this third replicator requirement the *non-triggering requirement*. When a string of DNA is produced, this string has its structure not merely because the presence of the original DNA initiates the occurrence of a similar DNA-molecule, whose structure is determined by something else. The original DNA determines the structure of the copy. As Sperber has put it in informational terms, B is only then a copy of A, if the process that generates B obtains the "information that makes B similar to A from A" (Sperber 2000: 169). For this reason, contagious laughter is usually not considered as a process of copying or replication. My laughter merely triggers yours. It is similar, but it is similar because laughing (how it looks like, how it is done) is innate, not because you copy my laughter. You are not copying me; you are merely triggered by me to do the same as me.

The latter two requirements show that it is not only similarity that is relevant for the concept of a replicator. The kind of mechanism that causes the similarity, "the *causal* relation linking the copy to the copied" is important as well (Godfrey-Smith 2000: 405; Emph. in the orig.). This comes even more to

---

[40] Sterelny (forthcoming a) makes the same point in requiring for a "true copying process" that the copy must be a copy of a particular original. If the copy is drawn from numerous sources, it is not true copying.

[41] See also Dawkins (1982a: 95).

the fore, if we look at the main reasons Dawkins cites for excluding other entities from being replicators, such as organisms or nests, although they can be said to reproduce. Looking at these reasons helps to make the similarity requirement more precise and will lead to further replicator requirements.

*(a). Organisms do not make copies*. Dawkins wants to exclude any kinds of organisms (sexual and asexual) from being replicators. His first argument is the temporary existence of organisms: "[T]hey do not persist through making copies of themselves" (Dawkins 1989 [1976]: 34). The point is not that individual organisms die. As I mentioned above, individual DNA-molecules die as well. The point is that they do not make copies of themselves. But this holds only for sexually reproducing organisms, since asexually reproducing organisms make copies of themselves. Furthermore, even sexually reproducing organisms produce *similar* organisms: 'like begets like' due to heritability. As mentioned in section 2.1, heritability means that parents and offspring are correlated in terms of similarity. It means that parent and offspring are *more* similar than other pairs of individuals, and that this is due to shared genes and not due to environmental influence. Since traits of organisms show heritability, Lewontin (1970) and others stated: Heritability is enough for evolution to occur.[42] They admit that reproduction does not exhibit such a high 'copy-fidelity' as the replication of genes, but they stress that copy-fidelity is a measure of degree. For them, there is no qualitative difference between genes, organisms, or other types of entities that reoccur generation after generation. As Lewontin writes:

> "The heritability is highest in units where no internal adjustment or reassortment is possible since such units will pass on to their descendent units an unchanged set of information. Thus, cell organelles, haploid organisms, and gametes are levels of selection with a higher heritability than diploid sexual genotypes, since the latter do not perfectly reproduce themselves, but undergo segregation and recombination in the course of their reproduction." (Lewontin 1970: 8).

The only thing that changes, if we have individuals as units of heredity, and therefore as units of selection, is the rapidity of selection. Although rapidity of

---

[42] For instance, Sober & Wilson (1994: 538), Godfrey-Smith (2000), Gould (2001: 214). See also Hull (1980) on this position.

selection depends upon the degree of heritability, a lower heritability does not make selection impossible. It only makes it slower. Thus, the objection is that one either has to count organisms as further replicators (widening the similarity requirement), or one admits that evolution is possible without replication in the narrow sense. In both cases, organisms would also have to count as units of selection: units that persist long enough to be the base for evolutionary changes. I will not take a stance on the issue. I merely wanted to point out that there is a problem whose solution – how much similarity is enough – is controversial. The solution itself is not important here.

*(b). Dawkins Weismannian reply.* What is important, however, is that this problem is the reason why Dawkins cites a second main reason for the exclusion of organisms as replicators. According to Dawkins, the “fundamental reason” (Dawkins 1989 [1976]: 274) why organisms cannot be replicators is that acquired changes, phenotypic changes acquired during lifetime, are not handed down to the descendants of the organism.[43] Thus, the argument is that since Lamarckian inheritance of acquired characteristics is impossible, no organism – be it a sexually or asexually reproducing organism – can be a replicator. As outlined already in section 2.2, inheritance of acquired characteristics is one of the assumptions in Lamarck’s theory of evolution. Neo-Darwinism (or Weismannism) is strongly opposed to it: Only non-acquired, genetic characteristics are assumed to be heritable. This is the so-called “central dogma” of Neo-Darwinism (Crick 1958: 153). In contrast to this, inheritance of acquired characteristics was accepted almost universally at Darwin's time (Zirkle 1946). Even Darwin believed in the inheritance of acquired characteristics and tried to explain it with his hypothesis of ‘gemules’ produced and inherited through pangenesis (Darwin 1868). Before Weismann (1892) introduced the strict distinction between germ and soma,[44] it was common belief that the hereditary material in the sex cells is produced from the material of the developed body. Given such a system, changes that the

---

[43] See also Dawkins (1982a: 97-99, 1982b).

[44] Later, Weismann’s germ-soma distinction has been replaced by the genotype-phenotype distinction.

organism acquired during lifetime could be inherited. Since the material that makes up the organism constantly changes, the hereditary material was also considered to be ‘soft’, i.e., pliable. Heredity was considered to be based on so-called *‘soft inheritance’* (Mayr 1982: 959). Weismann, however, stated that heredity is ‘hard.’ He claimed that the germ plasm contains the hereditary material and that this germ plasm is a special substance that is not produced by somatic tissue. The hereditary material is present from the start, not made by or out of the material of the organism, and protected against any changes that occur in the somatic tissue. Biological inheritance is *hard inheritance*: The hereditary material is ‘hard,’ i.e., isolated, present from the start, persistent throughout the whole life-cycle, and stable against acquired changes that occur during the life of an individual.

According to Dawkins, it is Weismannism that excludes organisms from being replicators. An entity can only be a replicator, if changes to the entity are inherited. Through pointing to hard inheritance, which implies the non-inheritance of acquired characteristics, Dawkins can exclude even asexual reproducing organisms from being replicators. The characteristics of the organisms are not copied. Only their genome is copied. As Sober & Wilson (1994: 538) answer, given this argument, gene selectionism relies on the universal truth of Weismannism. If we empirically find out that there are organisms for which the central dogma does not hold (organisms whose acquired changes are inherited), then we might well consider them as replicators. In such a case, heritability of whole organisms would be secured on the basis of soft inheritance, leading to the replication of these organisms. Now, the following question arises naturally: If we assume, for the sake of argument, that there can be asexually reproducing organisms that reproduce through soft Lamarckian inheritance, could we then count them as replicators? According to Dawkins, we still could not, since they do not self-replicate.

*(iv). Self-replication requirement.* Dawkins explicitly requires self-replication for true replication. He does this, for instance, to argue against nests as replicating entities (Dawkins 1982a: 99). Birds build nests and these nests reoccur at each generation of birds, but Dawkins assumes that nests do not

replicate. He reacts to a claim of Bateson (1978), opposing Dawkins gene selectionism that a bird (and its nests) is merely a 'gene's way of making another gene.' According to Bateson (1978), who objects to giving genes so much causal priority, a gene could as well be regarded as a 'nest's way of making another nest.' Dawkins answers:

> "There is a causal arrow going from gene to bird, but none in the reverse direction. A changed gene may perpetuate itself better than its unmutated allele. A changed nest will do no such thing unless, of course, the change is due to a changed gene, in which case it is the gene that is perpetuated, not the nest. A nest, like a bird, is a gene's way of making another gene. […] The special status of genetic factors rather than non-genetic factors [in evolution, MK] is deserved for one reason only: genetic factors *replicate themselves*, blemishes and all, but non-genetic factors do not" (Dawkins 1982a: 98f; Emph. added).

Nests do not count as replicators, since ancestor nests have not the required direct causal connection to descendant nests. As the lineage and the non-triggering requirement, the self-replication requirement points towards a certain causal connection between original and copy. Nests get reproduced only as a consequence of the replication-machinery of something else. Before a nest can reoccur, the nest-building organism has to reproduce and this requires that genes replicate. Genes seem to self-replicate in the sense that one gene can lead directly to another one. That is why Dawkins considers them as self-replicating. Birds and nests do not self-replicate in this sense.

Yet, as critics, like Lewontin (1991: 48) or Griffiths & Gray (1994; 1997) have pointed out, it is a very crude oversimplification to say that genes self-replicate. Genes replicate, but they do this by a complex machinery. The complex machinery is so important that it would be more precise to say that genes are replicated *by this complex machinery* in the same way as nests are replicated by the complex machinery of the habits of birds. Whether this counts as a critique depends on what one means by 'self-replication.' If Dawkins answers that the complex machinery is not an argument against what *he* means by self-replication,[45] since any causal process that secures similarity between

---

[45] Dawkins implicitly does this through counting xeroxing sheets of paper as replication, which definitely relies on heavy copying machinery (Dawkins 1982a: 83). Dennett, defending Dawkins, answers explicitly in this way. Any causal chain connecting original and copy is enough for counting the entity as a replicator (Dennett forthcoming).

copy and original of a gene is enough for self-replication, then he ends up with a broad concept of replicators that has to count organisms, nests, and a number of other things (like thumbs) as replicators (Godfrey-Smith 2000). Yet, this would not only let the concept "collapse," since it becomes too broad, as Godfrey-Smith (2000: 410) claims. It would also undermine the central claim of Dawkins gene selectionism, namely that organisms cannot be replicators. However, Dawkins offers a further way to single out genes as the sole replicators: Replicators are active.

*(v). Active replicator requirement.* Dawkins introduced the concept of "active replicators" in the *Extended Phenotype* (1982a: 83). Genes now appear as a special sort of replicators, namely active germ-line replicators. Dawkins defines

> "a *replicator* as anything in the universe of which copies are made. Examples are a DNA molecule, and a sheet of paper that is xeroxed. Replicators may be classified in two ways. They may be 'active' or 'passive', and, cutting across this classification, they may be 'germ-line' or 'dead-end' replicators." (Dawkins 1982a: 83).

Dead-end replicators (active or passive) can only be copied for a finite number of times. DNA-bits of body cells count as dead-end replicators. A germ-line replicator (active or passive) "is potentially the ancestor of an indefinitely long line of descendant replicators" (*ibid.*: 83), such as bits of germ-line DNA. More important than the distinction between dead-end and germ-line replicators is Dawkins' distinction between active and passive replicators.

> "An *active replicator* is any replicator whose nature has some influence over its probability of being copied. For example a DNA molecule, via protein synthesis, exerts phenotypic effects which influence whether it is copied […]. A passive replicator is a replicator whose nature has no influence over its probability of being copied. A xeroxed sheet of paper at first sight seems to be an example, but some might argue that its nature does influence whether it is copied, and therefore that it is active: humans are more likely to xerox some sheets of paper than others, because of what is written on them, and these copies, are, in their turn, relatively likely to be copied again" (Dawkins 1982a: 83).

The xeroxed sheet of paper as a replicator shows that self-replication in the end is not central to Dawkins replicator concept, since a xeroxed sheet of paper clearly does not self-replicate in any sense.[46]

However, what is most important at this point is the thesis that replicators are taken to have certain phenotypic effects. That genes are active germ-line replicators means for Dawkins that genes have causal priority over organisms and environmental or developmental factors, although the latter also influence the growth of organisms. To distinguish the influence of genes from the other factors, Dawkins often uses stronger language, in order to make the case for the gene-as-replicator as a primary causal agent in evolution. DNA is said to 'program' or 'code' for phenotypic effect, they are the primary 'controllers' of developmental processes etc. Through 'building' organisms they control their own destiny. There could be given many references for this language and many withdrawals from it as well. The most famous phrases are on humans as "lumbering robots" and "survival machines" of selfish replicators (e.g. Dawkins (1989 [1976]: 19f). All of this often highly metaphorical language amounts to the following: Organisms are mere "units of phenotypic power of replicators" (Dawkins 1982b: 167). Although Dawkins grants that selection acts on interactors, that they thereby are "functional units of great importance" (Dawkins 1982a: 112), he stresses that they are mere *vehicles*: "A vehicle is an entity in which replicators (genes or memes) travel about, an entity whose attributes are affected by the replicators propagation" (*ibid.*: 112). Groups and species can also be vehicles. Vehicles have differential success in survival, since they are selected during interaction with their environment. But this success is „differential success of vehicles in

---

[46] Sober & Wilson (1994: 538) do not distinguish between Dawkins' self-replication requirement and his requirement that replicators must be active, when they criticize that a page fed into a copying machine is not self-replicating and therefore not "active." That the page is not self-replicating in the sense of 'able to replicate even without a further complex machinery,' as genes and not nests are regarded to do by Dawkins, does not make them totally passive in the sense of Dawkins. However, if 'self-replication' is interpreted as excluding those units that are reducible to other units that code for it, then the self-replication requirement is reducible to the active requirement. I do not want to say that this interpretation is wrong. Dawkins can indeed be interpreted in that way. But if one does so, than it becomes totally mysterious why Dawkins distinguishes between self-replication and active replication at all.

propagating the replicators that ride inside them“ (Dawkins 1982b: 166). Last but not least, the vehicles or phenotypic effects do not even have to be a part of the respective organism: A nest is a phenotypic effect of the genes coding for nest building behavior. It is not a part of the organism, but a part of the “extended phenotype.”[47] To summarize, the fundamental reason why, according to gene selectionism, organisms, nests, thumbs etc. cannot be replicators lies in the idea that they are a mere consequence of a unit that has a replicating and phenotype-building ‘power.’

### Genes’ phenotypic effects and bookkeeping

Although Dawkins repeatedly admits that genes always interact with each other and with environmental and developmental factors, he nonetheless claims that the phenotypic power of interactors can be reduced to the genes coding for that behaviour. The phenotypic power of interactors can be reduced to the “phenotypic power of replicators“ (Dawkins 1982b: 167). And he has to claim this, since only then can he say that a gene is selected *for* a certain phenotypic trait. Genes can only be units of selection, if they are selected *for a specific phenotypic effect that reoccurs each generation of organisms*. Hence genes can only be *selected* for this or that, if they *are ‘genes for’* this or that. Units of selection must be *functional units*. Only through the reduction of phenotypic traits to the causal power of genes is it possible to say that, if an organism is selected for having an eye, this eye ultimately serves the benefit of those genes that cause this organism to have such an eye. “If we wish to speak of adaptations as being ‘for the good of’ something, that something is the active, germ-line replicator” (Dawkins 1982a: 113). They eye of the organism is just

---

[47] The concept of an extended phenotype is a further and, according to some, the most powerful argument that Dawkins offers against organisms as units of selection. See Dawkins (1982a), and summarized in Dawkins (1989 [1976]: ch. 13). He tries to show that an organism is not even necessary for evolution, as long as a replicator is accompanied by some interaction with the environment that makes replication differential. This can be through the replicator itself, being replicator and interactor at the same time, through a phenotypic effect in an organism containing the replicator, through an entity caused by the replicator in the environment, such as a nest, or through an effect of the replicator on a host organism, such as in the case of viruses. Since this argument is not so much relevant for this investigation and would have to be dealt in detail, I can only point to the relevant literature debating the issue, e.g.: Sterelny & Griffiths (1999), Jablonka (2004), Laland (2004), Dawkins (2004).

the vehicle making the survival of genes differential. Therefore, only the gene can be the unit of selection: it (1) survives long enough to benefit from evolution, and it is (2) the unit that ultimately causes those traits for which organisms are selected directly. Genes are the units whose *fitness* (in ultimately having properties that cause their survival to be differential, i.e., in being active) explains the spread of genes and the spread of adaptations, which are exhibited by the organisms having these genes. In other words, the causal function of vehicles, actually exhibiting adaptations and interacting with the environment, can be reduced to the gene as the ultimate unit that is causally responsible for this interaction, as Lloyd (forthcoming 2005) puts the issue.

But, and this is the core of the issue, if units of selection are those units that are selected for, the question is whether selection really can discriminate *for or against single genes*, given that genes are defined as replicators, i.e. as strings of DNA that are potentially long-living, not broken up by cross-over etc. This question is at the center of the debates that address the causal issues involved in the units of selection debate. Until today, no consensus has been reached. I can only review the core arguments. The main argument of critics is that attributing to single replicators a fitness and phenotypic consequences of their own is distorting the causal process. Genes are only ‘bookkeeping’ what happens on the organismic level.

Today, it is uncontroversial to define evolution as a change in gene frequencies in a population. As countable units, genes keep the book of evolutionary change. If this is the case, than single genes do have a fitness value of their own that can be “treated algebraically,” as the arithmetic mean effect of the respective alleles in a population. Nonetheless, this does not qualify them as the causally responsible units of selection, as critics claim.[48] As Gould summarizes the opposition to gene centrism: Gene centrism rests on a “confusion of bookkeeping with causality” (Gould 2001: 208). The problem is that genes would need attributable context-independent effects on phenotypes: effects for which *they* are selected, effects that are their effects, and not the

---

[48] Sober & Lewontin (1982), Wimsatt (1980), Sober & Wilson (1994), Gould (2001), Lloyd (forthcoming).

effect of a larger junk of DNA, in the limit the whole genome. That Dawkins' replicators do *not* have such a context-independent effect is suggested by the fact that phenotypes are caused by a complex network of gene interaction, which differs in each organismic context, since the gene meets different genes in different organisms.[49] The bits of DNA that are singled out by Dawkins' evolutionary concept of genes as replicators might not have such a context-independent effect. Dawkins and others have countered this critique with the argument that we can treat other genes as a background condition, as a genetic environment influencing the fitness of a single gene.[50] Genes are difference-makers with a context-independent effect, if we regard other factors as constant, i.e., if we stick to 'ceteris paribus.' Whether this suffices to show that the context-independent phenotypic effects of genes and the respective averaged fitness of individual genes are more than mere mathematical "artifacts," as Sober & Lewontin (1982) object, cannot be answered here. The details of the arguments involve assumptions about causation, emergence, ceteris paribus clauses, and the like, which would lead too far away from the goal of this investigation. Important here is not the answer to the units of selection debate. Important is rather that it does not suffice to claim that replicators exist and that evolution can be represented through counting gene frequencies. You can track evolution through regarding genes as the ultimate *units of heredity,* which preserve their structure with a high copy-fidelity. But in order to regard them as units of selection, as the entities, which selection discriminates for or against, you need further assumptions about the causal picture that are contentious.

---

[49] Note that this is different from Mayr's (1963, 1975) and Brandon's (1982, 1999) argument that the causal influence of phenotypic characteristics 'screens off' the causal impact of their genotypic basis. According to them, the organism is a more direct cause of selection and is thus the unit of selection. The latter addresses the relation between phenotype and genotype, while what I have described addresses the relation between single genes-as-replicators and genes in the developmental sense (those bits of DNA that code for a single phenotypic trait) or the relation between single genes and the whole genotype. This difference has been made clear already by Sober & Lewontin (1982). What is a problem for gene selectionism is not whether the direct or ultimate causal agent is the unit of selection, but whether there exists such an ultimate causal agent that has a context-independent phenotypic effect and replicates at the same time.

[50] Sterelny & Kitcher (1988), Waters (1991), or Sterelny et al (1996).

### Conclusion

The units of selection debate is about the roles certain entities play in Darwinian evolution. I introduced the replicator concept, which formalizes the role genes play in evolution, and which was meant to show that genes are the units that are selected for by evolution. According to Dawkins, a replicator in the narrow sense has to meet five requirements: (i) the similarity requirement, (ii) the lineage requirement, (iii) the non-triggering requirement, (iv) the self-replication requirement, and (v) the active-difference-maker requirement. The latter concept is connected to the claim that genes have a fitness of their own, and a context-independent phenotypic effect. A replicator in a broad sense only has to fulfill the similarity requirement. According to this concept, almost anything that reoccurs in form of copies would count as a replicator. The narrow concept of replicators is therefore necessary to single out genes as the one and only replicators of biological evolution. In chapter 3, I will put into question whether memes are replicators in the narrow sense. In chapter 5, I will show that attributing *memes* a fitness of their own cannot serve as a basis for considering them as the ‘selfish’ units of cultural selection. The reason is similar to why genes as explanatory prior units of selection are controversial: To attributing them a fitness of their own might be mere bookkeeping.

## 2.6 SUMMARY

We saw that a Darwinian explanation of change has the following structure: It explains systems that exhibit variation, fitness differences, and heredity by referring to a sorting process, which is in the case of biological evolution mainly natural selection. But drift is an alternative sorting process compatible with a Darwinian paradigm. Variation arises through a process that is decoupled from selection, leading to undirected variation. Selection is imposing direction on this variation ex post facto and can only work in an opportunistic manner, leading to local adaptation, according to immediate pay-off, and doing this in the manner of a ‘tinkerer.’ Fitness is a propensity for survival and reproduction. If the concept of fitness is not connected to the causal basis of this propensity, i.e., to traits of the respective entity that make up his adaptedness, then the principle of ‘survival of the fittest’ is tautological, explaining survival via survival. Different entities that can be distinguished in biological evolution can play different roles in this evolutionary process. Organisms are the paradigmatic interactors, and genes are the paradigmatic replicators. The controversial issue is whether the causal role of interactors can be reduced to the causal role of replicators.

# 3 ONTOLOGICAL ANALOGY: GENES AND MEMES

## 3.1 UNITS OF CULTURE

### The gene-meme-analogy

The ontological analogy between biological evolution and culture compares genes and *ideational units of culture*, such as ideas, beliefs, rules for behavior, or values. In the 19th century, it was organisms (or organs) and artifacts that had been compared by natural theology or evolutionists. Paley, for instance, draws an analogy between the eye and the telescope because of the perceived design (complexity and functionality) in both of these entities. Although the analogy between organisms, organs, and artifacts was at issue in the debates about the argument from design and is still at issue in current debates about function ascriptions,[1] it is of no importance in the two approaches that are at issue here: the Darwinian approach to creativity and memetics.

As the evolution of species is defined as change in gene frequencies and counted through genes, since genes are as replicators the 'bookkeeping' basic units of heredity in biology, it is assumed that culture involves similar basic building blocks: units of cultural heredity that can be tracked and counted, that can therefore be used to 'keep the book' of cultural change, and that are replicators. Dawkins (1976) called these building blocks *memes*: ideational units, such as ideas, beliefs, rules for behavior, values, concepts and the like. The ontological analogy at issue here is an analogy between genes and memes.

This ontological analogy is based on three assumptions: First, it rests on the assumption (1) that there *are* such ideational units of cultural heredity. Second, it rests on the assumption that (2) we can *count* these units. I will call this second assumption *traceability condition of memes*. Third, it rests on the

---

[1] See, for instance, McLaughlin (2001), Lewens (2004).

assumption that (3) both are *replicators*. This third assumption will be called *replicator condition*.[2]

## The anthropological concept of culture

The first assumption, (1), that there are basic ideational building blocks of culture, simply mirrors the contemporary anthropological concept of culture. Since memeticists and most critics do not care much about the history of the concept of culture, as it has developed in humanities and social sciences, especially anthropology, they miss this point.[3]

Etymologically the term 'culture' comes from the Latin term *colere*. It means the tending of natural growth or 'husbandry'. Cicero, who used the term *cultura animi* for the tending of the soul, made the first application beyond the sphere of agriculture. This concept of culture, referring to the enhancement or perfection of the soul, is still found in the French 'Enlightment' concept of culture as civilization – the universal progress of humanity. In contrast to this, a more 'Romantic' and nationalistic concept of *Kultur* was developed in Germany: Culture means national identity. That each nation had its culture means that "[e]ach *Volk* had its own *Geist*, and its specific destiny," as Adam Kuper (2002: 88) puts it. Herder is one of the most well-known philosophers in this tradition. In parallel to these two concepts, there had always been a third, more elitist or 'classic' concept of culture: only art, philosophy and the like were culture. Culture is 'high culture.' These three concepts were prevalent in all the humanities.

In the 19th century then, a distinct scientific concept of culture developed during the formation of anthropology as a scientific discipline. Edward B. Tylor was the pioneer of this development. The two volumes of *Primitive Culture* (1871) are considered as the founding texts of modern anthropology as a scientific discipline. The first volume, entitled *Origins of Culture*, begins with Tylor's now classic anthropological definition of culture:

---

[2] The fourth assumption – a part of memetics but not a part of the ontological analogy – is that memes and not minds explain why we think what we think. It will be addressed in chapter 5.

[3] For exceptions see Plotkin (2002) and the critique of memetics of Kuper (2000) and Bloch (2000).

> "Culture or Civilization, taken in its wide ethnographic sense, is that complex whole which includes knowledge, belief, art, morals, law, custom, and other capabilities and habits acquired by man as a member of a society" (Tylor 1958 [1871]: 1).

This definition is important for two reasons: (i) First, the definition implicitly establishes an opposition between *biological inheritance and social learning*. Culture is 'acquired by man as a member of a society' and thus not innate. Hence, the concept that culture is a separate system of inheritance, based on social learning, which I introduced already in chapter 1, goes at least as far back as to the 19th century. This is the first of two important aspects of the contemporary scientific concept of culture that will be essential for the evaluation of the ontological analogy. The second aspect, which will be relevant for that goal, is: (ii) Tylor enumerates diverse *units of culture* that are organized into an integrating 'complex whole,' abstracted from the conglomerate of these units.

In 1952, the anthropologists Alfred Kroeber and Clyde Kluckhohn (1952) tried to summarize the developments from Tylor until then. They listed 164 definitions and several statements about culture. Based on this review, they suggested a summarizing formula as a kind of consensus for the anthropological concept of culture at their time:

> "[C]ulture consists of patterns, explicit and implicit, of and for behavior acquired and transmitted by symbols, constituting the distinctive achievement of human groups, including their embodiments in artefacts; the essential core of culture consists of traditional (i.e., historically derived and selected) ideas and especially their attached values; culture systems may, on the one hand, be considered as products of action, on the other as conditioning elements of further action" (Kroeber & Kluckhohn 1963 [1952]: 357).

Let me reconstruct this formula with respect to the above mentioned two aspects: (i) As consisting of 'products of action' that are at the same time 'conditioning elements of further action,' as 'acquired and transmitted by symbols,' being thereby the 'distinctive achievement of human groups,' culture is *created* by man, transmitted through *social learning*, and thereby maintained in history through *traditions*. (ii) As consisting of 'patterns of and for behavior,' as transmitted by 'symbols,' culture has its 'essential core' in '*ideas* and especially their attached values,' which are merely 'embodied' in artifacts. These two aspects have already been present in Tylor's definition. But Tylor did not put so much stress on ideational units. Even if the concept of culture still is subject to controversial debates, not much has changed since Kroeber & Kluckhohn, at least not with respect to the two aspects at question here.[4] I will now turn to a more detailed description of these two aspects.

*(i). Culture as created by humans and transmitted through social learning*. The contemporary scientific concept of culture assumes that humans either create cultural units or learn them from others. The way a traditional pot is made in Poland in the 20th century is not specified in the genes of Polish people. Somebody invented the pattern of making the pots and others have learned these patterns. Culture is thus based on individual learning, creativity in a more narrow sense, and social learning. (The distinction between individual learning and creativity will be explained in chapter 4.) Traditions, which conserve culture over time, are thus maintained via non-genetic transfer from individual to individual. Culture is something 'added to' our biological design or equipment: We create things that are not innately specified and hand them down to others through means that are not part of biological inheritance.

This is, as you might recall, also the aspect that distinguishes the analogical applications of Darwinism from the literal extensions mentioned in chapter 1. Evolutionary psychology uses a wide notion of culture: Cosmides & Tooby, for instance, define culture as "any mental, behavioral, or material commonalities shared across individuals […] regardless of why these commonalities exist" (Cosmides & Tooby 1992: 117). Culture is behavior. According to them, culture is merely 'evoked' through experience in the world and basically innate. As indicated in chapter 1, the problem with this wide notion of culture is that it ignores that behavior is caused by different factors that can hardly be isolated in practice but have to be distinguished in theory: genes and other biological factors, environment, social structure, and culture – as consisting of socially transmitted ideational units, such as ideas, values etc. To say that every behavior is culture and that culture in this sense (i.e., behavior) can be reduced to genes and mere triggering factors is a sleight of

---

[4] See Ingold (1994), Kuper (2000), or Fox & King (2002) for other historical and current issues about the culture concept.

hand that reduces culture by defining it away. It is ignoring the “critical generative dimension” of culture, as Weingart et al (1997: 301) write, which can only come into focus, if we define culture as a second system of inheritance of ideational units.

This directly leads us to the second aspect, namely the units of culture. Before I turn to that, I want to point to the implied relation between cultural items and humans: On the basis of a concept of culture that does not ignore the generative dimension of culture, humans are creators of culture, they select cultural items and they are the culture bearers.

*(ii). Culture as based on ideational units.* Culture is related to entities that fall in three ontological categories: ideational units, social units, and observable units. *Ideational units* include mainly mental states that include a cognitive content (i.e., beliefs, ideas, values, etc.; skills, habits, behavioral rules, and the like). But not all mental states can be cultural units, since these contain, by definition, *sharable* cognitive contents: Color sensations and emotional states, for instance, cannot be shared and are therefore excluded from the cultural realm. Nonetheless, we might share (i.e., transmit to others) our knowledge about colors. *Social units* comprise mainly social institutions, like laws or universities, and social roles, like being a minister. The *observable units* are behavioral patterns (such as rituals or actions), and artifacts (such as books, the university building, or other created objects like pots).

The contemporary concept of culture gives ideational units a predominant role. This ideational concept of culture, used already by Kroeber & Kluckhohn (1952) was furthered when Talcott Parsons distinguished between the social and the cultural. Parsons (1951) tried to develop an interdisciplinary, but unified theory of social action. For this reason, he tried to establish a new division of labor for the social sciences: Culture and society are interdependent but separate systems (Kroeber & Parsons 1958). Thus, social institutions are not a part of culture in this ideational sense. Social institutions belong to the social system that – besides other things – structures the space in which cultural items in the narrow sense (i.e., ideational units) are shared and ‘travel’ from person to person – through being learned, i.e., diffused.[5] Parsons’ program has gained dominance through such important students of him like Clifford Geertz and David Schneider, major figures in the anthropology of the second half of the 20th century.[6] For Geertz and Schneider, culture is a ‘symbolic system.’ Culture is for them analogous to a shared text that has to be interpreted with the same methods as any literary text. Their theories were the base for the breakthrough of the so-called ‘ideational theories of culture.’ Keesing (1974) distinguishes their theories from two other main schools that developed along an ideational concept of culture: Ward Goodenough defined culture as a ‘cognitive system,’ as a mental competence that is analogous to a Chomskian language faculty through which humans acquire knowledge that must be known in order to be part of a society. Claude Levi-Strauss and his followers considered culture to be a ‘structural system.’ It consists of symbolic structures, such as myths. Through deciphering these symbolic structures, they believed to find universal principles of the human mind that generate these symbolic structures. I cannot go into the details of the differences between these approaches. What is important here is that all three schools more or less assume that the basic building blocks of culture are ideational units, more or less ‘ideas.’ This ideational concept of culture considers ideational units as generative and as more basic than behavior and artifacts: Behavior and artifacts are a mere consequence of ideational units. Thus, the basic building blocks and units of heredity of culture are the ideational units.

---

[5] For more on Parson and the distinction between culture and society see Kuper (2002: 95ff). Some certainly would object the distinction between the social and the cultural and would prefer to talk about the ‘sociocultural.’ Whether the distinction between the social and the cultural is really clear-cut cannot be answered here. On the one hand, social roles and institutions are somehow represented mentally. On the other hand, because of the complexity of social systems, the hope to describe and explain the change of these systems as part of cultural change – the change in differentially distributed ideational units – is not very realistic, since the entities and structures cannot easily be translated into ideational units, as Luhmann (1997: Part 1, ch. 3, esp. 536ff), for instance, assumes. He therefore distinguishes between his claim of social evolution as the evolution of self-organized autopoietic systems, analogous to organisms, and claims about partial systems of society, like the ‘evolution of ideas.’ A similar distinction is made by Runciman (1998) about the difference between an evolutionary account of ‘social roles’ and cultural units, e.g. beliefs.

[6] See again Kuper (2002), for more on these developments.

This shows that the first assumption of the Darwinian ontological analogy, which was that (1) there *are* basic ideational units of cultural heredity, mirrors the two central aspects of the contemporary scientific concept of culture, as developed in anthropology: (i) that culture is created by humans and transmitted through social learning; (ii) that culture is based on ideational units.

**Identification and replication**

This said, it is important to realize that the contrary holds for the second assumption of the gene-meme-analogy. Anthropologists usually doubt that (2), the traceability condition of the gene-meme-analogy, holds. We cannot easily *identify* and *count* these basic building blocks of culture. As ideational units they can only be inferred and are subject to interpretation. And this is the basis of a common critique against the gene-meme-analogy. I will discuss this critique in my analysis of a set of *identification problems* in section 3.3. Last but not least, the ideational concept of culture does also not entail the third assumption: that social learning, the transmission of these ideational units from person to person, is analogous to *replication,* as defined by Dawkins. This issue, comprising a couple of *replication problems*, will be dealt with in section 3.4.

Before I can address these issues, it is essential to draw a precise picture of what memes are, in order to understand the position of the memeticists and the critics alike. The most frustrating feature of memetics, however, is that even memeticists have different concepts of memes.

## 3.2 WHAT AND WHERE MEMES ARE

**Memes made more precise**

Since the 1970s, when Dawkins first introduced the idea of memes as replicators, the precise nature of these memes has been a controversial issue. What is a meme? Can we find memes only in brains, or are memes also ‘in’ behavior and artifacts? Is there an analogue to DNA and one to phenotypes? As Susan Blackmore, well-known defender of memetics, writes: “[T]he terminology of memetics is in a mess and needs sorting out” (Blackmore 1999:

63). Since these questions have caused and still cause confusion, I have to go into some conceptual detail, which are often enough ignored in the hot ‘meme wars’ between defenders and critics. I will concentrate on the main approaches of Dawkins, Hull, Dennett and Blackmore, who settled on an *ideational concept of memes*. Nonetheless, there are still differences between them: mainly with respect to what the analogue to DNA is, and with respect to what the analogue to phenotypic expressions of genes is. There are other memeticists that do not define memes as ideational units. To make the ideational concept more precise, I will thus introduce the alternative of a *neuronal* concept of memes, presented by Aunger (2002). A further alternative, namely to define memes as *behavioral* units, will be presented at the end of section 3.3: This alternative can partly be interpreted as a reaction to the identification problems. Memes become identified with something else *in order to* restore the analogy between genes and memes, i.e., in order to find *something* in culture that is easily identifiable and mirrors gene replication more closely than the transmission of ideational units. As I will claim, the price you pay for this is a certain trivialization.

### Dawkins’ ideational memes

Dawkins (1976) started memetics with a rather imprecise definition. According to him, the meme is a “new kind of replicator,” a “unit of cultural transmission, or a unit of *imitation*” (Dawkins 1989 [1976]: 192). After saying this, all what he adds as a clarification at this point is the following:

> “Examples of memes are *tunes, ideas, catch-phrases, clothes, fashions, ways of making pots or of building arches*. Just as genes propagate themselves in the gene pool by leaping from body to body via sperms or eggs, so *memes propagate themselves* in the meme pool by leaping from brain to brain via a *process which, in the broad sense, can be called imitation*. If a scientist hears, or reads about, a good idea, he passes it on to his colleagues and students. He mentions it in his articles and his lectures. If the idea catches on, it can be said to propagate itself, spreading from brain to brain” (Dawkins 1989 [1976] :192; Emph. added).

This means that, first, the entities sanctified as new replicators belong to various ontological categories: ideas and skills on the one hand, externally observable behavior and artifacts on the other hand. Second, imitation in the broad sense is self-propagation and analogous to the process of replication. I will say more on the latter in section 3.4.

After explaining why memes are selfish, Dawkins states that we can, third, divide a meme-complex (e.g., a whole symphony or a theory), into single memes, as we can divide a gene-complex into single genes. He then makes clear that, fourth, we actually have to distinguish between the meme as an essence of something (i.e., the "idea-meme") and its interpretations in the minds of diverse individuals. Darwinism is his example. He writes:

> "[W]hen we say that all biologists nowadays believe in Darwin's theory, we do not mean that every biologist has, graven in his brain, an identical copy of the exact words of Charles Darwin himself. Each individual has his own way of interpreting Darwin's ideas. […] Yet […] there is something, some essence of Darwinism, which is present in the head of every individual who understands the theory. If this were not so, then almost any statement about two people agreeing with each other would be meaningless. An 'idea-meme' might be defined as an entity that is capable of being transmitted from one brain to another. The meme of Darwin's theory is therefore that essential basis of the idea which is held in common by all brains that understand the theory. The *differences* in the ways that people represent the theory are then, by definition, not part of the meme." (*ibid.*: 195-6; Emph. in orig.)

With this statement Dawkins want to secure that there is one and the same meme in different heads at all. It is important to realize that this 'essentialism' is similar to the essentialism that is hidden in the modern concept of hard inheritance. As explained in section 2.5, this concept assumes that genes are the heritable units, which are hidden inside of bodies and 'hard' – strictly separate from the somatic tissue, present from the start, continuously existing, and protected against acquired changes. Acquired changes are those that arise from the influence of the external environment during development, which builds a context that varies from organism to organism. Acquired changes are not part of the genes, do not become integrated into the genetic heritage, and are thus not heritable. Acquired changes are mere temporary, arbitrary, or even 'imperfect' realizations of the gene. These phenotypic realizations change according to the context, but the gene stays the same – except, of course, when mutations occur.

If memes are analogous to genes, and if there is something (i.e., the interpretation of the meme) that changes from context to context, whereas the meme stays the same, then memes have a context-dependent phenotypic 'expression.' Now, the following interpretational problem arises: Do they also have a material realization in a 'meme-DNA', which is invariant across contexts? If there is no such context-independent material realization of the essential Darwinism-meme, than this means that memes do not have a DNA, as genes have a DNA that stays the same in different organisms, despite different phenotypic realizations. Thus is the meme a purely abstract entity *without* a clear material identification? Dawkins statements in 1976 allow no definite answer to this question about a DNA-analogue. Nonetheless, what is clear is that Dawkins regards memes as abstract, essential entities: memes have an essence that is shared despite concrete differences in the minds of individuals.

Dawkins was more precise in his *The Extended Phenotype* (Dawkins 1982a: 109-112). He writes:

> "I have previously supported the case for a completely non-genetic kind of replicator, which flourishes only in the environment provided by complex, communicating brains. I called it the 'meme' […] I was *insufficiently clear about the distinction between the meme itself, as replicator, on the one hand, and its 'phenotypic effects' or 'meme products'* on the other. A meme should be regarded as a *unit of information residing in a brain* […]. It has a definite structure, realized in whatever *physical medium* the brain uses for storing information. […] This is to distinguish it from its *phenotypic effects*, which are its consequences in the outside world […]. The phenotypic effects of a meme may be in the form of words, music, visual images, styles of clothes, facial or hand gestures, skills such as opening milk bottles in tits, or panning wheat in Japanese macaques. They are the outward and visible (audible, etc.) manifestations of the *memes within the brain*" (Dawkins 1982a: 109; Emph. added).

Only ideational units are regarded as memes. Variations in physical realizations are not part of these memes; they are phenotypes. The analogues to bits of DNA are brain-patterns. That there are different internal phenotypic expressions, interpretations, of these memes in the minds of individuals, and not only different external realizations, is not addressed anymore.

In his *Foreword to Blackmore*, Dawkins (1999) again slightly changes his concept of memes. On the one hand, there are memes, which have a non-memetic phenotype, understood as imperfect external realizations of an idea or 'instruction.' On the other hand, there are also memes where the "phenotype in every generation is also the genotype" (Dawkins 1999: xi). The example he uses for the first kind of memes are "instructions" (*ibid.*: xi) for making a Chinese paper junk, an origami model of a flat-bottomed Chinese sailing-ship.

The example of the second kind of memes is an artifact, a drawing of a Chinese paper junk. The transmission processes of these two types of memes are different, especially with respect to the copy-fidelity of the transmission. The transmission of the first kind of memes is truly Weismannian (i.e. hard inheritance), while the transmission of the second is 'Lamarckian.' Lamarckian in this context means that the imperfect 'phenotypic' realizations are copied as well. Hence, imperfections accumulate and lead to a total different drawing after a couple of transmissions. I will come back to this distinction between two kinds of memes and two kinds of transmission in section 3.4, where I will say more about copy-fidelity of memes. Important at this point is that the constant back and forth, between defining memes as ideational units and defining them as including observable artifacts and behaviors as well, does indeed create some of the confusion about what memes are.

### Hull's ideational memes

David Hull was the first author who took Dawkins suggestion about cultural units as replicators seriously. Already in *The Naked meme* (Hull 1982) and in later papers (Hull 2000, 2001), he suggested that memes – the sought-for "units of sociocultural evolution" that can play the role of replicators – "can exist in brains, books, computers, and a wide variety of physical vehicles of knowledge" (Hull 1982: 276). Note that the term 'vehicle' here, does not have the special meaning it has in Dawkins' gene selectionism (introduced in section 2.5). Hull is very clear about this, at least in later papers, where he writes: "Using vehicle to refer both to interactors [to which Dawkins vehicles belong, MK] and to the physical basis of replication begs for misunderstanding, and misunderstanding comes along easily enough on its own. One need not beg for it" (Hull 2001: 33). 'Vehicle' in the above quotation just stands for *any physical "substrate" of memes*, the latter being "ideas" (Hull 1982: 310) or "information" incorporated in these substrates (Hull 2000: 58).[7] Hull thus

---

[7] The concept of information is now often used, replacing 'representations' or 'mental content' or 'ideas,' even though its meaning is usually not clarified. As Hull (2000: 58-61) states, the concept of information that should stand for contents that are in minds, books, etc. is still a dubious one. It cannot be the information thermodynamics deals with. It cannot be the

differs from Dawkins in that he does not regard brain-patterns as primary material substrate of memes. While DNA is usually considered to be the one and only material substrate of genes, memes have diverse physical substrates.[8]

Hull (1982: 310) also distinguishes these physical substrates from the phenotypic expressions of memes.[9] Symbolic systems (written or spoken language, musical scores or geometric representations) should not be counted as phenotypic, as Dawkins does, since the relationship between memes and these physical substrates is more similar to the relationship between genes and DNA, than to the relationship between genes and phenotypic characteristics. The reason he offers is that the relationship between symbolic systems and mental content is "structure-retaining," just as the relationship between genes and DNA. The relation between a word and the respective concept, for instance, counts as structure-retaining. Words would therefore count as mere physical substrates of memes, as DNA is a mere physical structure-retaining substrate of genes.[10] Memes are coded in diverse physical substrates, while genes are coded only in DNA. Furthermore, genes and memes are not only coded *in* a physical substrate, they code *for* phenotypic characteristics. As Dawkins, Hull takes memes to 'code for' behavior and products of behavior. But only those instances and products of behavior that can count as a non-structure retaining behavioral "application" or "implementation" of memes are phenotypic expressions of memes. The relation between the idea of peaceful behavior and peaceful behavior itself would thus not count as structure-retaining. Peaceful behavior would merely be a phenotypic expression of the idea, as an eye is a phenotypic expression of genes. The same holds for the

---

information mathematical information theory (Shannon & Weaver 1949) talks about. It must be a semantic concept of information that still waits for its discovery. See Wilkins (1998) for a first step in the direction of a precise concept of semantic information in memes.

[8] See also Hull (1982: 310, 2000: 58).

[9] See also Hull (2000: 58-61, 2001: 33).

[10] Whether spoken words, for instance, still have the same structure as the respective ideas is an important philosophical question, which cannot be addressed here. I will simply take it for granted that there is a difference between, let's say, the phenotypic expressions of such an idea like peaceful behavior, and the mere physical realizations of this concept in the written word 'peaceful behavior.' The word for the concept is more similar to the concept itself than its expression in behavior. However, see Heyes & Plotkin (1989), who criticize Hull for assuming a Fodorian language of thought, i.e., that all mental contents are propositional.

relation between scientific theorems and their application (Hull 1982: 310), and for the relation between a recipe and a cake. The former is the meme, coded in any kind of physical substrate, for instance letters on paper, while the latter is analogous to a phenotypic expression of genes in a respective organism.

The distinction between structure-retaining and non-structure-retaining physical substrates is similar to Dawkins' (1999) distinction between memes with a clear genotype-phenotype distinction (instructions to make a Chinese paper junk) and those memes without a clear genotype-phenotype distinction (the drawing of the Chinese paper junk). Hull's conception, however, has the advantage not to end up with two kinds of memes. According to his conception, there would be one meme (the idea of a Chinese paper junk) that can be realized either in structure-retaining symbolically coded behavior (instructions), or in resulting artifacts (the drawing of a Chinese paper junk or the Chinese paper junk itself), which do not necessarily retain the structure of the meme.

### Dennett's ideational memes

Dennett moves away even further from Dawkins. He is not very clear on what he regards as the analogues to 'phenotypic effects of memes' or whether he really wants to transfer the dichotomy between genotypes and phenotypes at all. In *Consciousness Explained* (1991),[11] Dennett defined the meme, the "new replicator," as

> "roughly, ideas. Not the 'simple ideas' of Locke and Hume (the idea of red, or the idea of round or hot or cold), but the sort of complex ideas that form themselves into distinct memorable units – such as the ideas of wheel, wearing clothes, vendetta, right triangle, alphabet, calendar, the Odyssey, calculus, chess, perspective drawing, evolution by natural selection, Impressionism, 'Greensleeves,' deconstructionism" (Dennett 1991: 201).

As genes are "invisible,"

> "carried by gene vehicles (organisms) in which they tend to produce characteristic effects ('phenotypic' effects) by which their fates are, in the long run, determined" (*ibid.*: 203),

memes are also

---

[11] With respect to the quotations used here, the section on memes in Dennett (1991) is almost identical with Dennett (1990), where he first introduced his version of memes.

> "*invisible*, and are *carried by meme vehicles* – pictures, books, sayings (in particular languages, oral or written, on paper or magnetically encoded, etc.). Tools and buildings and other inventions are also meme vehicles. A wagon with spoked wheels carries not only grain or freight from place to place; it carries the brilliant idea of a wagon with spoked wheels from mind to mind. A meme's existence depends on a *physical embodiment in some medium*" (*ibid.*: 204; Emph. added).

Dennett talks of behavior (e.g., oral language) and artifacts (e.g., the wheel) as "meme vehicles," explicitly in analogy to phenotypic effects or "phenotypic expressions" due to genes (*ibid.*: 206). At the same time he talks of these things as "physical embodiment," as genes have their embodiment in DNA. Evidently, he does not draw a difference between physical substrate and phenotypic expressions, as Dawkins (at least in 1982) and Hull did, although there clearly is one in most cases of biological entities. This ambiguousness is literally reproduced in Dennett's famous *Darwin's Dangerous Idea* (1995: 342-352).[12]

However, as he writes, in order to "consider more carefully what Dawkins' memes are or might be" (Dennett 1995: 342), he tried to add "important improvements" (*ibid.*: 343) in *Darwin's Dangerous Idea*. He more clearly specified his ideational concept of memes. Memes are now

> "*information* – in a media-neutral, language-neutral sense. Thus the meme is primarily a *semantic* classification, not a *syntactic* classification that might be directly observable in 'brain language' or natural language. In the case of genes, we are blessed by a gratifyingly strong alignment of semantic and syntactic identity: there *is* a single genetic language, in which *meaning* is (roughly) preserved across all species. Still, it is important to distinguish semantic types from syntactic types" (*ibid.*: 353f; Emph. in the orig.).

Two aspects of his account are important for this study:

*(i). Memes consist of information that have multiple physical substrates*, in Dennett terms, "mediums," or "physical vehicles," (*ibid.*: 348) or "phenotypes (the 'body design' of memes)" (*ibid.*: 355; *cf.* 349). As we should not "identify genes with their vehicles in DNA" (*ibid.*: 353), we should not identify memes with their vehicles, such as brain-patterns. According to Dennett, memes show best "the separation between information and vehicle," for instance in brain-patterns, since "it is very unlikely – but not quite

---

[12] The section "Invasion of the body-snatchers" (Dennett 1995: 342-352) contains large portions literally reprinted from Dennett (1991).

impossible – that there is a uniform 'brain language' in which information is stored in different human brains" (*ibid.*: 353). Furthermore, memes can have multiple 'vehicles': books, wheels, or any other physical entity.

(ii). *Memes are semantic.* Memes are connected to something they '*are about*,' as genes are connected to their "phenotypic effects – what they are 'about' (such as making hemoglobin, or eyes)" (*ibid.*: 354). Thus, in addition to having various physical substrates, each meme is about something. The 'wheel'-meme (the idea of a wheel) can be 'coded in' brain-patterns, words, artifacts, or behavior such as making a wheel. They are 'coded in' these entities as genes are 'coded in' DNA. The meme is, however, about one thing only, namely the wheel, the artifact, in which it is coded or realized at the same time. The meme of 'peaceful behavior' can be coded in brain-patterns, in words, in artifacts like a peace treaty, or in peaceful behavior. And it has a meaning, which is not easy to find out. A 'moon'-meme can equally appear in diverse physical substrates, but it clearly is about the moon. There are memes that are about artifacts, about behavior (e.g. clothing), about entities, about matters of fact, about institutions …. all the kind of things our intentionality can focus on.

Although it can be doubted that it makes sense to speak of 'genes being *about* phenotypic effects,' as ideas are about something, I would like to focus on a different issue.[13] What is now the cultural analogue to DNA and what the cultural analogue to 'phenotypic effects'? According to Dennett, all kinds of things are cultural analogues to DNA: from brain-patterns, to symbolic artifacts, to behavior and non-symbolic artifacts. With respect to the cultural analogue to phenotypic effects it is unclear what Dennett wants to say. Is it the thing memes are *about* or the diverse physical substrates? Dennett is highly confusing in his usage of the concept of 'phenotypic effects.' On the one hand, Dennett says that memes 'are about' something, as genes '*are about'* phenotypic characteristics. On the other hand, memes *have* 'phenotypic

---

[13] To talk of genes 'coding for' or 'being about' phenotypic characteristics rests on the concept of genetic information. This concept is highly controversial. There have been many articles on this issue in the last couple of years. See Maynard-Smith (2000) and the replies to this article as entrance to the debate. See also Griffiths (2001) for further discussion.

effects,' i.e., 'vehicles' (e.g. words, brain-patterns, books, artifacts), *although the memes are not necessarily about these phenotypic vehicles*.

From an analytic standpoint, what is so confusing in Dennett's version of the gene-meme-analogy is that he ignores an important difference between genes and memes: genes are (if at all) not only 'about their phenotypic effect,' they *causally play a role in building* these 'phenotypic effect of genes.' This is not the case for all the things 'memes' can be about. The 'moon'-meme is not causally involved in bringing about the moon, as the 'eye'-gene is causally involved in bringing about an eye. (If at all, it would be just the other way round: If at all, the moon is causally involved in bringing about the moon-meme, which would be analogous to an eye bringing about its gene). One should not forget that 'coding for' in biology is not a mere '*semantic relation'* between genes and their phenotypic expression – if it is one at all; it is – first and foremost – a *causal relation between the two*. If Dennett is taken seriously, genes do cause their phenotypic effects, since they code for them, while memes, first, 'cause' and are represented in physical realizations of the meme (i.e., the DNA analogue), and, second, can mean (i.e., 'code for') something else. In other words, if a gene produces a phenotypic characteristic (an eye), the gene is certainly involved in producing this 'phenotypic effect' and we might say that the gene 'codes for' or 'is about' the eye. If I, however, produce outward behavior (expressing the moon-meme by saying the word 'moon'), the meme is certainly causally involved in producing these 'phenotypic effects.' In this sense the meme 'codes for' this *behavior* (me saying the word 'moon'). However, and this is very important, the meme is not 'about' that behavior (me saying the word 'moon'). It is about the moon, which, in turn, exists without any causal influence of the meme. This difference between genes and memes is totally ignored by Dennett.

To recapitulate, instead of being clear about what he means by 'phenotypic effects' of memes, Dennett uses the concept of phenotypic effects of genes (or memes) in (i), his claim that memes consist of information that have multiple physical substrates, with respect to the *causal* relation between genes (or memes) and the respective phenotypic effects, as did Dawkins and

Hull. In (ii), his claim that memes are semantic, he uses the concept of phenotypes in a different way, namely with respect to the supposed analogous *'semantic'* relation between genes (or memes) and their phenotypic effect. Dawkins and Hull only used the causal relation between genes and phenotypes for the gene-meme-analogy: In Dawkins' case, 'genes causally influencing the occurrence of phenotypes' equals 'memes causally influencing the occurrence of behavior and artifacts.' In Hull's case, 'genes causally influencing the occurrence of non-structure retaining phenotypes' equals 'memes causally influencing the occurrence of non-structure retaining behavior and artifacts.' Whether all this analogical reasoning from nature to culture makes sense at all has not yet been at issue in this descriptive section; however, to be clear about and to decide how to use the concept of 'phenotype' in the analogical reasoning from nature to culture is a prerequisite for any serious consideration of the ontological analogy.

If one cannot decide for what the concept of 'phenotypic effects' should be used, it is better to reject that the concept can be applied to culture properly. This is what Blackmore (1999: 62, 64-66; 2002: 715f) decides for, wrongly stating, however, that Hull and Dennett have the same position as she does. Nonetheless, I interpret Dennett to have at least intended the same as Blackmore, namely that the concept of phenotype is best not used at all for the gene-meme-analogy. If Dennett is taken in this way, his concept of memes includes that: (i) There are multiple substrates of memes, multiple DNA-analogues, since memes can be found in brain-patterns as well as books, behavior etc; (ii) memes are semantic units; (iii) there is no clear analogue to a phenotype that stands in the same relation to memes as biological phenotypes stand to genes, having a semantic relation that is at the same time also a causal relation between genes and their phenotypic effects, a causal relation of the sort specified above.

## Blackmore's ideational memes

Susan Blackmore is ambiguous as well. In her book *The Meme Machine* (1999), she defines memes as "instructions for carrying out behavior, stored in brains (or other objects) and passed on by imitation" (Blackmore 1999: 17).

Although explicitly denying that there is a clear distinction between genotypic and phenotypic aspects of memes (Blackmore 1999: 62, 64-66; 2002: 715f), she nonetheless uses the supposed semantic relation between genes and their phenotypic expression implicitly, for instance in the following statement: “It is tempting to consider memes as simply ‘ideas’, but more properly memes are a form of information. (Genes, too, are information: instructions, written in DNA, for building proteins)” (Blackmore 2000: 65). What ‘information’ is and in which sense genes are information *about* something is – as usual – not explained at all. A further employment of the concept of phenotypic effects can be found in the following: Blackmore criticizes Dennett for using the term ‘vehicle’ for artifacts such as a wheel. She wants to use the term ‘vehicle’ as restricted to a relation between genes and their phenotypes not highlighted so far: Genes are carried around by their vehicles. According to Blackmore, memes are not carried around by such artifacts as a wheel (Blackmore 1999: 65). This conflicts with Dennett’s claim that the wheel is a ‘meme vehicle.’ According to Blackmore, the ‘wheel’ is a product of the wheel-meme, but does not carry the meme. The background of this is that she distinguishes between ‘instructions’ and their ‘products.’ I regard this as similar to Hull’s distinction between structure-retaining physical substrates and non-structure-retaining phenotypes of memes. Products of memes, such as a wheel, do not retain the structure of the instruction, and they thus do not carry the meme.

I have analyzed the vagueness and various differences in the concept of memes used by the main defenders of the idea of memes as replicators. Neither Blackmore, who, at least, tries to illustrate the differences in the concepts of memes employed by the various memeticists (Blackmore 1999: 63-66), nor Dawkins, Hull, Dennett, or other memeticists, recognize that some of the differences are due to different assumed relations between genes and their phenotypes: causal, ‘semantic,’ or ‘carrying around,’ When one uses analogies one should be precise in which sense one uses them, otherwise analogies only lead to confusion.

### Aunger's neural memes

Aunger (2002) regards memes as brain-patterns and is thus closest to Dawkins meme concept.[14] However, he differs from Dawkins and the others in regarding the replicator – genes and memes – not as abstract entity (the type) that is merely realized in different physical substrates, but as lineage of material tokens. Recall what I have said in section 2.5 on replicators as types and as tokens. A singular gene token (i.e., a singular string of DNA) does not persist over cycles of replication. As I understand Dawkins, what persists according to him is an abstract entity – the type. But what persists can alternatively also be understood to be the lineage of tokens. If the replicator concept is understood in the latter sense, then Aunger is correct in saying the following: "Most definitions of memes are abstract, couched in terms of information or the mental representation that results from imitation. But replicators exist *as* specific substrates, as physical complexes" (Aunger 2002: 193; Emph. added). They are not only 'carried' or realized in physical substrates – they *are* these specific substrates. He is, however, not quite correct, since, if understood as material entities, replicators only exist as a *lineage* of physical substrates, as Hull has insisted (Hull 2001: 33).

The neuronal concept of memes points thus towards an issue that has not been in focus so far. How can we understand the ontological status of these abstract types? Blackmore (1999: 29) points to this as well, when she criticizes Popper for his concept of cultural units residing in a 'world 3' (Popper 1972). Popper is treated as a kind of forerunner of memetics, but criticized since he thought of abstract ideas in a 'world 3' that is distinct from both the mental and the material world. The problem is, she says, that it is unclear how these abstract entities can cause anything. Her answer to how memetics can get out of such a problem is clearly deficient. She merely says that "[I]n memetic terms, all that happens – whether in science or art – is selective imitation" (Blackmore 1999: 29), while it is unclear whether she is on Dawkins' or Aunger's side with respect to these issues and what Dawkins position would

---

[14] For a similar account see Delius (1991).

amount to in terms of a supposed ‘world 2,’ ‘world 3,’ or similar Platonic realms. Dennett (1995: 356-361) has a definite answer: Memes as semantic units are actually “intentional objects” that can be ascribed as being part of the natural world from an intentional stance (Dennett 1987). As such intentional objects they are real. What the ontological status of memes as abstract entities is will not be addressed further. I merely wanted to point out what memeticists themselves assume. I will neither defend nor criticize any answer to the question of the ontological status of these abstract entities. It would lead us too far away from the Darwinian analogies. My critique of the assumption that there are such abstracts units, presented at the end of section 3.3, will go in a different direction.

### Conclusion and further outlook

I chose Dawkins, Hull, Dennett, Blackmore and Aunger because they represent typical solutions that can also be found in other authors. To review all of them, in the detailed way needed to prevent confusion, would consume too much space. Before I end this section with a note on the relationship between persons and memes, I would like to summarize what we have found so far.

There is a purely *neuronal concept of memes*, regarding memes as tokens of brain-patterns that replicate and get thereby inherited from brain to brain (Aunger). There is an *ideational concept of memes*, identifying memes with abstract types of mental content, often called ‘information,’ as genes are often regarded as abstract types of information coded in DNA-tokens (Dawkins, Dennett, Hull, Blackmore). However, within this group only Dawkins regards memes as having a single specific physical substrate, analogous to genes having a single physical substrate in DNA. This substrate is the brain. Everything else is equivalent to phenotypic effects of memes, somehow ‘caused’ by these memes. Dawkins thereby employs an analogy between the causal relation between genes and their phenotypes as basis for his concept of memes. According to Hull, Dennett and Blackmore, physical meme-tokens exist in brains, books, spoken language, and the like. Hull clearly distinguishes these physical meme-tokens from their non-structure retaining phenotypic consequences, which are nonetheless caused by memes (also employing thereby the causal relation between genes and their phenotypes as basis for this terminological decision). Dennett regards all kinds of things as physical substrates ('vehicles') of memes, which are consequences of these memes and carrying these around, employing thereby implicitly a kind of causal relation between memes and their phenotypes, and a further carrying-relation as basis for his description of memes and their existence in the material world. However, he mainly uses a semantic relation between genes and their phenotypic effects, to distinguish between memes themselves and what they are about. Blackmore also treats all kinds of material entities as physical meme-tokens, but does not want to use the concept of 'phenotypes.'

In addition to the neural and ideational meme concept, there are revisionists, defining memes as outward behavior or in terms of the consequences of behavior (artifacts). As mentioned already, with this *behavioristic meme concept*,[15] they react to the problem that ideational memes cannot be identified easily (identification problems). Although they can react to the identification problems, as I will argue after I have discussed these problems in section 3.3, what is left over from the gene-meme-analogy is almost trivial.

Before I proceed to the identification problems, I want to add a note on how memeticists construct the relation between memes and human individuals. What I have said in this descriptive section relates to old philosophical and anthropological questions. What is the ontological status of ideas, mental content etc.? Are they essences or not? How do they relate to brain-patterns and behavior? I have clarified what memeticists say on these issues: what memes are, in which relation memes stand to brain-patterns, behavior and different products of behavior. However, in addition to these relations, memes also have a relation to the persons having these memes in their minds. What is the assumed relation between memes and people? This last question creates a link to the explanatory units of selection analogy, since Dawkins, Dennett and Blackmore assume that a human minds are mere 'vehicles' or 'hosts' of

---

[15] A term I borrowed from Blackmore (2002).

memes: built by memes and carrying memes around with them. A human person is only a consequence of memes and thus derivative for explaining the change in frequency of memes. This is employing the causal *and* the carrying-relation between genes and phenotypes at the same time, but now with respect to whole persons. Anthropologists used to express a similar point in saying that humans are 'culture bearers.' Yet, one can question whether this is the only role humans play in culture. As mentioned in section 3.1, the anthropological concept assumes that a person is also the creator and selector of the cultural items a person bears around with him. I will address the relationship between memes and minds in chapter 5. The issue is negligible for the ontological analogy between genes and memes, since the ontological analogy is independent of the explanatory units of selection analogy in the following sense: Even if the units of selection analogy fails and minds are not reducible to their role of 'carrying around' memes, memes may still be replicators in the narrow sense. In this chapter, I will explain that whether they are replicators depends on the other relations discussed in this section: the relation between memes and brain-patterns, memes and behavior, memes and artifacts.

That memes are replicators has indeed been criticized heavily. In the latest reply to critics, Dawkins mentions three "favourite objections" to the gene-meme-analogy: (i) "the vexed question of how large a unit deserves the name 'meme';" that (ii) "nobody really knows what a meme physically is;" that (iii) "memes have insufficient copy fidelity," and therefore do not qualify as replicators (Dawkins 1999: xiv).[16] Although (i) and (ii) involve a more complicated problem of identification, and although (iii) involves more then the question of copy-fidelity, these objections mirror the essential problems the ontological analogy faces: (i) and (ii) belong to what I have above called *identification problems*; (iii) belongs to what I call the *replication problems*.

---

[16] Blackmore (1999: 53-62) has a similar listing.

## 3.3 IDENTIFICATION PROBLEMS

### Traceability condition and identification

If memes should be cultural units of heredity that can be counted in order to track cultural change, then they should be easy to identify. This has been doubted on the base of a couple of objections. The debate about meme's identification is messy, since different problems of identification are involved. Memes are doubted to exist, because we do not know what they are made of, or because they can only be abstracted from other entities. It is objected that we cannot single out discrete, particulate memes as we can single out and identify discrete, particulate single genes. The fronts of the debate are harsh: Memeticists often accuse critics of not knowing enough about the concept of genes, genetics ,and its history;[17] and critics in turn accuse memeticists for not knowing enough about culture.[18] Both do this without clearly distinguishing between different problems of identification.[19] I will present a systematic overview and single out three distinct problems of identification: the *boundary problem*, the *holism problem*, and the *material identification problem*. The boundary problem is about finding the length of single memes: i.e. about how to partition the whole meme-complex into single memes. The holism problem is about finding a context-independent effect/meaning of single memes. The material identification problem is about identifying memes in material substrates. I will say less on the first two problems, since others have treated these in a convincing way. This is different for the material identification problem. I will analyze this third problem in detail. By specifying in which sense there are disanalogies between genes and memes, and in which sense

---

[17] See for instance Hull (2000), Blute (2005).

[18] See for instance Kuper (2000), Bloch (2000).

[19] Blackmore (1999: 53ff), for instance, confuses the problem that genes can be defined in various ways (as evolutionary units being replicators, as molecular units coding for one protein and as functional units, as 'genes for' a phenotypic trait) and the alleged analogous problem for memes, with the problem of finding the boundaries of the gene or meme as replicator, a problem that appears only with respect to the evolutionary gene concept. The question how large a gene-as-replicator is different from the question what a gene is in conceptual terms: the replicator, the cistron, or a functional unit.

there are not, I offer a new way through an insufficiently structured debate about the gene-meme-analogy.

## Boundary problem

A first kind of objection, which has been addressed by memeticists from the start, states that we cannot define the boundaries of single memes, whereas we can easily do this with single genes. After introducing memes, Dennett writes:

> "Intuitively these [memes, MK] are more or less identifiable cultural units, but we can say something more precise about how we draw the boundaries – about why D-F#-A isn't a unit, and the theme from the slow movement of Beethoven's Seventh Symphony is: the units are the smallest elements that replicate themselves with reliability and fecundity" (Dennett 1991: 201).

If the meme is defined as a replicator, it is identified as the gene-as-replicator by a relative measure: as the smallest replicable unit of a complex that can survive transmission undivided quite long. For the example of a symphony, the singular meme would have to be "sufficiently distinctive and memorable to be abstracted from the context of the whole symphony" (Dawkins 1989 [1976]: 195). What a sufficiently distinctive and memorable unit of a symphony is depends on the respective symphony and is sometimes surely not easy to answer. The splitting up of a whole into parts is even harder if we look at other cultural items, for instance, a theory. Using Darwinism as an example, Dawkins says that the boundary problem can in principle be solved in the following way:

> "If Darwin's theory can be subdivided into components, such that some people believe component A but not component B, while others believe B but not A, then A and B should be regarded as separate memes. If almost everybody who believes in A also believes in B – if the memes are closely 'linked' to use the genetic term – then it is convenient to lump them together as one meme" (Dawkins 1989 [1976]: 196).

Believing A and not B is treated as analogous to segregation in biological inheritance: Segregation means that a certain segment of the chromosome of one parent is inherited independent of the inheritance of another segment.

It is indeed not easy to find the boundaries of a single meme, i.e., to identify a single meme out of a complex of memes. However, as just indicated, the same problem holds for genes. As shown in section 2.5, if a gene-token is not defined as that string of DNA that codes for one protein (molecular gene concept), but as that string of DNA that is very likely to survive long enough through not being divided in reproduction (evolutionary gene concept), then genes have the same boundary problem. It is not easy to say which string of DNA out of the genome is the evolutionary gene (i.e., the replicator). The boundaries can only be defined statistically or ex post facto, after the survival of a string of DNA over a significant time-span.

Therefore, although the boundary problem exists, it cannot serve as an argument against the analogy between genes and memes as replicators. As Hull writes the "definition of evolutionary genes is just as difficult to apply as is its memetic correlate. In general, critics of memetics assume standards so high for scientific knowledge that few, if any, areas of science can possible meet them" (Hull 2000: 48). If the rough-and-ready definition of replicators is accepted for genes, which faces equal boundary problems, than it should be good enough for memes.[20]

### Holism problem

Another objection against the gene-meme-analogy is that memes cannot be regarded as particulate isolates, as independent entities so to speak, since memes have their meaning only in the context of other memes (Bloch 2000, Kuper 2000). As I also illustrated in section 2.5, the same *holism problem* exists for genes. Whether genes have a context-independent phenotypic effect depends on how one looks at the relation between genes themselves, between genes and phenotypes, and between genes and the environment. Be it as it may, the context-independent effects of genes are not easily found out. Sober (1992) addressed this issue with respect to the critique that memetics and other Darwinian approaches to cultural change 'atomize cultural characteristics.' He concludes that the same problem occurs for biological evolution of distinct traits based on distinct genes:

> "Having two children rather than five, or being a kamikaze pilot, are characteristics that are abstracted from a rich and interconnected network of traits. The worry is that by singling out these traits for treatment, we are

---

[20] Similar answers can be found in other defenses of the gene-meme-analogy: Dawkins (1989 [1976]: 195, 1999: xiv); Dennett (1995: 344), Blackmore (1999: 53-55); see also Laland & Brown (2002: 225-226) or Gil-White (forthcoming) for the same line of argument.

> losing sight of the context that gives them cultural meaning. It is worth mentioned that precisely the same question has been raised about various models in genetic evolution itself. If you wish to understand the population frequency of sickle cell anaemia, for example, you cannot ignore the fact that the trait is correlated with resistance to malaria. In both cultural and genetic evolution it is a mistake to think that each trait evolved independently of all the others. Of course, the lesson to be drawn from this is not that one should not atomize characteristics, but rather that the atoms one identifies should be understood in terms of their relationship to other atoms" (Sober 1992: 31).

In this sense, the gene as a functional unit is definitely context-dependent, as is a meme. Bloch (2000) criticizes memeticists for wrongly assuming that memes are "discrete," "distinguishable," having a "defined existence." At one place, he writes:

> "At first, some [memes, MK] seem convincing as discrete units: catchy tunes, folk tales, the taboo on shaving among Sikhs, Pythagoras' theorem, etc. However, on closer observation, even these more obvious 'units' lose their *boundaries*. Is it the whole tune or only a part of it which is the meme? The Sikh taboo is *meaningless* unless it is seen as part of Sikh religion and identity. Pythagoras' theorem is a part of geometry and could be divided into smaller units such as the concept of triangle, angle, equivalence, etc." (Bloch 2000: 194; Emph. added).

He definitely mixes up the boundary problem with the holism problem. The Sikh taboo is definitely a unit that can exist and can be isolated from the whole, even if its meaning is dependent on the whole. The same situation holds for genes. A certain part of a chromosome that can be singled out as a replicator, as an evolutionary gene, does have its effect on the phenotype (i.e., its 'meaning' as Dennett would call it), only in context of all the other genes of the respective organisms. Identifiable units might well be dependent in their effects (or meaning) and still be identifiable as single units. Bloch (2000: 197ff) has a point, when claiming that each meme that enters a culture (or a single mind) is made consistent with the other memes in this culture (or mind) and that its meaning changes through this integration and is therefore context-dependent. But he ignores that there is nonetheless *something* that is made consistent. Therefore, it is important to distinguish between the claim that there are no memes as distinguishable units (i.e., that their boundaries are often unclear) and the claim that their meaning is context-dependent. One should also distinguish these two problems from a further one, which I call the *material identification problem*. Nonetheless, although different, the latter will lead us back to the former.

## Material identification problem

Although both, genes as well as memes, face a boundary problem and a holism problem, there is indeed a difference between genes and memes with respect to their identification. It is usually ignored that the evolutionary gene concept presupposes that genes are indeed identifiable, namely as bits of DNA, even though it is – at the same time – admitted that genes face a boundary and holism problem. The evolutionary gene, the gene-as-replicator, is defined as a type of similar strings of DNA tokens, or as a lineage of DNA tokens that are not likely to be divided during reproduction.[21] Through looking at DNA, we identify a replicator and say that these bits of DNA over there are the same as the bits of DNA over here, proving thereby that there is similarity between the two bits and proving thereby that they are tokens of the same type. The same holds for the gene as a functional unit – those bits of the genome that are involved in the expression of a single protein or phenotypic characteristic. In a nutshell, *genes do have a definite material substrate that allows to identify them and to count them*. And that is, first and foremost, why they are at all considered to be the 'bookkeepers' of biological evolution. We can identify them because genes are materially realized exclusively in DNA, and are always 'coded' in DNA in the same way. If the genes are different, then the DNA is different; if the genes are the same, the DNA will be the same. This is a one-to-one relationship between the abstract, informational gene and its material realization. There is one material substrate with one way to 'code' the genes in this material substrate.

Memes, however, are doubted to have such a universal material substrate that allows to identify and to count them in the same way. Note that we not only need a DNA-analogue; we need a DNA-analogue *that allows to identify memes*. This has been contentious and is what I call the *material identification problem*. In contrast to genes, memes seem to be subject to this problem.

---

[21] Harms (1996: 359) makes a similar point.

*(i). Memes and identification through brain-patterns.* As Dawkins writes, "[m]emes have not yet found their Watson and Crick; they even lack their Mendel" (Dawkins 1999: xii). If we regard brain-patterns as the exclusive material substrate of memes, it stands to question whether we can infer the same memes from looking at two brain-patterns in two humans. Although this is part of the philosophical mind-body problem, an issue that definitely has no rough-and-ready answer, nobody involved in debates about memetics believes that we can find for a given idea the same brain-pattern in different heads of people. Imagine two people and imagine that we can convincingly show that they believe in exactly the same meme, namely the proposition that the apple in front of their faces is red and not blue. Now, could we ever find out that they believe in the same proposition by looking at their brain-patterns? Not likely. The first person's belief might be stored or 'coded' in a totally different brain-pattern than the second person's belief. Hence, we clearly cannot identify memes through their alleged material substrate in brains. One and the same meme, adopted or held by different people, can be correlated with diverse brain-patterns in these people's heads; and, *vice versa*, one and the same brain pattern of different people's minds can be correlated with diverse memes. This *many-many-relationship between brain-patterns and memes* is usually admitted by critics as well as defenders.[22] It is beyond doubt that we cannot observe memes through looking at brain-patterns. Scholars involved in debates about memetics do not differ on the fact that mental contents do not map on to a definite, universal brain pattern in all kinds of individuals. But, and this is important, they differ on what *follows* from this for the gene-meme-analogy.

*(ii). Memes identified through other obersvables.* Dawkins, for instance, answers that, although there is a disanalogy with respect to a definite one-to-one-mapping of memes in brain-patterns, which makes identification of memes in brain-patterns impossible, memes can nonetheless be identified and counted through their phenotypic effects, or – as other memeticists would say – through other material substrates, such as written words. After admitting that memes

---

[22] See, for instance, Dawkins (1982a: 109, 1999: xii), Dennett (1995: 352-354), Gatherer (1998), Wilkins (1999).

cannot be identified and counted through brain-patterns, Dawkins writes: “As with genes, we track memes through populations by their phenotypes” (Dawkins 1999: xii). The problem is that the relation between genes (or memes) and their respective phenotypic expression is *also* characterized by a many-many relationship.

Let me explain this first with respect to genes: In the past it might have been considered by some researchers as a good way to track genes by tracking phenotypic characteristics, however, today it is not considered anymore as a reliable way to track genes, partly because of the holism problem (introduced in section 2.5 and mentioned above). Furthermore, not only the holism of gene interaction poses a problem. In principle, one particular phenotypic characteristic can be correlated with diverse genes, and one gene can be correlated with diverse phenotypic characteristic. One phenotypic characteristic can be correlated with diverse genes, since, for instance, an eye might be caused by that string of DNA in one species and by a different string of DNA in another species. Even within a species, it is principally possible that two different genes cause the same phenotypic effect. This is because the expression of genes is not only influenced by the presence of other genes and the interaction with these. The expression is also influenced by environmental factors: A gene has a *norm of reaction*, a spectrum of phenotypic effects that vary with the context of their environment. In one case, the phenotypic trait might be causally influenced by certain bits of DNA having a reaction norm that comprises this phenotypic effect, given a certain environmental interaction. In another case, it can well be caused by different bits of DNA having a different reaction norm, but one that also comprises the respective phenotypic effect, given a certain environmental interaction. The reaction norms can in principle overlap. Therefore, different genes can *in principle* cause the same phenotypic effect. In turn, given the same gene, different phenotypic traits can result, precisely because genes do not code for a specific trait but have reaction norms, a range of phenotypic effects. Which phenotypic effect results in the end is dependent on the environment in which a gene is expressed. In short, the inference from phenotypic traits to certain genes has to be made very carefully. It is an empirical question whether a certain phenotypic characteristic is always (within and between species) causally influenced by the same gene.

However, and this is very important, *in principle* we can nonetheless find out whether one phenotypic characteristic is correlated with the same gene, across individuals and across species. We simply need to *double-count*: phenotypic characteristic and the DNA and look at the causal processes producing the phenotypic characteristic. To find out which gene or genes cause a certain phenotypic effect, given that we can identify genes as molecular structures independently of their phenotypic effect, is still difficult, mainly because of the holism problem. However, to find out which genes cause a certain phenotypic effect *without* direct access to genes is terribly hard.

And there's the rub: double-counting is precisely what we *cannot* do in the case of memes. I will offer two kinds of examples of memes that show that there is a many-many-relationship between memes and their alleged 'phenotypic' effect. I will then claim that because of the lack of access to something analogous to DNA, the identification problem cannot be solved in the case of memes, whereas, as just stated, it can in principle be solved in the case of genes. First, as Boyd & Richerson (2000: 155) write: "For any phenotypic performance there are potentially an infinite number of rules that would generate that performance." In this case, rules would be the memes and a certain behavior the 'phenotypic performance,' produced by these memes. Their convincing example is taken from the generativist model of phonological change. Pronunciation is governed by complex rules. If pronunciation changes in adults, they simply add a rule at the end of a chain of existing rules, adjusting their behavior through this procedure. If children, however, learn a pronunciation they simply "induce the simplest set of grammatical rules that will account for the performances they hear" (*ibid.*: 156). Let us look at their example. "In some dialects of English, people pronounce words that begin with *wh* using what linguists call an 'unvoiced' sound while they pronounce words beginning with *w* using a voiced sound. […] Now suppose that people who speak such a dialect come into contact with other people who only use the voiced *w* sound" (*ibid.*: 156). If the people now totally adopt the pronunciation of the other group, then we end up with the following situation: After adults learned to get rid of the unvoiced sounds, children will never hear it and will induce from observation the simplest set of rules, namely one that does not distinguish between the pronunciation of words with 'wh' (as in whether) and those with 'w' (as in 'weather'). Adults still maintain rules for the difference and, in addition, a rule for pronouncing the former nonetheless like the latter. Although "there is no difference in the phenotypic performance among parents and children, children do not acquire the same mental representation as their parents" (*ibid.*: 156). In this example, a given 'phenotypic trait' is then due to different 'memes.'

If we look at a second kind of memes, memes that are not cognitive instructions for a behavioral pattern (such as the pronunciation of a word), the same problem occurs. The meaning of the word 'moon,' the 'phenotype' of the alleged moon-meme, may vary from person to person. Weingart et al (1997: 301-312) called this the *meaning problem*. How do we find out about the 'meaning' of memes? Memes are not physically observable, neither as ideational units, nor as brain-patterns. We therefore track them through behavior and artifacts. However, there is not always a stable one-to-one mapping between behavior and meaning of memes. Memetics thus faces "the fundamental problem in the social sciences of relating attitudes to behavior" (Weingart et al 1997: 309).

If this problem exists, than it is not only the case that we cannot identify memes and their meaning through brain-patterns. It is also the case that we cannot identify them easily through behavior or artifacts, which are either 'physical substrates' or material 'phenotypic expressions' of the meme. We have to *infer* the memes from the observables of culture and this involves interpretation. The problem is that the inference from similar behavior to similar memes is precarious. Examples where such an inference from similar behavior or artifacts to similar meaning is dubious are abundant. As Bloch stresses, even if Italians imported the behavior of making and eating pasta from the Chinese, it does not necessarily mean the same for them. What 'pasta' means for Italians includes, for instance, the reasons for the maintenance of the behavior of making and eating pasta, as he claims. And these reasons include facts about other “beliefs, symbolism, economy, agriculture and perhaps family organization” (Bloch 2000: 198). The Italian concept of ‘pasta’ is different from the Chinese one, although both make and eat pasta. The same differences in meaning has been shown for basic shared social norms even within a culture, for instance for the meaning of the Ten Commandments in contemporary America (Atran 2001).

Note that it is not enough to cite other examples where the inference is quite safe. Such examples certainly exist and we use them in our everyday inferences. Take the above-mentioned example of two persons asserting that they see a red apple in front of them. We believe that if the two persons do or say similar things, then they must also believe in similar things. We do this because we further assume that they speak the same language and have the same background beliefs. Thus, when I say ‘red’ and you say it as well, it is rather safe to conclude that we mean approximately the same, given that we share enough other memes, for instance, the same belief in the natural causes of colors and so on. Checking enough other beliefs can test this. However, the meaning problem reminds us that we have to be careful with inferences from similar behavior to similar memes. We have to be as careful as biologists, when they infer a ‘gene for’ from observable facts of phenotypes without direct access to the genes themselves.

Now, the meaning problem sounds quite similar to the holism problem mentioned above: A meme has its meaning only in context of other memes that influence what I mean when I say ‘moon.’ Indeed the meaning problem is the holism problem. However, the reason why the meaning problem is a problem for memetics is not the holistic aspect of meaning as such, which holds for genes as well. Although I think that the just mentioned authors are correct in claiming that there is a many-many-relationship between memes and the respective behavior or artifacts, they ignore the following complication: The important point is not that one behavioral characteristic (for instance, me using the word ‘moon’ and somebody else doing the same) *can be* due to different memes, since the analogous problem exists for genes, as shown above. The important point is that it is *so hard to find out whether a specific behavior is due to a different meme or due to the same meme*, whereas we can in principle find it out for genes, since we have, besides looking at the phenotypic characteristic, a *second* access through directly looking at DNA. The problem is that we do not have an access to memes that is *independent* of their phenotypic behavioral effects – an access comparable to the access we have to genes due to our access to DNA.[23] Only the material identification problem shows why the holism/ meaning problem is a problem for the gene-meme-analogy.

Nonetheless, memeticists could choose a last strategy: They can argue, that in the past we did also not have such an access to DNA. Memetics is still in its infancy that should be judged according to this infancy-status. Dennett (1995: 344) suggests, as Dawkins did, that we can identify genes via their "uniformity of the phenotypic effects," and that this was the method used to identify them before we discovered DNA. From this he concludes that there is neither an identification problem for genes nor for memes. As I said with respect to the similar claim of Dawkins, the latter simply ignores all the complexities of the causal effects of genes and the analogous problem for memes. As Weingart et al write, the situation with memes

> "is strikingly similar to the early and ill-fated strategy of certain camps of genetics who, driven by enthusiastic speculation, identified all kinds of common sense attributes of people, especially racial ones, as 'traits' whose genetic determination was postulated by them and then searched for in vain by way of inference from genealogies. The crucial task of establishing 'heritability' is already a highly complex endeavor on the level of human organisms, and even more so on the level of cognitive entities. Evolutionists and anthropologists have to beware so as not to make the same mistakes as the eugenicists" (Weingart et al 1997: 313).

---

[23] Hull (2000: 60) replies in a similar way to Boyd & Richerson: For genes and memes there is a many-many-relationship between them and their 'phenotypic' effects. However, he concludes that we therefore do not have an identification problem. My position differs from Hull in claiming that although there is an analogous many-many-relationship between genes and their alleged 'phenotypic effects,' there is still an important disanalogy. In the case of genes, we can double-check: Although in principle one phenotypic characteristic *could* be due to diverse genes, we can find out which gene *does* in fact produce it.

There definitely was an identification problem for genes, but DNA now provides a rather good way out of it. The claim that genetics had the same problem before the discovery of DNA (and the implicit claim that memetics will solve it one day as genetics did), does not help. If it is correct, as Dennett concedes, that there is no one-to-one mapping between memes and brain pattern, we will *never* have an independent access to memes. In addition, before the discovery of DNA, a one-to-one mapping between genes and phenotypes was at least assumed as a viable hypothesis for genetics. Not even this is the case for the analogous hypothesis in memetics. One wonders where the confidence of memeticists about the future of their new science comes from.

To conclude: The wonderful thing about genes is that there is a one-to-one mapping between physical substrate and abstract gene, as described above. In addition we have a direct access to that physical substrate, independent of the phenotypic effect of genes. We can double-count genes and phenotypes. This allows a clear identification of genes. For memes the case is different. There is no one-to-one mapping between memes and their phenotypic expression. We do not have an access to memes that is independent of the phenotypic expression of these memes. We cannot double-count memes and their phenotypic expression. Indeed, we cannot count them at all, we can only *infer* memes from the observable behavior and artifacts and *interpret* these observables as representations of memes. Therefore, compared to genes, we do not have the same possibility of a clear identification. This is a disanalogy between genes and memes: Memes have a material identification problem, while genes do not. I turn now to a detailed description why this disanalogy is important for the evaluation of the ontological analogy.

## Consequences for the ontological analogy

First of all, as Dennett (1995: 353f) says, memetics will certainly not be a science in the sense as genetics is, since this would require that neuronal mind-reading is possible, as DNA-reading is possible.

*(i). The traceability condition and memetics as a science*. If this is the case, as admitted by almost everybody in the debates about memes, then

memetics will not be a science comparable to natural sciences such as biology. However, ignored by Dennett and others, memetics will not really be a science for a further reason that follows from the first: memetics relies on unobservable ideational units, which can only be inferred, which involves interpretation of what a certain behavior (speaking out 'moon,' for instance) means for the person engaged in the behavior. Therefore, my first main conclusion is that memes fail to fulfill the *traceability condition*, which is evidently fulfilled by genes. The ontological analogy is *wrong* with respect to this central aspect of the analogy.

This has further consequences for the status of memetics in midst other sciences. Since the attribution of unobservable memes does involve *interpretation*, and since this is usually considered to be the watershed between hard natural and 'soft' social sciences or humanities, memetics will have to take its place in midst of the humanities and social sciences. Despite being an application of evolutionary theory, memetics is not a new naturalistic frame that can give the latter disciplines some new 'scientific' tools of thinking. On the other hand, this simply means that memeticists are in good company with 'soft' scientists, who also try to deal with the meaning problem since a long time. The problem of identification is not only a problem for memetics. It is the standard problem of social sciences and humanities. Therefore, it would be unfair to judge memetics harder than any other theory of culture, which inevitably faces the same problem. The theoretical postulates about ideational units of heredity are at the same level as the assumptions of other schools of thought in the social sciences, such as anthropology. They assume the same ideational concept of culture and end up with the same problems. My critique is not that the gene-meme-analogy uses these assumptions and has these problems. My critique is that memetics is *heuristically trivial*, not adding anything new and ending up with the same meaning problem as any social science relying on an ideational ontology. In addition, the meaning problem and the material identification problem affects the memeticists more severely than other scholars who also assume ideational units. While social scientists usually know about the meaning problem and try to find methods to secure that the inferences from a certain behavior to a specific mental state are legitimate, memetics seem to lack any recognition of the complexities that are involved in such inferences.

*(ii). The similarity requirement for replicators*. Although memeticists may already be not amused already, I want to make an additional point – a point that leads to a more far-reaching critique of memetics. Can the material identification problem serve as critique that affects memetics in particular but not other theories that try to understand culture? I have claimed that genetics differs from memetics in having the chance to double-check DNA and phenotypes. We cannot do that in memetics.

This is an *essential difference* between genes and memes that harms the ontological analogy between genes and memes in terms of descriptive adequacy and explanatory force in a way not mentioned so far: Because of the meaning problem and the material identification problem, it also becomes questionable whether memes fulfill the *replicator condition*. If the traceability condition is not fulfilled, than memes may fail to meet the first replicator requirement (introduced in section 2.5), namely the *similarity requirement*, which secures the longevity of replicators. For the neuronal meme it is safe to say that it is not replicated, since each meme is very likely to be coded in a *different* brain pattern. If memes are these brain-patterns, they do not replicate.[24] For the ideational meme the same *can* hold, since a certain behavior, exhibited by two different individuals, is not necessarily accompanied by the same mental content. Therefore, the learning of a certain behavior might well not be a replication *of* memes, since memes might not have been copied. The question whether a given meme is a replicator or not depends thus on each individual case. And the answer to this question is hard

---

[24] Since the neural pattern does not replicate, Maynard-Smith & Szathmary (1999) and – in a more systematic way – Szathmary (1999: 5-8) define the meme as a "phenotypic replicator." They maintain that the meme is a replicator that fulfils the similarity requirement, but they understand the meme as the effect of its 'genotype,' which is specified as the corresponding neuronal brain pattern: "Genes specify structures or behaviours – that is, phenotypes – during development: in inheritance, the phenotype dies and only the genotype is transmitted. The transmission of memes is quite different. A meme is in effect a phenotype: the analogue of the genotype is the neural structure in the brain that specifies the meme" (Maynard-Smith & Szathmary 1999: 140).

to find out. It depends on whether the mental content is also copied, when a behavioral pattern or an artifact is copied. It therefore depends on a one-to-one mapping between memes and their phenotypic expressions that holds across individuals. Genes, on the contrary, are replicators, even if there is no such one-to-one mapping between them and their phenotypic effect. That one biological phenotype can in principle be due to different genes, does not imply that – if a given phenotype does reproduce by biological inheritance – its genes can fail to replicate. That a certain behavior can be due to different memes, however, implies that the meme can fail to survive when the respective phenotype of the meme (behavior or artifact) is transmitted from person to person.

That is why the material identification problem refers to a very important disanalogy: Memes can fail to meet the similarity requirement because of the complex relationship between ideational units and their consequences in behavior and artifacts. This is my second main conclusion with respect to the identification problems. The claim that memes are replicators is in danger of being *false* for all cases where memes cannot safely be inferred from similar behavior, since in such cases the similarity requirement is in danger of not being fulfilled. Whether the claim about memes as replicators is false for a specific meme is an empirical question that is difficult to answer. In addition, since memetics cannot offer memes as identifiable units that would make answering the question easier, the ontological gene-meme-analogy is *heuristically trivial*.

### Something-is-preserved-arguments

I will now turn to two tricky defense arguments against my critique about the traceability and replicator condition. I call them '*something-is-preserved-arguments.*' Against my last objection, memeticists can counter that something evidently is preserved in social learning and this something shows – if preserved – necessarily similarity across persons. Although the meaning problem seems to be plausible, there evidently is *something* shared between cultures and in cultures, even in cases where the meaning of a meme varies across contexts. As Mark Jeffreys writes: "Clearly, information is preserved across minds, even across cultures and centuries […] *something* is preserved" (Jeffreys 2000: 233; Emph. in the orig.). Now, the essential question is what this "something" is. The behavior that can be correlated with different memes certainly is transmitted and maybe changed later on. Take again the pasta example. Italians imported from Chinese the habit to make and eat pasta. If Bloch (2000) is right, then the cultural meaning attached to pasta-eating was not transmitted thereby. In order to reject the meaning problem, memeticists merely need to say that the meaning, which is attached to pasta-eating by some and not others, is *not part of the meme*. There are two ways to do that. The first option comes from scholars that suggest to define the meme as a behavioral instead of an ideational unit. The second comes from within the ideational concept of memes.

*(i). Observable units as preserved.* The meaning problem has lead some memeticists to define memes not as ideational unit but as observable units, i.e. as behavior or artifacts.[25] These units can be identified and counted; and one can decide whether they are shared between peoples or not. But the problem with this solution is that we would then end up with memes that are not the basic building block of culture. Just like evolutionary psychology we would end up with a concept of culture that ignores the *generative* building blocks of culture. As argued for in section 3.1, these are ideational units. Gatherer (1998), for instance, opts for the meme as a behavioral unit. In addition, he says that only behavioral units are cultural, while cognitive units are not. He thereby creates a dualism between mind and culture, totally separating culture from mind. This would be a step backwards – to a concept of culture that ignores that behavior and artifacts are mere consequences of ideational units. Although we might then have something that is observable, that can be counted, and whose similarity can therefore be judged much more easily, we end up with counting things that might not represent culture in its full generative sense. In a nutshell, we might thereby get countable replicators, but lose our target, namely culture. As Weingart et al (1997) have put it, to treat

---

[25] For instance: Gatherer (1998), Benzon (1996), Deacon (1999).

behavioral patterns as the basic units of culture, would be “a mistake roughly tantamount to treating genes as traits and ignoring pleiotropy, dominance, and other developmental complexities of the geno-phenotype connection“ (Weingart et al 1997: 301). Hence, defining memes in a behavioral way will not help to get rid of the meaning problem.

Therefore, my third main conclusion with respect to the identification problem is: By defining memes as behavioral units, the claim that memes are replicators would become *true*, but would nonetheless loose a lot of descriptive adequacy, since it would leave out an essential part of culture. Through this, the ontological analogy between genes and memes would also loose its explanatory force, since an explanation of culture necessarily requires to take the generative dimension of culture into account. In addition, the analogy would then fall short of the level at which an explanation is thought for culture by others, who take the ideational dimension of culture into account. Through a ‘behaviorist’ something-is-preserved argument, we would restore the analogy, but only by paying the price of triviality.

*(ii). Ideational essences as preserved.* However, defenders of the ideational concept of memes can react with a different ‘something-is-preserved-argument’: They can state that there must be something that is *cognitively* shared. Dennett, for instance, writes that we can identify the meme as a semantic property through looking at what is common between different physical vehicles. We can identify the meme that lies beyond an English, German or French translation of ‘West Side Story.’ We can identify it as “the story, not the text” (Dennett 1995: 356). Although it is correct that we often can *abstract* a common core of different observable units of culture, he still misses the point of the meaning problem, since what the story ‘means’ to this or that person can still be different at the cognitive level. My inner understanding of ‘West Side Story’ is certainly different from yours. Dawkins made a similar step, but directly referring to the cognitive level, when he referred to an essence of Darwinism, the Darwinism-meme, that is shared between scientists, even if these scientists all have different interpretations of

Darwinism in their mind.[26] We infer these different cognitive variants (i.e., variants in different people's minds) from what these people do, write and say. We then abstract a common core. The problem is that to discover this essence is a task that is even more complicated than just to infer the different variants from the observable differences in behavior. It can even be doubted and has been doubted that there is such an essence of theories and similar entities. As I already mentioned at the end of section 3.2, I cannot answer here whether there are such abstract 'essences' of cultural units and what their ontological status would be.

What I wanted to object to the gene-meme-analogy with respect to the described identification problem and the ideational something-is-preserved argument is more restricted. If memeticists merely retreat to 'essential' memes, then the ontological analogy is heuristically trivial: In such a case, the analogy would not add anything to the common assumptions about culture, and would not contribute to the solutions of the problems arising from these assumptions. Memeticists state that there must be something ideational that has to be abstracted from the behavior and artifacts that make up the observables of culture. This 'something' is assumed to be shared and maintained between people despite different interpretations of a text, despite different cultural meanings of a certain cultural habit. They thus state that there are basic building blocks of culture. As described already, this simply mirrors the contemporary ideational concept of culture. The ontological status of this 'something,' however, is not made clear through superimposing the gene-meme-analogy. The analogy does thus not have a heuristic value in this sense. On the contrary, the postulate of memes ends up with the same problems any theory of culture has to face and cannot offer any new insights. Philosophy of language, psychology, and anthropology have developed diverse theories that try to solve the problem. Memeticists are often ignorant of these theories. But even if they were not ignorant of them, they would still merely reinvent the wheel. As Bloch (2000: 191) writes, what memeticists do when they announce that they have found the basic building blocks of culture and call them memes,

---

[26] See the quotation at the beginning of section 3.2.

can be interpreted as similar to a sociologist announcing in the year 2000 that he has discovered that there are genes. This is my fourth concluding claim against the merits of the ontological analogy.

### Conclusion

My arguments against the gene-meme-analogy with respect to the identification problems can be summarized in the following way: Although the meaning problem is a common problem of all theories dealing with culture, it shows two central disanalogies: that (i) the traceability condition is definitely not fulfilled because we cannot double-count memes and their phenotypic expression; that (ii) the similarity requirement for replicators may not be fulfilled in some cases of memes, whereas it is evidently fulfilled in the case of genes. If we then try to circumvent this by moving back to a behavioral concept of memes, then (iii) the analogy can be restored but only for the price of trivialization, since a behavioral concept of memes ignores the generative ideational dimension of culture. These are the first three of my concluding claims about memes and their identification. My fourth concluding claim for this section is: (iv) Even if the generative dimension is included, memetics mirrors the standard assumption of the existence of basic ideational units of culture and has not managed to add anything but confusion to the *solution* of the traditional identification problem that arises from the meaning/holism problem.

## 3.4 REPLICATION PROBLEMS

### Replicator condition

So far, we have seen that the ontological analogy rests on three assumptions: (1) that memes are basic ideational units of culture; (2) that they can be tracked; (3) that they are replicators. The material identification problem has not only shown that (1) is trivial, and that (2) is false, but also that (3) is in danger of not being fulfilled. The latter is connected to the following further objection against the ontological gene-meme-analogy: memes lack a copy-fidelity that is high enough for being replicators. I will call this the *copy-fidelity* *problem*. It is related to the meaning problem, but it addresses not only an epistemological problem, about how to find out whether a meme is shared, given that people exhibit the *same* behavior. It gives a *reason* why the meaning problem exists at all: Memes change during the process of social learning. In addition, since the corresponding phenotypic expression of memes might change as well, as a consequence, the copy-fidelity problem also addresses cases where the outward behavior and artifacts are *different* after transmission, since they change in line with the alleged memes.

However, just like the meaning problem, the copy-fidelity problem has been answered by memeticists by referring to memes as ideational units with 'hard' essences. Although we met this argument already in the last section, I will nonetheless summarize what memeticists have said with respect to copy-fidelity. Their answer shows that memeticists end up assuming inferences to memes – inferences that are made by the learning individuals themselves, and not only by the researches looking at social transmission. These inferences have been considered as providing a basis for two further arguments against memes as replicators: One argument doubts that memes fulfill the lineage requirement (*lineage problem*); another argument doubts that memes fulfill the non-triggering requirement (*triggering problem*). Since replicators in the narrow sense do not only have to fulfill the similarity requirement but also the lineage and non-triggering requirement this provides a serious problem for the analogy between genes and memes as replicators. Both arguments claim that in most cases of social learning – even in cases where there is reliable transmission of behavior – the process of transmission of the underlying supposed memes is not analogous to replication: Memes are not replicated, they are *inferred* by the learning individual from multiple encounters with similar ideas. In a nutshell, the critique is that, even if we assume that memes have sufficient copy-fidelity, in order to spread across a group of individuals, they do this not via *replication* in the narrow sense.

### Copy-fidelity problem

The copy-fidelity problem points towards the reasons why there is a meaning/holism problem at all. Most memes are considered to fail to meet the similarity requirement, since they are changed when they enter the mind of a new ‘host.’ They are made ‘consistent’ with what is already there or changed for other reasons. This change can affect the ‘phenotype’ of memes (i.e., the consequences in behavior), but does not have to. The important thing is that from this changeability of memes, a disanalogy is derived: Memes are too soft, too pliable to change to serve as analogous to genes, which are not soft, but hard – hard enough to count as replicators and thus able to survive long enough in identical copies to serve as units of a multiple-step selection process. As Dennett has put it with respect to biological evolution: “Raise the mutation rate just a bit too high and evolution goes haywire; natural selection can no longer work to guarantee fitness over the long run” (Dennett 1995: 354).

Before I show how the issue has been discussed by memeticists, I have to add two notes of caution. First, from the point of view of general selection theory, it is unclear how much mutability or softness is still compatible with a multiple selection process. Although Dennett is correct with respect to biological evolution, in general a high mutation rate is compatible with the possibility of a selection process. The immune system is a perfect selection system with a *very* high mutation rate, as Hull & Wilkins (2005: 6) and Henry Plotkin (2000a: 77) make clear. It depends on the strength of selection in the respective system how much mutability is compatible with multiple-step selection. To say that a mutation rate is too high for a respective system requires an empirical proof of this claim, as Blackmore (1999: 58) answers. We have met the same problem with respect to Dawkins dismissal of organisms because they lack sufficient copy-fidelity: it is notoriously unclear how much is enough. Second, it is often unclear whether the changeability of memes is due to a process that is analogous to mutation, or to other processes. Apart from attributing a ‘high mutation rate,’ the softness of memes has also been regarded as due to ‘blending’ or ‘Lamarckian inheritance.’ Let me explain this by illustrating how Dawkins reacts to the copy-fidelity problem.

Dawkins addressed the issue about the copy-fidelity of memes already in *The Selfish Gene* (1976). He wrote:

> “At first sight it looks as if memes are not high-fidelity replicators at all. Every time a scientist hears an idea and passes it on to somebody else, he is

> likely to change it somewhat. I have made no secret of my debt in this book to the ideas of R.L. Trivers. Yet I have not repeated them in his own words. I have twisted them round for my own purposes, changing the emphasis, blending them with ideas of my own and of other people. This looks quite unlike the particulate, all-or-none quality of gene transmission. It looks as though meme transmission is subject to continuous mutation, and also to blending." (Dawkins 1989 [1976]: 195)

Nonetheless, according to him, the disanalogy is an illusion. As genotypic properties 'blend' at the level of phenotypic properties, interacting to build these phenotypic effects, ideas blend *only* at their phenotypic level. The Darwinism meme blends with other things at its 'phenotypic level'. The differences in interpretation of different versions of Darwinism, maintained by different people, are not part of the essential 'Darwinism meme' (Dawkins 1989 [1976]: 195f). The differences are interpreted as occurring only at the level of the phenotype of memes, whatever that is, where one meme blends with others and where the meme gets muddled with all the phenotypic noise of interpretation and additional aspects. At the same time, the essence of the meme resides unchanged somewhere in the mind, as genes are understood to stay unchanged in the nucleus of cells since Weismann postulated the non-inheritance of acquired characteristics.

In 1982, Dawkins changed his opinion and accepted the changeability of memes as indeed providing a disanalogy. Besides pointing to other differences between genes and memes, he admitted:

> "There are, of course, significant differences between meme-based and gene-based selection processes […]. The copying process is probably much less precise than in the case of genes: there may be a certain 'mutational' element in every copying event […]. Memes may partially blend with each other in a way that genes do not. […] The equivalent of Weismannism is less rigid for memes than for genes: there may be 'Lamarckian' causal arrows leading from phenotype to replicator, as well as the other way around. These differences may prove sufficient to render the analogy with genetic natural selection worthless or even positively misleading." (Dawkins 1982a: 112)

I will come back to this 'recantation' of Dawkins, since he nonetheless sees some value in the analogy, not for culture theory, but for gene selectionism itself. At this point only the following is important: Mutation, blending and Lamarckian causal arrows are forces that would change hereditary material. They can clearly be distinguished for biological evolution. Mutation happens to particulate genes and happens relatively rarely. Blending and Lamarckian causal arrows could so far be excluded to exist in biological evolution. Blending would mean that during reproduction the hereditary material mixes and that the outcome is a kind of average between the two ‘genes’: big wings of one parent and small wings of the other result in mediate size wings. Mendel has proven that this is wrong. Hereditary factors do not blend. Lamarckian causal arrows refer to the inheritance of acquired characteristics, through which the progeny inherits the changes that a phenotype acquires over its lifespan. Weisman first established the claim that this is impossible, a claim that has been confirmed since then and found its most concise formulation in the ‘central dogma’ of molecular biology, as briefly described in section 2.5. The common core of blending and Lamarckian inheritance is that both would render ‘hard inheritance’ impossible. Although a high mutation rate has to be distinguished from the latter two, it can also lead to a hereditary material that might be too soft to serve as a material basis for a selection process. Dawkins (1982a) does not explain in which sense change through mutation, blending, or Lamarckian causal arrows can be distinguished in the case of memes. Be it as it may, the difference will be important at a later point of this section.

Although Dawkins regards high mutation rate, blending and Lamarckian causal arrows as providing a disanalogy between memes and genes in 1982, he changed his mind later on. In 1999, replying directly to the objection that memes are not stable enough to be proper units of heredity and selection, Dawkins then distinguishes between his two kinds of memes, which I introduced already in section 3.1. He introduces them in order to partly restore the analogy. If a child learns how to fold a Chinese paper junk because a teacher instructs the child how to do it, then the meme (i.e., the instruction) has a non-memetic phenotype, namely the paper junk, an imperfect realizations of the ‘idea’ or ‘instruction.’ Blackmore (1999: 59-62) called this transmission process “copy-the-instructions.” This is, according to Dawkins, analogous to replication. If the teacher, however, does not transmit instructions but merely shows the child a drawing of a paper junk, which is the meme in this case, and asks it to copy it, then the meme – the drawing – is genotype and phenotype at the same time. In such a case, Blackmore calls the process “copy-the-product,” as opposed to copy-the-instructions. The transmission processes are consequently different, most importantly with respect to copy-fidelity: The first kinds of memes are, according to Dawkins, based on Weismannian or Neo-Darwinian inheritance and show a high copy-fidelity, whereas the latter do not and are inherited in a Lamarckian manner. 'Lamarckian' in this context means that the imperfect phenotypic realizations are copied. Hence, in the case of copy-the-product, imperfections accumulate and lead to a total different drawing after a couple of transmissions from teacher to child to another child and so on. According to Dawkins, in the case of learning the instructions, what is copied and is therefore the sought-for meme is an "idealized task" (Dawkins 1999: xii) and it is replicated since it shows a high copy-fidelity. As Dawkins writes, "Plato would enjoy it: what passes down the line is an ideal essence of junk, of which each actual junk is an imperfect approximation" (*ibid.*: xii). The child that should learn the task does thus not slavishly copy the product of the behavior of the teacher, which always shows arbitrary details that might not have been intended. The child tries to *infer* the intentions beyond the verbally given instruction and the exemplar produced by the teacher. The learning individual will, for instance, try to fold all four corners of the paper into the exact center of a perfect square, even if the teacher folded it not exactly at the center. That is why the transmission of the "inferred Weismannian instruction" (*ibid.*: xii) has much more copy-fidelity than the second meme-transmission in the example, namely the transmission of the drawing.[27] We thus get memes (or more precisely behaviors and artifacts) with a high-copy-fidelity, and others with a less high-copy-fidelity.

Genes are assumed by almost everybody in evolutionary debates to exhibit a very high-copy-fidelity. Now, we cannot infer a general disanalogy between genes and memes as such: Genes mutate as well and it is unclear how

---

[27] Dennett's (1995: 354ff) solution goes in the same direction, while mixing up different kinds of how minds change memes when they first confront them, work with them and in which sense this has an analogue in biological evolution, be it Lamarckian causal arrows or blending or mutation. Blackmore also answers with the "gist of a story" that is maintained (Blackmore 1999: 6, 43) or with the above described distinction between copy-the-instructions as opposed to copy-the-product (*ibid.*: 59-62, 213-216), when confronted with the changeability of memes.

much copy-fidelity is enough for a selection process to occur. The copy fidelity problem thus does not lead to a strong argument that shows that memes are in principle different from genes. Furthermore, as stated in the last section 3.3, we often do not even know exactly how much similarity the memes have, since we do not have direct access to them.

My critique against the gene-meme-analogy is rather that the copy-fidelity-problem shows that both – copy-the-instructions as well as copy-the-product – involve an *inference* to a Weismannian instruction. This is ignored not only by Dawkins and Blackmore but also in all the meme debates. Copy-the-product is more akin to changes only because it makes the inference to the intentions of the person showing something much harder, although not impossible. To make the same Chinese paper junk, if you only have the junk of somebody else as a model, is much harder compared to a case where you have in addition somebody explaining to you what you have to do. That is why we need teachers to teach our children so that they learn as quickly and efficiently as possible what we want them to learn. That is why reverse-engineering in technology is hard and why we have copy-right laws that help to keep the recipes (i.e., the instructions) for making Coca-Cola and other goods secret so that copying is made harder. Nonetheless, it is important to realize that both processes, copy-the-instructions as well as copy-the-product, involve inferring a "Weismannian instruction." Even in copy-the-product, the learning individual has to reconstruct the instruction from the observable verbal and non-verbal behavior of its teacher. If the inference is safe, the copying process of the same behavior can exhibit a high copy-fidelity. In all cases, and this is important, copy-fidelity depends on the reliability of these inferences and not on the kind of meme or transmission per se.

In addition, that all cases of transmitting memes involve inferences, shows that all cases of meme transmission involve reconstruction of memes: If we transmit a meme, the meme is not transmitted directly. On the contrary, it is reconstructed. This means that meme transmission is *always* more similar to what Dawkins has called 'Lamarckian inheritance,' and what I would like to call the *epi-memetic inheritance of memes*: What is directly transmitted in copy-the-product as well as copy-the-instructions are not memes (i.e., Dawkins' ideal essences) but behavior or artifacts. This is what the child observes and takes from the teacher. If there are ideal essences at all, they become reconstructed, in copy-the-instructions as well as in copy-the-product, by the learning individual through inference from the respective observables. If at all, memes are inherited through an epi-memetic channel, by a process of inferential reconstruction.[28] This is evidently not the way genes are transmitted from person to person. Genes are not reconstructed from transmitted phenotypic characteristics. And that they are not reconstructed is central for Dawkins replicator concept, since otherwise he could not exclude organisms from being replicators by pointing to non-Lamarckian inheritance, as shown in section 2.5. Therefore, the epi-memetic inheritance of memes provides a central disanalogy between genes and memes.

Nonetheless, my main critique lies somewhere else: The process of inferential reconstruction, occurring in copy-the-instructions and copy-the-product, is the basis for the lineage and triggering problem, which address whether the way we learn these instructions, or any kind of memes, is analogous to gene replication in the narrow sense.

Before I proceed to that issue, I want to make a concluding point with respect to the issue about copy-fidelity. As we saw already in the last section, the similarity requirement for memes is claimed to be fulfilled by simply defining memes as that 'something' that reoccurs reliably across people, so that what the learner has in mind, after he learned something, can be considered as similar to what the teacher had in mind. Since learning evidently does take place, it is assumed that there are ideational units – memes – with a high copy-fidelity that stay the same despite differences in their 'phenotypes.' As mentioned before, whether such essential ideal units really exist in the minds of people is not indubitable. One could also say that the way memeticists answer the issue about copy-fidelity is selling a definition as a hypothesis: They secure

---

[28] Epi-memetic inheritance actually is not really 'Lamarckian inheritance' either. If we transfer Lamarck's concept of inheritance to culture, we have to be careful since it can be transferred in different ways, as I have illustrated in more detail in Kronfeldner (forthcoming). The details are not important here.

fidelity and similarity through stripping off all differences that might result from the transmission processes between persons as non-memetic, as external to memes, as mere arbitrary details of the abstracted ‘idealized task.’ As Allen Orr has replied to Dawkins defense strategy:

> “For one thing, the low fidelity of memes is a simple observation: we all know that ideas change as they pass through many minds. This brute fact is unchanged by clever argumentation about how memes *could* be replicated with high fidelity; the fact is they often aren't. To put it differently, Dawkins's attempted fix is at best relevant to a subset of memes” (Orr 2004: 28; Emph. in the orig.).

Thus, as I said with respect to the material identification problem: There might well be memes that do not fulfill the similarity requirement. Whether they do is hard to find out.

A second reply offered by Orr leads to the next issue on whether the process of inferring the intentions of the teacher is replication, even if the meme shows high fidelity. Orr proceeds:

> “Finally, Dawkins’s fix only seems to work because he’s smuggled in a battery of mental processes like inference, intuition, and idealization: the child figures out the ‘inferred instruction,’ intuits what the ‘instructor intended,’ and correctly identifies certain ‘idealized tasks.’ The problem is that it’s all this inference, intuition, and idealization that does the heavy lifting in Dawkins’s scenario, not memes. It’s hardly surprising that if every child infers the same implied task, all children will pass on instructions for the same task. But this leaves wholly unexplained why and how each child infers the same thing—and *this* is the source of high-fidelity copying in Dawkins’s scenario. While I wouldn’t claim that this objection is fatal, it at the least suggests that, if you want to understand the mind, you’re probably better off trying to understand inference, intuition, and idealization than memes” (Orr 2004: 28; Emph. in the orig.).

I have already illustrated in the last section in which sense this shows that the gene-meme-analogy is heuristically trivial, since to refer to ‘idealization’ in the process of learning does not add anything to what social scientists and philosophers have worked with so far, namely the idea that common ideational units lie behind common behavior. Furthermore, others have indeed claimed that the invocation of inferential reconstruction is fatal for the analogy: It shows that the causal process of social learning is not analogous to *replication* in the narrow sense, as defined by Dawkins (see section 2.5).

### Meme replication as a process

If we take for granted, for the sake of argument, that there are cases where ‘something’ ideational is shared between persons, then this something fulfills the similarity requirement for replicators. The similarity requirement was, however, only one of the requirements a replicator has to meet. Similarity is not enough; replication as a process implies further requirements. There must be a certain causal connection between the original and the copy. As explained in section 2.5, something can be a copy without being produced by a process of replication in the narrow sense. In addition to similarity, the narrow concept of replication included the *lineage requirement*, the *non-triggering requirement*, the *self-replication requirement* and the *active-difference maker requirement*. With respect to the ontological analogy, critics mainly addressed the lineage and the non-triggering requirement. Since the status of the self-replication requirement is already unclear for genes, it is usually not debated at all. A word, as a sheet of paper in a copying machine, clearly cannot self-replicate, it always needs minds in order to be copied. The active-difference maker requirement is not essential for the ontological analogy, since it is that requirement that would make minds – as phenotype of memes – to mere consequences of memes. This issue will therefore be addressed in chapter 5. Relevant for this chapter is not whether memes are active replicators but whether memes are replicators at all, i.e., whether social learning is analogous to replication in the narrow sense. Therefore, the essential question is: If we take for granted that there is something ideational shared through social learning, *how* is it transmitted – through replication or not?

In order to answer this question one has to look more closely at the actual mechanisms of social learning, i.e., at the details of the cognitive processes that make social learning possible, processes which evidently involve ‘inference, intuition, and idealization.’ A review of what social learning is will then lead us to the critique that in most of these processes of social learning the lineage requirement (lineage problem) and the non-triggering requirement (triggering problem) are not fulfilled.

## Social learning

There are two levels of analysis that help specify different kinds of social learning. The first level of analysis is a populational one, the second a psychological one. At a *populational* level, Cavalli-Sforza & Feldman (1981) distinguished between three *types of social transmission*: vertical, horizontal, and oblique transmission.

> “*Vertical* transmission is used to denote transmission from parent to offspring and *horizontal* transmission denotes transmission between any two (usually unrelated) individuals. […] We will, however, use the term horizontal as restricted to members (related or not) of the *same* generation, and in addition we introduce the word *oblique* to describe transmission from a member of a given generation to a member of the next (or later) generation who is not his or her child or direct descendant.” (Cavalli-Sforza & Feldman 1981: 54; Emph. in the orig.)

Different concrete “*modes*” of transmission, such as parental teaching, sib-sib-interactions, peer learning, teaching, enculturation through social hierarchy, political indoctrination, mass communication, can be ordered according to these three major *kinds* of transmission. The distinctions are analytical distinctions, since in practice all the modes “interact and produce transmission matrices of great complexity” (*ibid.*: 59).

Cavalli-Sforza (2000: 179-187) later refined this typology. He now distinguishes between vertical transmission and three different forms of horizontal transmission. Vertical transmission occurs between parents and their biological or adopted children. Horizontal transmission now includes all pathways between biologically and socially unrelated individuals, i.e., individuals whose contact and relationship is not as enduring and stable as between parents and children. It can be split up into three types that are distinguished with respect to the *number of sender or receiver* in the transmission of a cultural item, or ‘trait’ as Cavalli-Sforza prefers to say. If the sender belongs to an older generation, transmission is still called oblique. The first type is a one-to-one communication pattern, as in peer-to-peer communication. The second type, ‘magistral’ transmission, involves a one-to-many communication pattern, as in mass media, or as in cases where an authority enforces a cultural item on a population by decree or other political or social pressures. The third type, which Cavalli-Sforza has called ‘concerted,’ involves a many-to-one communication, where a group of people exerts social pressure on new members.

What is important for our discussion here is that magistral and concerted transmissions depart significantly from the replication of genes, since the copies of a cultural item and their original do not form lineages. The transmission is comparable to making many copies of a book from one original or superimposing many versions of an original in order to produce one copy, a kind of blend of the different versions. As shown in section 2.5, such processes are excluded by Dawkins from being true replication processes, since the lineage requirement is not fulfilled, which is important for the possibility of cumulative evolution. Already this shows that memes cannot be replicators in the narrow sense, given that some social learning evidently relies on magistral and concerted transmission. Memes that are transmitted in this manner are thus not replicators in the narrow sense. But note that even vertical transmission might fail to fulfill the lineage requirement. This becomes evident, when we look at the psychological level of learning.

At a *psychological* level, cultural transmission can be due to diverse *cognitive types of learning*. There are many psychological classifications of kinds of learning. Avital & Jablonka (2000: 90ff) for instance have counted thirty different terms to distinguish between different forms of social learning, whereas many overlap.[29] The types of social learning differ with respect to the cognitive demands and with respect to what exactly is learned. Social learning in general does only require that “the presence of one relatively experienced individual increases the chances that a naïve individual will learn a new behaviour,” as Avital & Jablonka (2000: 90) have put it. This distinguishes social learning from pure individual learning. However, it is common to distinguish between social learning that is merely learning *through* others on the one hand and learning *from* others on the other hand. Learning through others means that the novice learns *through* observation not the behavior itself, but only something *about the environment*. This kind of learning has also been

---

[29] See Heyes (1994) attempt to bring some order in the terminology about social learning. See Heyes & Galef (1996) for an overview of debates about different kinds of social learning.

called 'stimulus enhancement,' or 'local enhancement.'[30] The individual still learns on its own, it 'reinvents the wheel' but it is *enhanced* through being exposed to the relevant stimuli. Learning not only through, but also directly *from* others involves learning something *about the behavior* of the experienced individual. This has often been called 'imitation in the broad sense.'[31] As Avital & Jablonka put it, "[w]ith socially influenced learning, an animal learns *what* to do as a result of its association with others; with imitation it learns both what to do and *how* to do it" (Avital & Jablonka 2000: 93).

There is considerable debate over how much animal learning is due to imitation in this broad sense, and not due to mere enhancement learning. A famous example is the spreading of the habit of opening of milk bottles by the Great tits in Britain. It has been shown that the method was not imitated by these birds, but learned by trial-and-error by each individual tit. It has been shown that some tits learn the behavior, even if they do not watch the behavior but merely the outcome of the behavior – the opened milk bottles.[32]

Learning *from* others (imitation in the wide sense) comprises simple *observational learning* as well as kinds of learning that are more complex, since they are based on understanding of symbols and attributions of intentionality. Pure observational learning has also been called "imitation in the narrow sense."[33] Humans imitate when they slavishly copy bodily movements. Imitation in the narrow sense is only a minor force in culture, since, for complicate tasks, humans have developed more efficient ways of learning. Tomasello (1999) insisted on such a type of social learning. According to him it is peculiar to humans. He calls it *cultural learning*. Cultural learning involves role-taking and sharing intentions. Cultural learning is not possible without joint attention that is not exhibited by children before the so-called nine-month-revolution. It is also not exhibited by animals. According to Tomasello,

---

[30] For instance, by Maynard-Smith & Szathmary (1999), or Boyd & Richerson (2000).

[31] See, for instance, Avital & Jablonka (2000: 92), Blackmore (1999: 43), or Dawkins (1989 [1976]: 206).

[32] See Sherry & Galef (1984); see also Tomasello (1999) for review of the debate and further examples.

[33] Plotkin (2000a: 75f), Laland & Brown (2002: 210). Calling observational learning imitation goes back until Thorndike (1898), as Plotkin mentions.

cumulative cultural evolution does not only require imitation in the wide sense, but also understanding the intentionality of actions (goals of actions, functions of artifacts). The child learns not only *what* to do (as in enhancement learning through others) and *how* to do (as in learning from others) but also *why* it should do it. According to Tomasello, without the latter the cognitive separation of goal and means is not possible. Yet this separation is necessary for an intentional variation of the already developed means in order to intentionally improve on these means for a certain goal, or in order to improve on the goal itself. Furthermore, without cultural learning language acquisition or the acquisition of any symbolic culture would be impossible. The arbitrariness of symbols makes it necessary that one understands the communicative intention in the use of symbols. Only on the base of cultural learning, humans are able to learn symbolically. Symbolic learning is very effective, since the respective learning situations can be represented symbolically. Even if the behavior is not performed, the novice can learn it through explaining what, how and why he has to do this or that. This is the psychological basis of the difference between Dawkins' copy-the-instruction learning and copy-the-product learning. Since only humans are engaged in cultural learning and massively use symbol systems, cultural learning has been taken by Tomasello to explain why only humans managed to build up such impressive cultural systems like religion, art, technology, science etc.[34]

To recapitulate, at the psychological level we can systematically distinguish between three cognitive types of learning: *enhancement learning* (learning merely through others); *observational learning* in the narrow sense (learning through and from others by simply repeating behavior); insightful *cultural learning* processes (learning through and from others, involving attribution of intentions, use of symbols and inferences with respect to the

---

[34] In addition to Tomasello (1999), see Cavalli-Sforza & Feldman, who also rest human uniqueness on this capacity and add that therefore culture can be retained over time and over space without direct contact of individuals: "[W]hat may be unique to man is the capacity to transmit knowledge to other individuals remote in space and time by means of such devices as writing, mainly, transference of abstract instructions and explanations in ways that do not require face-to-face observation and direct imitation" (Cavalli-Sforza & Feldman 1981: 4).

meaning and function of symbols or actions). In order to keep things clear, I will use these three terms –enhancement learning, observational learning, and cultural learning – to distinguish between these three types of learning. The reason for this terminological decision and the detailed review of types of learning is that the use of the term imitation has caused much confusion: it is sometimes used for all kinds of social learning, sometimes for learning from others – observational or cultural – and sometimes only for observational learning.

Before I proceed to the lineage and triggering problem, I have to add a note on the explanatory force of postulating these different kinds of learning. All these three psychological types of learning might well comprise different cognitive mechanisms that could be specified further. Learning a motor pattern or learning how to dance, even if both might well be instances of observational learning, could involve different mechanisms. Furthermore language acquisition, learning to use mathematical symbols, learning to understand and work with social constructions such as ‘money,’ or learning to understand Newton’s laws on the basis of other basic competences and already acquired knowledge will also differ and involve different cognitive mechanisms, as Plotkin (2000a) stresses, although they all belong to the category of cultural learning. For our purposes, however, it suffices to distinguish between these three cognitive types of learning. Explanation of social learning clearly has reached this level in psychology of learning. If we want to explain how a single tit learned to open milk bottles, we say that it learned it by enhancement learning, or imitation. Given the distinctions about social learning, we can now ask whether social learning is analogous to replication.

### Lineage problem

I claimed above that in cases of magistral and concerted transmission, the lineage requirement is not fulfilled. However, at the psychological level, magistral and concerted transmissions do not so much differ from one-to-one transmission. Even if we assume that we exclusively learn something from one person alone, we might not learn it instantaneously. We might well learn it through multiple encounters with an outward realization of that ‘something’ that we should learn. Even if one ends up with a cultural item that is very similar to what others maintain, e.g., a certain pronunciation of 'whether,' one did certainly not acquire this 'copy' through a process of copying one's 'copy' from a single template-original – whether one learned it from one person alone or from many people. In such cases, the 'meme' has not been transmitted through replication in the narrow sense. Thus, in all cases of learning that rely on multiple encounters with tokens of a type that is said to be copied, the process is not a true replication, since the lineage requirement is not fulfilled in these cases.

As Dawkins himself has recognized for genes, if the lineage requirement is not fulfilled, then cumulative change becomes impossible, since a change that has been introduced newly in one copy is then not passed on in subsequent copies. And this is what happens in cases of multiple encounters, since the newly introduced variant is averaged out. If a child acquires the pronunciation of 'whether' through multiple encounters, critiques of the gene-meme-analogy claimed, the child will 'blend' or 'average' over all the information it receives from others.[35] Through this blending it is secured that the child ends up with the average of the population. Heredity, and therefore social learning and similarity at least in a wide sense, are therefore secured. Yet at the same time a newly introduced change, for instance in pronunciation, occurring in one individual in the population, has almost no chance to spread, since it is averaged out: i.e., it does not survive the averaging-process because of blending. As said, that is the reason why the lineage requirement was so important for the replicator concept and also for the concept of Darwinian evolution as such: Changes occurring at one point must be heritable; they are not heritable if blending is operative. Mendel showed, although not yet at a molecular level, that heredity does not involve blending. It is particulate.

---

[35] Boyd & Richerson (2000), Henrich, Boyd & Richerson (forthcoming), Godfrey-Smith (2000: 419), Sterelny (forthcoming a). Boyd & Richerson (2000), Henrich, Boyd & Richerson (forthcoming) and Gil-White (forthcoming) try to explain this blending process at a cognitive level. Boyd & Richerson, for instance, claim that heredity is secured partly by a conformity bias, partly by other biases.

Therefore, the lineage requirement is, first, a central aspect of Darwinian evolution and, second, it is one of the essential requirements for replicators.

From this it follows that if memes are often blended in the sense that they do not fulfill the lineage requirement, memes are in many cases not transmitted through replication.[36] For these cases there is an important disanalogy between genes and memes. Without being able to offer empirical evidence here, I guess that many cultural units are transmitted through a diffuse blending process of multiple encounter with multiple templates. In culture, heredity is usually secured differently than in nature. Therefore my main conclusion with respect to the lineage problem is: In cases of social learning that involve multiple encounters, memes are not replicators. The analogy is *false* for these cases of social learning, since the lineage requirement is not fulfilled.

Yet, for clarity, I have to add how change is possible nonetheless on the basis of blending inheritance. Is it really true that only particulate inheritance can secure cumulative evolution? Boyd & Richerson and other critics of memetics, who want nonetheless to maintain that culture evolves in a Darwinian variational manner, claim that in the absence of replication, blending heredity can lead to the accumulation of changes, as required for Darwinian evolution. Replication is not needed for a Darwinian account of culture.[37] According to them, what changes is the statistical mean of a certain cultural item. The mean (of phenotypic realizations of alleged ideational units) can be tracked by researchers, and the mean can be said to change and spread differentially in a culture. I do not want to go into the details of their defense here, since the only thing important for this study is not how cumulative change is possible at all, even if memes do not replicate as genes do. The

---

[36] Note that referring to blending is different from claiming that a meme pool is subject to constant crossing and joining of branches or lineages of species, while “biological evolution is a system of constant divergence without subsequent joining of branches”, as Gould (1991: 65) objected against memetics. Genes may replicate and hence fulfill the lineage requirement, even if some genes in a population come from a different species, a phenomenon that does in fact sometimes happen in biological evolution, as Hull (1982) argued.

[37] See Henrich, Boyd & Richerson (forthcoming), Gil-White (forthcoming). Their solution has been criticized and refined by Sterelny (forthcoming b).

important thing is that the latter question occurs only because blending shows that *memes are usually not replicators in the sense defined by Dawkins*.

### Triggering problem

As outlined in the subsection on the copy-fidelity problem, memes do not only blend, they are also inferred and reconstructed. Our example was learning to make a Chinese paper junk. According to Sperber (1996, 2000) and Atran (2001) the invocation of inference provides a further argument against the claim that memes are replicators, since the invocation of inferences shows that memes fail to fulfill the non-triggering requirement. Recall that the non-triggering requirement says that replication requires that 'information' is transferred and not only triggered. If it is only triggered, the information was there already. The example from section 2.5 was laughter. If laughter spreads in a group of people, it is not replicated in the narrow sense; it is triggered. Sperber explains the difference between replication and triggering with a thought-experiment:

> "First case: ten sound-recorders with the same repertoire of melodies in each have been fixed so that they are activated by the sound of the last five bars of any melody in their repertoire, and then play this very melody. They are placed in such a manner and at such a distance of one another that the first one activates the second, the second the third, etc. The first recorder plays melodies in random order at appropriate time intervals. Second case: ten sound-recorders have been fixed and placed so that the second-recorder records sound from the first, and then replays it, the third recorder records sound from the second and then replays it, and so on. Only the first recorder has a ready repertoire of melodies, and it plays them in random order at appropriate time intervals" (Sperber 2000: 169).

According to him, in the first case, "only triggering takes place and no copying at all" (*ibid.*: 169). Since replication requires what Sperber here calls 'copying', no replication has taken place. Applied to the question whether social learning is an instance of replication, analogous to gene replication, the answer for Sperber is 'No,' since according to him triggering is always involved in learning.

If we learn something and do this through inferring certain intentions, as in Dawkins example of a child learning by instruction from a teacher how to fold a Chinese paper junk, then some of the information that the child uses to perform the task is actually not learned *at that very moment*. Part of the information necessary to understand the teacher has been learned long before. For instance, the instruction to fold the paper exactly at the middle of the paper to a square relies on general knowledge about squares, folding etc. This general knowledge is triggered in the learning situation, and the task itself is inferred by the child from what it knows already, and from what it sees and hears at that very moment. According to Sperber, the pre-existing knowledge can be acquired or innate; yet, according to him, it usually is innate, specified in cognitive modules.

Note, that Sperber does not (and cannot) claim that literally nothing is transmitted in such cases.[38] It is not an either-or-question. It is a question of more-and-less. The influence of pre-existing knowledge might for instance be less important in observational learning, but it certainly is very important in cultural learning, such as in learning the grammar of language or learning how to use symbols for sounds in music. How much triggering is involved in an instance of learning is thus dependent on each individual case. Therefore, although I agree with Sperber on his general point that triggering is not replication, it is hard to decide what to do with mixed cases, i.e., cases that do involve some copying, but also triggering. Sperber merely states that

> "[f]or memetics to be a reasonable research program, it should be the case that copying, and differential success in causing the multiplication of copies, *overwhelmingly* plays the major role in shaping all or at least most of the contents of culture. Evolved domain-specific psychological dispositions, if there are any, should be at most a relatively minor factor that could be considered part of background conditions" (Sperber 2000: 172; Emph. added).

Two points are important here:

First, how much is "overwhelmingly"? This is indeed unclear. Therefore, I suggest not to appeal to any relative measure in order to evaluate the claim about memes as replicators. Genes clearly do not involve *any* triggering, even if they cannot self-replicate. The form of DNA copied from a previous DNA-token definitely stems from the latter, which is a template, and not from proteins or any other entities involved in the complex molecular

---

[38] See Sperber (2000: 172f). See also Atran (2001: 375f) who merely adds details to Sperber's argument with respect to this point.

replication machinery. Therefore, the triggering problem seems to show a second central disanalogy between genes and memes with respect to the causal process that connects a serious of similar tokens of a type, be it a meme or a gene. The disanalogy would be that, contrary to genes, memes almost always include *some* triggering.

In order to show how memeticists can react to this charge, let me add a second note. Not only innate pre-existing psychological knowledge is triggered and not replicated. If Sperber's argument (i.e., that triggering proves that the transfer of knowledge is not an instance of replication) is correct, then the argument holds for *any* pre-existing knowledge (i.e., for any knowledge that is pre-existing in the mind of the novice at the very moment where the novice learns a specific cultural item). As shown in chapter 1, not everything in culture is innate. There evidently are things that are definitely learned. I therefore disagree with Sperber on the following point: that evidence for innate knowledge provides an argument against memetics. If *only* triggering of innate knowledge were involved in a learning situation, the situation would in fact be no instance of culture as contemporarily defined. The contemporary definition of social learning, which I presented in section 3.1, entails that the content that makes up culture is *socially* transmitted and not genetically. Therefore, for those cases that clearly do involve social learning, Sperber's form of the triggering argument does not provide a critique against memetics, who only deal with instances of social learning. The triggering argument can only be defended for cases that clearly are cases of social learning. Now, the question is whether those instances of acquiring knowledge that definitely include social learning (i.e., *cultural* items) are *learned by replication* or whether they are learned by triggering of previously *learned* knowledge.

That is the snag of the whole issue. From the latter it follows that the triggering problem that can be objected to memetics applies to cases where the pre-existing knowledge that makes learning possible has been acquired earlier in life. However, since even in such cases the issue is not an either-or issue, memeticists could answer to the triggering problem in the following way: There is always *something* that is transmitted *at that point in time*, even if pre-existing knowledge helped to transmit it. In other words, a memeticist simply needs to answer that what he regards as the meme is that ‘something’ that has been transmitted in that moment of learning, i.e. that something that a novice was able to add to his repertoire, given the background of pre-existing knowledge that made this addition to his repertoire possible. Plotkin (2000a) does exactly that. There are memes that are acquired early in life. Some of them are ‘deep-level memes,’ like the concept (or schema, frame, script, as others prefer to say) of a restaurant. A person can learn that Italian restaurants regularly serve pizza, only if a person first acquired the ‘deep-level’ restaurant-meme, to take a simple example. Nonetheless, the proposition ‘Italian restaurants regularly serve pizza’ has been learned at that moment and not only triggered through pre-existing knowledge. In other words, the information that has been learned at that moment, has been ‘bootstrapped,’ but not triggered, as Gil-White (forthcoming) objects in a similar way to Sperber’s triggering-argument against memetics. This holds even if the learning involves inferences to what a speaker means when he says ‘restaurant,’ ‘Italian,’ etc.

However, memeticists are responsible for not being precise in specifying what a meme is. As illustrated in this chapter, often they merely repeat that there must be ‘something’ that is similar or transmitted, for instance in learning by instruction of how to make a Chinese paper junk. As Dawkins has said, the meme is the instruction to make the paper junk. But this leaves undecided whether this meme includes or excludes the knowledge that has been learned before – knowledge that is needed for the child, in order to infer from the observed verbal and non-verbal behavior what his teacher intends the child to do. If a child learns to make the Chinese paper junk, then a lot of the things, which are required for the child to make it, has not been transmitted and therefore has not replicated.

In practice, it might be hard, if not impossible, to find out which part of a newly acquired skill has been transmitted and which has been acquired before. However, the distinction is important, in order to get a foothold on the triggering problem. For that ‘something’ that is indeed transmitted at the moment of learning, the non-triggering requirement is fulfilled. Only that ‘something’ that is learned can be the content of the newly learned ‘meme’ in the mind of the novice.

Therefore, if it is made precise what the actual meme is in an instance of learning, and if previously acquired aspects are excluded from the respective meme, then there is no difference between genetic replication and social learning with respect to triggering. This is my concluding claim on the triggering problem. In principle, the analogy can thus be restored. Yet the triggering problem shows again: It all depends on what is specified as the meme. If the skill of *folding* a Chinese junk is the meme, then the meme has not been replicated, as long as, for instance, the general skill of folding papers has been acquired before. If we, on the contrary, identify the meme as being the meme of folding this *particular* Chinese paper junk, then this particular meme has been replicated – given that the other replicator conditions are fulfilled as well. Therefore, whether the non-triggering requirement is fulfilled, and the analogy justified, on that ground, is dependent on how we define the respective meme. Memeticists usually do not care about this issue and maybe cannot, since it is practically almost impossible to dissect the learning situation in such a way. My conclusion on the triggering problem includes therefore the following critique: First, as with the similarity requirement, it is hard to decide empirically whether a meme fulfills the non-triggering requirement; second, the claim that ‘something’ is learned in each learning situation, the only claim that can be derived from calling memes replicators, is no news for anybody, except for radical nativists. In this sense the analogy is *heuristically trivial*, even if it can be made precise in a way that circumvents the triggering problem.

## Widening the concept of replication

I will now turn to a last resort some memeticists have taken in order to answer the different replication problems all together. Their answer is that replication in the narrow sense is not required. Similarity is enough. The concept of replication is thus widened so that any kind of cultural transmission counts as replication, reaching thereby an immunization against the critique that memes do not replicate in the narrow sense, as genes do. This defense strategy, however, leads to a trivialization of the claim that memes are replicators. As with the behaviorist something-is-preserved argument (discussed in section 3.3), the analogy thus becomes true but trivial, if we take for granted, for the sake of argument, that usually some ideational units are transmitted through social learning of similar behavior.

Dennett (forthcoming) states that the "invocation of intelligent, semantically-sensitive, intention-attributing agents in the purported replication process" does not as such provide a claim that the process is not an instance of replication, even if genetic replication is "mindless" and "mechanic," whereas social learning is not. Although this is correct, I think that the invocation of semantically-sensitive, intention-attributing agents nonetheless provides a problem for the claim that memes are replicators. According to Dennett, this is not the case: "Darwin (and Fisher, and Williams, and others) saw the need for a sufficiently 'strong principle of inheritance' to keep evolution going, but nothing has been said about *how* that fidelity is to be maintained, *mechanically*" (Dennett (forthcoming), Emph. added).[39] This might have been the case for Darwin and Fisher, but it is a crude oversimplification of the concept of replication, made precise by Dawkins on the basis of Williams' gene selectionism. Dennett's statement ignores all the issues of the contemporary units of selection debate. Heredity is not replication (as I illustrated in section 2.5 and above). Replication has much stronger requirements than heredity, requiring higher copy-fidelity, lineage building, non-triggering, self-replication, and a causal influence of the replicator on its own replication. Memes are often not replicated, since they often do not build lineages; and if they are not defined carefully, they do not fulfill the non-triggering requirement.

Oversimplification is the deeper, hidden problem that many memeticists face, if forced to the details of *how* the transmission of memes happens. Although they claim that memes do replicate, they tend to claim at the same time that the details are not important. They widen the concept of replication or

---

[39] The second reason he cites is that the mind is itself an effect of a mindless evolutionary process, which proves nothing with respect to the point at issue.

are unclear about what is demanded for replication. Again, according to Dennett, "[y]ou can finesse your ignorance of the gory mechanical details of how the information got from A to B, at least temporarily, and just concentrate on the implications of the *fact that some information did get there* – and some other information didn't" (Dennett 1995: 359; Emph. added). According to him, merely the similarity across people is important. Thus, for him there seems to be no difference between the following two specifications, which he uses simultaneously. At one point, for instance, he mentions that others have invented alternatives to the term 'meme' for the 'new replicator' and writes:

> "But since the word meme has secured a foothold in the English language, appearing in the most recent edition of the *Oxford English Dictionary* with the definition 'an element of culture that may be considered to be passed on by *non-genetic means',* we may conveniently settle on it as the general term for any culturally based *replicator* – if there are" (Dennett 2002: E-85; Emph. added).

Not every unit that is cultural, and therefore *by definition* not transmitted by genetic means, qualifies as a replicator in the narrow sense. Dennett and other memeticists constantly ignore this.

Dawkins also cited the English Oxford Dictionary for defining 'meme' as "[a]n element of a culture that may be considered to be passed on by non-genetic means, esp. imitation" (Dawkins 1999: viii). In a shortened reprint of this passage (Dawkins 2003), he changes this to "a self-replicating element of culture, passed on by imitation." One often does indeed not know what the exact claim behind the gene-meme-analogy is: that culture is based on social learning in the general sense, i.e., transmission through non-genetic means; or that culture is based on imitation, in the narrow sense of observational learning or in the wide sense that is tantamount to social learning in the general sense; or that imitation in one of these senses is analogous to replication.

The same confusion arises from statements of Blackmore. According to her, cultural inheritance (i.e., social learning) is based on imitation and imitation is analogous to replication. Blackmore (1999: 47-52) explicitly excludes 'local enhancement' as an instance of imitation and replication. She does this with the same argument Sperber has used to exclude cases of learning from being instances of replication in the narrow sense, namely with the argument that triggering is included. According to Blackmore, ‘local enhancement’ relies heavily on triggering and is therefore not replication. If a novice learns merely *through* others but not from *others*, since he is merely exposed with a greater frequency to the stimuli that make individual learning more likely, the new behavior or cultural idea itself has not been replicated. It has been triggered, in this case by the environment. However, observational learning (i.e., imitation in the narrow sense) and other more complex processes of learning (i.e., what Blackmore calls imitation ‘in the broad sense’) are considered by her as analogous to replication, despite the triggering problem for most cases of cultural learning:

> “Imitation is a kind of replication, or copying, and that is what makes the meme a replicator and gives it its replicator power. You could even say that ‘a meme is whatever it is that is passed on by imitation’ – if it didn’t sound so awkward. We may (and will) argue about just what counts as imitation but for now I shall use the word *‘in the broad sense’*, as Dawkins did. When I say ‘imitation’ I mean to include passing on information by using language, reading, and instruction, as well as other complex skills and behaviors. Imitation includes any kind of copying of ideas and behaviors from one person to another” (Blackmore 1999: 43; Emph. added).

What she offers later on does not add much clarity.[40] What is important is, first, that, although she excludes enhancement learning from being replication because of triggering, all other forms of social learning are taken to fulfill the non-triggering requirement. Second, she has nothing new to say on the “mechanisms underlying imitation” (*ibid.*: 51). She repeats the standard kinds of social learning, which I introduced above. According to her, the fact that memeticists make their claims without knowing “the mechanisms for copying and storing memes” does not matter. She offers two arguments for this position: First, Darwin started without having any reliable knowledge of how

---

[40] See Blackmore (1999: 46- 52): On the one hand, she admits that similarity between two tokens of a cultural unit is necessary but not sufficient for replication. She writes that to show that imitation is more than this, “we need to define imitation” (*ibid.*: 46). Using on the one hand, the narrow concept of imitation (observational learning), taken from Thorndike, she changes back and fourth to the wide concept of imitation. She then distinguishes between: (i) “contagion,” which is innately triggered behavior, such as laughing; (ii) “social learning,” such as enhancement learning; (iii) “true imitation,” i.e. learning from others (*ibid.*: 47-52). Since not every learning from others is observational learning, it seems that, on the one hand, all cases of social learning are true imitation; but on the other hand it seems that everything that fulfills the non-triggering requirement is considered as ‘true imitation,’ but this is not the case for all cases of learning from others.

heredity is secured, so can memetics. Second, “memes depend on being transmitted from one person to another and, by definition, this is done by imitation” (*ibid.*: 58). The first merely reaffirms that so far memetics has not reached the level of a science and it ignores the material identification problem. The second indicates that her concept of imitation is indeed a very broad one.[41] In such a case, memes are indeed – by definition – replicators: things that are transmitted from person to person through non-genetic means. If imitation in the broad sense (social learning in general) is replication for the sole reason that it fulfills the similarity requirement, then this contradicts her exclusion of enhancement learning as imitation and replication. Enhancement learning has been excluded because it does not fulfill the non-triggering requirement, despite it fulfills the similarity requirement. It is therefore totally unclear whether she requires more than similarity for a replicator or not. To require more than similarity for enhancement learning but not for higher forms of cultural learning is certainly no solution to the question whether social learning is replication or not.

Changing back and forth between a narrow and a wide concept of replication leads to confusion. In addition, widening leads to trivial statements about social learning. It leads to trivialization for the following reasons: (i) It conflicts with the narrow concept of replication, used to justify gene selectionism; (ii) it simply mirrors the anthropological concept of culture; (iii) last but not least, claiming that memes are replicators in the wide sense does not add anything to what social psychology has found out about social learning.

*(i). Conflict with the narrow concept of replication used to justify gene selectionism.* If memeticists answer to the objection that memes are not replicators in the narrow sense with a wide concept of replication, then any

---

[41] If, however, Blackmore would build her gene-meme-analogy on a narrow concept of imitation (observational learning), she would simplify culture by excluding all those kinds of social learning that are evidently different from simple observational learning. These kinds of social learning account for the more important parts of culture, e.g., social constructions like ‘democracy,’ ‘money.’ norms like the ‘Golden Rule’ or the Christian Ten Commandments, or mental scripts such as the concept of a ‘restaurant’ etc. See (Plotkin 2000a: 76, 2000b) on this line of argument.

kind of cultural transmission – by definition – counts as replication. This wide concept of replication, however, evidently conflicts with the claim that in the case of biological evolution only genes are replicators. If the loose standard that then holds for cultural evolution should apply to biological evolution, the difference between heredity and replication could not be justified (as I explained in section 2.5 and above in reaction to such a widening of the concept of replication in Dennett). Dawkins, when he withdraw from a strong analogy in 1982, wrote:

> "There are, of course, significant differences between meme-based and gene-based selection processes […]. These differences may prove sufficient to render the analogy with genetic natural selection worthless or even positively misleading. My own feeling is that its main value may lie not so much in *helping us to understand human culture* as in sharpening our perception of genetic natural selection" (Dawkins 1982a: 112; Emph. added).

If it is correct that memes are not replicators in the narrow sense, since they mutate too much, do not fulfill the lineage requirement, and – depending on individuation of memes – might not fulfill the triggering requirement, then memes cannot even be *used* to sharpen our perception of genetic replication, since memes are replicators only in a *wide* sense. If the wide concept of replication were used to illuminate the special role of genes, then the concept of replicators would become trivialized *even for genes*.

*(ii). Memes as replicators in the wide sense simply mirrors the anthropological concept of culture.* With a wide concept of replication, we are far away from illuminating the explanation of culture. Widening the concept of replication leads to a heuristically trivial reassertion of the anthropological concept of culture. Defending the claim that similarity for ideational units holds does not comprise special claims that would distinguish memetics from other approaches to social transmission. All of these approaches assume the basic anthropological concept of culture, i.e., they all assume that their objects of study are those generative ideational units of culture that show similarity across people, and that are transmitted through non-genetic means. The defense that memes are replicators in the wide sense states nothing more than that they are transmitted through non-genetic means. The less you require, the more trivial the analogy becomes. The gene-meme-analogy reinvents the wheel,

since memeticists neither developed the concept of culture with this analogy, nor did they contribute any new hypothesis about culture, nor did they provide any new solution to the identification and meaning problem. Everything that they say and that is correct can be found in any theory that tries to explain culture in the narrow sense. The gene-meme-analogy is thus *heuristically trivial* in this sense.

*(iii). Claiming that memes are replicators does not add anything to what social psychology has found out about social learning*. Finally, the gene-meme-analogy does not add anything to the explanation of how social learning works. Calling what we assumed all the time (that culture is based on the transmission of similar ideational units) 'replication' does not add anything to the explanation how this transmission works. Dennett and Blackmore state that we do not need this level of analysis for the analogy. We do indeed not need precise psychological accounts of how the information is transmitted in order to *state* the analogy. But whether the replicator requirements are fulfilled, whether the analogy is correct or not can only be found out by looking at the details of the psychological processes involved. Blackmore argues in one place for Dawkins gene selectionism, the view that only genes can be units of selection. She implicitly suggests that group selectionists were wrong since they "talked about evolution occurring for the 'good of the species' without worrying about the exact mechanisms involved" (Blackmore 1999: 4), whereas she takes our knowledge of DNA as replicating to show that these group selectionists are wrong. If this standard is used for biology, why not use it for culture – as a standard for judging the analogy?

A clear exposition of the different types of social learning, which have been distinguished by social psychology, although they are not yet an explanation at the level of true mechanisms, does allow to show that *not* all kinds of social learning are by definition instances of replication in the narrow sense. For many instances of social learning the lineage requirement is not fulfilled, even if we assume that these memes fulfill a quite loose standard of similarity. That this is the case is already visible at the populational level and could be backed up by looking at the cognitive process at the psychological level. The gene-meme-analogy needs to address these processes and the basic 'mechanisms' beneath these processes, as Plotkin (2000a) criticizes. Plotkin also wants to use the word 'meme' for an explanation of culture, but his account of memes takes all the points mentioned in this chapter into account, and he requires that memetics explain the mechanisms of transmission in order to have *explanatory value*. I disagree with Plotkin in claiming that – in order to proceed to this level – we do not need the gene-meme-analogy. The analogy is superficial at these levels of description and explanation. Social psychology works on these issues since decades and they do not need the meme analogy: First, they do not need the gene-meme-analogy to find out whether an instance of social learning is due to enhancement learning, observational learning, or cultural learning, the three basic forms I distinguished above; second, they do not need it to further explain how these processes work. At the level at which social learning is described and explained by describing and analyzing different processes of social learning, and at the level at which these processes might be explained further, the gene-meme-analogy is trivial, not offering anything that could count as explanation of how social transmission works.

### Conclusion

Considering all of the above, the gene-meme-analogy is wrong with respect to the narrow concept of replication, and the gene-meme-analogy is trivial with respect to the wide concept of replication: merely mirroring the definition of culture and not adding any new hypothesis to what we know already – an analogy with no fruitful role.

## 3.5 SUMMARY

There is heritability in culture, although the process *how* this is secured is different from the way heritability is secured in biological evolution by genes. Looking at the concept of culture, at the concept of memes, at social learning, and at the concept of replication in a precise way has proven to be decisive. The first basic assumption of the gene-meme-analogy, (1) that there are basic ideational units of culture, is heuristically trivial; the second assumption, (2) that they can be tracked, is wrong; the third assumption, (3) that they are replicators, is either wrong or trivial. The ontological analogy, claiming that culture is built and transmitted through basic building blocks that are observable replicators, in the same sense as genes are, is thus either wrong or trivial. It is trivial in its broad Darwinian guise, claiming that there is something transmitted, and wrong in its gene selectionism version, claiming that memes are countable replicators.

The only positive role the ontological gene-meme-analogy could play is to serve as a link between diverse disciplines (mainly philosophy, psychology, anthropology, and evolutionary biology), and to bring these disciplines together – to study mankind in its totality, as Hull (for instance in 2000) has claimed. Memetics can therefore be a philosophical perspective on culture that is valuable in this heuristic sense only – a *Wittgensteinian ladder* that can be thrown away as soon as we come to the details, i.e., after the different disciplines managed to talk to each other.

I have to add a last note on the relation between the ontological analogy and other applications of the Darwinian paradigm. I will illustrate the relation between the ontological gene-meme-analogy and the other two basic analogies at issue in this study in the respective chapters. However, as mentioned in chapter 1, there are numerous other applications that differ in many details from these analogies. It is not possible to evaluate them, their relation to, and dependency on the ontological analogy with the same necessary precision. I shall be content with a brief example. Although the ontological analogy has its defects in the sense shown in this chapter, it might nonetheless be a fruitful hypothesis to test whether cultures are – analogous to biological species – subject to fast evolutionary change if there are small groups, as Hull (1988a) claimed for science.[42] Such claims do not need the close analogy between genes and memes as replicators, as defined by Dawkins. What is important for

---

[42] See Hull (2000: 62ff) on more examples, or Blute (2005) for this typical manoeuvre. Laland & Brown (2002: 234-237) also cite a couple of empirical ways for detecting "natural selection of cultural variation". Memetics could for instance find out whether the phenomenon of convergent evolution occurs in unrelated cultures, i.e., whether they arrive at the same solutions independently, if confronted with similar problems. However, even if memetics would turn towards such empirical research, that research does not need a close analogy between genes and memes as replicators.

my critique of the gene-meme-analogy is that they therefore can also not be used to prevent the ontological analogy from the critique presented in this chapter.

# 4 ORIGINATION ANALOGY: DARWINIAN NOVELTY IN CULTURE

## 4.1 THE ORIGINATION AND PATTERN OF CHANGE IN CULTURE

### The origination analogy

Culture evidently changes in time. The question whether it changes in a Darwinian manner or not, can be approached in two different ways: First, is culture a variational system, with selection as a necessary part of the change? Second, does variation arise in an *undirected* way? That is, are the factors involved in the generation of novelty (i.e., in the process bringing about variation) and those involved in selection *decoupled* as in biological evolution or coupled, exhibiting a *Lamarckian correlation*? And is the correlation partial or even total, as in a Lamarckian instructive mechanism that guarantees that adaptive novelty arises? As explained in section 2.3, even if a system is a variational system, it can nonetheless include coupling, which is excluded in Neo-Darwinism. Thus, the two questions have to be addressed separately. There is a third question that, in principle, should also be addressed, when the analogy between origination of novelty in culture and in nature is at issue: In which sense does selection of cultural items differ from 'blind' (i.e., hindsighted and opportunistic) natural selection? I will address the issue about the hindsightedness of selection as part of my analysis of undirectedness of variation. In addition, I will address opportunism at the end of this section.

I will start with a detailed description about how the origin of novelty in culture can be approached, independent of Darwinian thinking. I will then shortly address the graduality of cultural change, the variational nature of culture, and intentional selection. That culture is often gradual, that it is basically a variational system, based on intentional selection, which differs from natural selection, are uncontroversial issues. Furthermore, the gradual and variational pattern of cultural change is usually not cited as justification of the Darwinian approach to creativity; and intentional selection is not cited as a critique of this approach. These aspects are not central for the origination analogy. That is why I will only shortly address these issues, as part of this opening section.

Most of this chapter is dedicated to a detailed analysis of the following issue: In which sense can creativity be said to be based on 'blind variation,' i.e. based on truly *undirected* variation as defined in section 2.3. This issue is at the center of the Darwinian approach to creativity and has caused considerable controversial debates over the last 45 years. The arguments have gone back and forth without much progress. I will lead through this zig-zagging in the following way: First, I will distinguish between three interpretations of the claim that creativity is based on 'blind variation' (section 4.2). Depending on what kind of blindness is meant, the evaluation of the analogy differs. I will then present the standard critique that culture and creativity do not rest on blind variation, but on *guided variation*. I will systematize and develop the critics' core argument, in order to show what exactly provides a problem for the origination analogy (section 4.3). After this I will introduce, analyze, and assess three kinds of *compatibility arguments* – arguments that state that the critique – although correct – does not destroy the analogy, since guided variation is compatible with drawing the analogy between origination in creativity and biological evolution (section 4.4 – 4.5). By distinguishing between three meanings of 'blind variation,' and by discussing these compatibility arguments, I will provide a new and well-balanced critique of the origination analogy that takes the points of critics as well as those of defenders of the analogy into account. However, my conclusions are not in favor of the origination analogy, since I will show that the three compatibility arguments either fail or end up in almost trivial claims.

Before I introduce how the origination of novelty in culture can be understood irrespective of Darwinian analogical reasoning from nature to culture, I want to add a note on the independence of the origination analogy from the ontological analogy. Culture relies on a certain degree of heritability of ideational entities. As illustrated in chapter 3, this heritability can exist, even if the process of transmission of these ideational entities is not analogous to replication. The close ontological analogy, that memes and genes are both replicators in the narrow sense, is neither necessary for claiming that culture is a variational system, nor necessary for claiming that creativity is based on undirected variation: First, the fact that there are memes does not entail that culture is a variational system. A Lamarckian transformational system could also rely on basic countable and replicating building blocks that are maintained over time. Second, it follows from what I have said in chapter 2, that Lewontin's three principles and the principle of undirectedness of variation can hold for culture, even if the units of selection are not replicators in the narrow sense.

### Origin of novelty in culture and creativity

The anthropological concept of culture (introduced in section 3.1) entails the conception that culture is created by individuals, maintained across individuals through social learning and intentional selection of newly introduced or socially learned cultural units. At issue in this chapter is how individuals create culture, i.e., how novelty is introduced into culture by individuals.

To *create culture* in its wide sense means to bring about something that is new as a token. Making a traditional kind of pot is creating culture, even if it is just a new token of a type of behavior (and artifact) that is already well-established in a certain culture. This kind of creating is a *conditio humana* and an ubiquitous activity. That is why I will call such activity *creativity in an anthropological sense*. Whenever we are creative in this sense, we bring about what I would like to call *anthropological novelty*: something is new in an anthropological sense, if it cannot come into existence without our creating activity. Even if a craftsman creates a traditional kind of pot over and over again, he is – in a wide, i.e., in an anthropological sense – creative in doing so. He adds something to the world that can only come about through this socially transmitted activity.

However, since creativity in this sense does not produce *historical novelty*, for instance through creating a *new* kind of pot, it is not the kind of creativity essential for an understanding of cultural *change*. What is essential for cultural change is what I call *psychological creativity.* It is the kind of creativity that is at the basis of each historical novelty.

A cultural unit can be historically new in two senses, relatively and objectively: It can be new for a group of people that are bound together diachronically and synchronically as a tradition (*relative historical novelty*), or even new in an absolute sense, i.e., new in the sense of first appearance in the whole universe (*absolute historical novelty*). The essential thing for this study is that every historical novelty must also be a *psychological novelty*, i.e., new for the person introducing the historical novelty. Something that is new in a historical sense (i.e., something that has never been brought about by *any* person before, in a certain culture or absolutely) must be new for the creative individual as well, since otherwise there would have been at least one person (i.e., the creative person) for whom the alleged historical novelty was not new. The novelty would in fact not be historically new, if it was not psychologically new. A historical novelty must thus always be a psychological novelty. The converse, however, does not have to be true, since something can be novel for a certain person but old in a historical sense. Nonetheless it follows that an explanation of historical novelty is tantamount to the explanation of psychological creativity: If we want to explain historical novelty, we always have to explain the occurrence of psychological novelty.

But in doing so, we have to exclude cases where the psychological novelty derives from a historical 'original,' since not everything that is psychological new, is conventionally called creative in a more narrow sense. If we copy something, we are not creative. Yet the reason is not the *existence* of a historical predecessor, the reason is that psychological creativity excludes the *'copying' from* such a historical predecessor. The exclusion of such a kind of 'copying' is thus a part of the contemporary concept of psychological creativity.

The contemporary folk-psychological and scientific concept of *psychological creativity*, distinguishes creativity (i) from individual learning (social learning and learning-by-experience) and (ii) from routine, 'mechanical' production. These two demarcations are implicit in contemporary psychological definitions of creativity. The creativity psychologists Robert J. Sternberg and Todd I. Lubart give the following definition:

> "Creativity is the ability to produce work that is both novel (i. e., original, unexpected) and appropriate (i. e., useful, adaptive concerning task constraints)" (Sternberg & Lubart 1999: 3).[1]

'Original' hints at (i) and 'unexpected' at (ii). Since these two demarcations are essential for evaluating the analogy between creativity and biological evolution, I will have to go into some detail about them.

*(i). Psychological creativity and learning*. If a psychologically novel idea is merely acquired through the causal influence of something that can legitimately be considered as an 'original,' then the idea is psychologically new, even though it did not come about through creativity in the narrow sense. Creativity in the narrow sense demands what I call *psychological originality*: a partial independence from the causal influences of any kind of 'original,' be it an influence of an 'original' through *social learning*, or through *individual learning-by-experience* of the external world.

Our intuition says the following: Although the potter who makes the traditional kinds of pots may be creative in the anthropological sense, he is not truly creative. But why? The answer normally is: If the potter did not come up with the idea on his own, considering the kinds of pot he makes, but learned to make them *from* others, he is not really creative. Again, it is not the existence of an original but the *copying from* an original that delineates social learning from creativity. If this intuition is correct, then the concept of psychological creativity entails that not every psychological novel idea is brought about creatively, since what I learn from others is – despite being a copy – psychologically new for the learning individual.

What I have said so far also holds for individual learning in the sense of individual learning-by-experience of the external world. In opposition to social learning, individual learning-by-experience is learning something about the external world because one has a *direct* experience of what one learns, whereas

---

[1] The criterion of 'appropriateness' is not central here, but see Kronfeldner (2001, manuscript) or Kronfeldner (2005).

the probability of the experience is not increased by the presence of individuals who have learned the respective thing already. The cognitive processes underlying learning-by-experience also lead to novelty in mind. Yet, individual learning-by-experience is conventionally also considered as opposed to creativity. As Plotkin writes: „Originality and creativity are, after all, defined by the independence of their products form direct experience" (Plotkin 1994: 64). Such a concept of creativity also lies behind Noam Chomsky's stress of the creative aspect of language use (Chomsky 1966: 3-31). According to Chomsky, language is produced "unbounded in scope and stimulus-free" and is in this sense creative (Chomsky 1966: 5). Stimulus-freedom is, however, most important for considering language production as creative, as Chomsky adds, since "[a] tape recorder or a person whose knowledge of a language extends only to the ability to take dictation has an unbounded output that is not stimulus-free in the intended sense" (Chomsky 1966: 77). 'Stimulus-freedom' means that the language production is "undetermined by any fixed association of utterances to external stimuli or physiological states" (Chomsky 1966: 5).[2] If we produce the entities that we produce with our mind only by experience and association of the experiences, then the part of the external world that becomes represented through individual learning-by-experience can also be regarded – with some poetic license –as a kind of 'original' that becomes '*copied.*' I think that is the reason why learning-by-experience is usually regarded as uncreative.

However, two points of clarification are necessary here: First, in this context the term 'copying' has to be taken in a wide sense, not in the sense of replication dealt with in chapters 2 and 3. If an outside world produces any representation of the external world at all in our minds, the similarity between the 'copy' and the 'original' is not the kind of similarity that holds for social learning or even for gene replication. The images, concepts, and knowledge our mind extracts from the external world (i.e., the 'copies') do not belong to the same ontological category as that external world of which a 'copy' is made from. This is also the reason why one clearly has to distinguish between

---

[2] See Chomsky (1966: 3-31) for a history of this opposition between learning and creativity since Descartes.

learning-by-experience and social learning, since in the case of the latter the transmission leads to the same kinds of entities: I ‘copy’ – through inferential reconstruction – the concept of peaceful behavior prevalent in our culture. Second, if I say that creativity is usually opposed to learning (learning from others or learning-by-experience of the external world), it should be clear that this statement depends on how one wants to use the term ‘learning’. If it can be shown that individual animal learning (animal innovation) is stimulus-free to some extent (i.e., not merely induced by an instructive, direct molding influence of the respective environment or due to social learning), then it is a terminological question whether we still want to talk about learning, caused by some special process, or whether we conclude that the respective animals do not learn but are in fact creative.[3] It doesn’t matter how we choose the terms. What does matter is that we see the difference between novelty through psychological originality and novelty through external ‘determination’ or ‘information transmission’ from something or someone. The reason for my choice of terminology is simple. It suits better our everyday usage of the terms: Although we say that I *learn from* experience of the external world or *from* someone, we do not say that I *create from* such an experience or from someone.

The following is an example that can illuminate what the opposition between learning and creativity amounts to. Friedrich August von Kekulé’s discovery that the benzene molecule forms a ring is usually taken as a paradigm example of creativity. The discovery was revolutionary. It made a whole new field of inquiry possible, namely the chemistry of aromatic compounds. The chemical composition of the benzene molecule was known. But its structure was unexplainable by the standards of chemical theory at that time, standards of a theory that were developed independently by Kekulé and Couper. This theory assumed that all hydrocarbon compounds are simple strings (Findlay 1968: 34-41). The problem was that the valences of benzene

---

[3] See Reader & Laland (2003: 16ff) for a list of the processes involved in animal innovation. Creativity is considered to be but one of these processes involved in animal learning. If we define learning very widely, creativity is one way of learning. If instead we define learning very narrowly, creativity and learning become excluding categories.

did not fit into this model. Now, Kekulé's more or less direct experience of benzene molecules did not suffice to find the solution for this problem. He could not learn the structure of the benzene molecule in the sense we learn other things by direct experience, for instance that sugar is sweet. Given his knowledge, he could not 'read off' the structure from the chemical features of benzene molecules. Although the benzene molecules exerted some influence on his mind, leading thus to learning-by-experience of the benzene molecule, this impact did not suffice to discover the structure of benzene. There must have been something else that complemented his empirical knowledge about benzene when he arrived at his new idea.

It is important not to misunderstand the opposition between creativity and learning. The opposition is only a gradual one. That learning was not sufficient for the invention of Kekulé's idea does not mean that he had not learned a lot about chemistry that had an influence on his mind when he thought about the structure. His knowledge as a chemist does not exclude that he is creative, in the same way as training in pottery does not exclude that a trained potter can be creative. On the contrary, Kekulé's previously acquired knowledge was necessary for the idea, necessary in order to come up with the idea and necessary to judge it as appropriate. That creativity is opposed to learning means that there must be *some* independence from the causal influences of an 'original' on its own re-presentation. It only means that Kekulé could not learn the *structure* of the benzene molecule. It only means that a potter did not learn the kind of pot he makes. The novelty in mind must transcend the learned and be independent in this sense only.

This can also be illustrated by the example of a student learning the structure of the benzene molecule at school. Through experiments the student has a direct experience of the object. Through teachers he gets a lot of additional information about chemical compounds etc. But as long as the teacher does not present the solution straight ahead to the student, the student has to be creative *to some degree*. The teacher defines the problems and gives the student everything that is needed – and this is surely much more than Kekulé himself had. Although the teacher can thus trigger the correct answer in the student, the student is creative to some degree, as long as the student finds the solution on his own. Since the student surely started with more ‘given’ knowledge and could use the guiding hints of the teacher, the student is nonetheless much less creative then Kekulé. Psychological originality is not an all or nothing. As I said, Kekulé did not discover the structure of the benzene molecule *out of nothing*. Kekulé also built his idea on experience and on already established knowledge that he learned from others, or that he invented himself beforehand.

To recapitulate this point on creativity and learning, psychological creativity, creativity in a narrow sense, demands psychological originality. It thus stands in a gradual opposition to learning – learning from others and from individual experience.

*(ii). Psychological creativity and routine production.* Psychological creativity also stands in a gradual opposition to routine production. It demands *psychological spontaneity*: the partial independence from the influence of already acquired knowledge of the creative person and thus from the intentional plans of the creative person resting on that knowledge. Even if Picasso’s paintings were not ‘copied’ from others, it could still be that Picasso knew exactly what’s going to happen in his mind when he was painting a certain picture. Imagine that he knew exactly what he is doing at that moment, in the same sense as a potter who creates over and over again the same pot. But would we then consider him as being creative in the moment when he paints this particular picture? Picasso himself is often quoted for having said: “*Je ne cherche pas, je trouve*,” indicating that, in the case of creativity in the narrow sense, originality is complemented by a certain passivity of the creative person. This passivity is the effect of psychological spontaneity – the partial independence of the cognitive processes involved in creativity from the intentional control and plans of the individual, which are based on already acquired or created knowledge.

That being creative in a narrow sense – by definition – presupposes spontaneity has been stressed since Plato stated in the *Meno* (80d-86e) the so-called *Meno problem*:[4] You cannot search for or intend to bring about what you do not know. If you, however, can search for something, then you know it already, it is thus not *really* new to you. Thus, if you don't know what you are looking for, you cannot look for *it*, even if you can look for *a* solution to the problem at hand, and even if you know to some extent the criteria that have to be fulfilled by the solution.

On this basis, the characteristic passivity involved in creativity, of which Picasso spoke, can be interpreted as a peculiar simultaneity: You only know that you looked for *it* at the very moment of insight. As Carl R. Hausman stresses this with respect to the activity of creative artists:

> "The artist begins a creative process without a preconceived plan or concept of the exact complex of qualities in the object which he will create. If he were to start with such a plan, then creation already would be complete in his mind. But the creator does begin with a certain talent and set of established habits of work. At first, he senses that certain elements are required in the future product, but he does not yet know these are. And as he creates, he somehow discovers what he wants to create. He formulates his plan at the same time that he comes to see what that plan is – at the same time that he sees what is required to complete the process he started" (Hausman 1984: 10f).

Hausman writes that if the artist were to start with a plan, "then creation already would be complete in his mind." The creation would not be creative, since the creation would have been entailed in the plan, it would have been already given. This is the reason why we do not call a potter creative who repeatedly makes the same kind of pot. When making the kind of pot that he made already several times, the actual making is entailed in his plan, even if only as a plan or as a possibility.

All this amounts to the following: In open-ended problems, where no already known procedure helps you all the way down to the final product, where these procedures give only rough guidelines, the person has to rely on spontaneous, i.e., unplanned occurrence of the solution. The opposition between creativity and routine production excludes an intentional planful or 'mechanical' production of the solution for a problem. 'Planful' in this context

---

[4] The problem is not only a problem for creativity, it is a problem for epistemology and reappears in philosophy as the problem of induction, how general knowledge can arise de-novo out of experience with a few particulars. See Gamble (1983) or Nickles (2003) on this issue.

means foresight and control: When acting planful, I know what I will do or want to do in order to reach my goal before I do it. Creativity presupposes the absence of such a control; it presupposes psychological spontaneity. But, as with psychological originality, the opposition can only be a gradual one, for the same reason outlined above for psychological originality. Creativity in the narrow sense thus requires a partial independence from already achieved knowledge of the creator. This independence excludes a methodic or routine plan of search. Creativity is not totally independent of these routines but transcends it at the same time.

This gradual opposition between creativity and routine production is not only a conceptual assumption of philosophers, it is a phenomenological datum of psychological research on creativity. Many famous protocols of creativity, taken from the hall of fame of history, show that creativity involves this kind of spontaneity. It is a phenomenological datum that shows up in at least three kinds of phenomena: insight, trial-and-error, and serendipity.

The history of science and art is full of great stories about sudden *Eurekas* – the moment of inspiration, the moment of sudden *insight*. Kekulé's discovery is an often-cited example. Insight appears, however, not only in great stories about creative genius but also in everyday creativity observed and analyzed in the psychologist's laboratories (Sternberg & Davidson 1995). In a famous paper, Poincaré (1982 [1908]) describes his mathematical inventions as based on such flashes of illumination. But he did not reduce creativity to this very moment. He described four stages, which are widely accepted by psychology of creativity as descriptively adequate. The whole process of creative cognition is assumed to consist of multiple overlapping and iterated stages: a preparation stage, an incubation stage, an inspiration stage, and an elaboration and evaluation stage.[5] The core of the overall process is the incubation and inspiration part. It is taken to account for the spontaneity of the

---

[5] Wallas (1926: 79-107) and Hadamard (1954 [1945]) adopted this four-stage-model and made it prominent. See also Csikszentmihalyi (1996: 79-106) and Boden (2004: 25-39) for suitable revision in the light of new data. Note that this stage model does not include a hard distinction between the so-called 'context of discovery' and the 'context of justification': evaluative aspects can be present to a different strength at the four stages.

overall process. The individual cannot intend and cannot control what happens in the incubation and inspiration part. From an intuitive point of view, insight is opposed to will: We can raise our arm by will, but we cannot come up with a solution for a difficult problem by the power of our will. We have to rely on that the ideas come to us.

Apart from insight, there are two other phenomena that can exclude the control of the creative individual over the creative process: Trial-and-error and serendipity.[6] Both involve an unexpected moment of finding the final solution, a passivity on the side of the creative person. There is a lot of trial-and-error in science, art, or other cultural spheres. Ignaz Semmelweis' discovery of the reason for the epidemics of childbed fever in a Vienna Hospital counts as an instance of the first. Alexander Fleming's discovery of penicillin is one of the famous cases of serendipity.[7] The difference between trial-and-error and serendipity is that in trial-and-error you are *looking for a solution to a certain problem*, even if the wish itself doesn't give birth to the solution. You have to try and the trials very often fail. In true serendipitous creation or discovery you aren't even looking for a solution. "Pseudo-serendipity" (Roberts 1989) is somehow between trial-and-error in the narrow sense and true serendipity. Whereas you intentionally produce trials in trial-and-error processes, you do not produce a 'trial' in pseudo-serendipity: You collide with a solution while looking for it. In true serendipity you are not even looking for it. The central mark of serendipity is that the solution was found by acknowledging or taking into account a coincidental event in the outside world. But what characterizes all three phenomena – serendipity, trial-and-error, as well as insight – is the following: it is the impossibility of generating the discovery or solution to a

---

[6] The concept of serendipity goes back to Horace Walpole, a British writer, who wrote in 1754 a letter to Horace Mann, describing a tale with the title "The Three Princes of Serendip". Serendip was the old name for Sri Lanka. The princes make discoveries 'by accident and sagacity, of things which they were not in quest of' (Lewis 1960: 407-411). The term serendipity found its way to standard English language through the work of Walter B. Cannon (1940), a physiologist, and the work of sociologist Robert K. Merton (1967 [1949]: 103-108). See on the history and current usage of the term Diaz de Chumaceiro 1999. See also Merton & Barber (2004) on the history of the term and the concept from the point of view of historical semantics and sociology of science. (The book was written already in the 50s but not published until 2003).

[7] See Boden (2004: 233f) on Fleming.

given problem in a ‘mechanic’ or intentional way. That you want a solution does not help, if you do not know how to find it. Serendipity is well documented.[8] It is as much a phenomenological datum of creativity as is insight. Trial-and-error is often the way scientists or even artists proceed when they explore the frontiers of their domain.

In practice, the distinction between insight, trial-and-error and serendipity is not clear-cut. A trial-and-error process can lead to a serendipitous discovery: It is possible that you look for a solution to a certain problem, you intentionally try one, and you thereby find a solution to another problem. Yet, as I will show in section 4.3, the distinction is important with respect to the question whether novelty occurs in an undirected way. Furthermore, insight can be but does not have to be the source of serendipity. Insight is characterized by an inner complex cognitive event as a part of a long creative process with one or all the four stages mentioned above, maybe iterated several times. True serendipity does not involve such a creative process, it is mainly characterized by a relevant new event in the outside world. Some might even want to exclude serendipity as an instance of creativity, since there is nothing creative in finding something. Others would answer that the finding usually also involves an insight that is creative. Archimedes, the father of the famous ‘Eureka!’, discovered how to find out whether a crown was really made of gold. He knew that if he were able to measure the volume of the crown, he could say whether the crown was pure gold. But he did not know how to measure the volume of such an irregular solid object. When he stepped into the bathtub in the public bath of Syracuse, he realized how this could be done: The volume of his irregular solid body equals the amount of water that overflowed when he stepped into it. He was so surprised so that he ran out of the bath, naked, through the town, shouting ‘Eureka! Eureka!’. Archimedes’ discovery counts for Roberts (1989) as an instance of pseudo-serendipity, since he looked for the solution but did – when he found the solution – not consciously produce a trial-situation in order to test it. Yet at the same time Archimedes’ discovery is one of the paradigm examples of insight. Fleming discovered the Penicillin

---

[8] For a collection of serendipitous discoveries see Roberts (1989).

by true serendipity, yet his discovery is sometimes considered as an instance of creativity and sometimes not. One solution for this interpretational problem could lie in the following: Insight appears as not necessarily relying on external events, while serendipity does. But on the other hand, insight can rely on external events that are incorporated by the creative person during incubation as clues for a solution. I do not want to give a final answer on this categorization problem.

The important points are the following: First, only those cases of trial-and-error that involve psychological spontaneity (and therefore insight) can be counted as creative. Trial-and-error can well be purely mechanical when the trials are produced mechanically and would thus be excluded from being creative in the narrow sense. Second, cases of serendipity might well do without insight and can therefore be doubted to be paradigm cases of creativity. Third, serendipity differs from trial-and-error in that the person was not engaged with a certain task for which it sought a solution.

To recapitulate: I have illustrated in which sense creativity in the narrow sense excludes routine production. Creativity requires thus a partial independence from an intentional plan of the creative individual, a partial independence from already acquired knowledge. We also saw that creativity presupposes a partial independence from the causal influence of an original – from a mental original in other people or from an external original in the outside world. The contemporary folk-psychological and scientific concept of creativity requires psychological originality and psychological spontaneity.

On the basis of this, it stands to question whether this includes important consequences for a psychological explanation of creativity. If Kekulé had just used a standard rule that says that, in the case of such a problem, try the structure of a ring, his idea would not have been creative, since the whole idea would have been just an application of a routine method. If someone had told Kekulé that the structure of the benzene molecule is a ring, this external influence would have made the process in Kekulé's mind uncreative, since he would have 'copied' the idea. If his discovery simply were the outcome of his strictly empirical knowledge of the benzene molecule, the process would not be creative, since he would have just passively re-presented the structure. From this it can be concluded that creativity cannot fully be explained by reference to *learning* and *previously acquired knowledge* alone. That is what makes an explanation of creativity special. Whether the Darwinian approach to creativity can offer a new and correct explanation of how we come up with new ideas will be addressed in the subsequent sections. At this point, it is important to realize that the contemporary narrow concept of creativity excludes a certain explanation of creativity.[9] However, it alone does not give an explanation yet: it simply specifies what the phenomenon to explain is.

Creativity in its narrow sense is one source of novelty in culture. Individual learning-by-experience – as defined above – is another major source. Errors in social learning can count as a further but minor source of novelty in culture. I will concentrate on creativity since it is the phenomenon that has been addressed by Darwinian approaches to creativity, and since it is the phenomenon that is – for conceptual reasons – most likely to fit a Darwinian frame.[10]

## Standing on the shoulders of giants

To attribute the occurrence of novelty to creativity does not necessarily mean that culture changes in big sudden leaps, made by great geniuses. Creativity is normally part of a gradual change, whereas each creative individual stands on the 'shoulders of giants.' A creator of a cultural item usually relies on many previously created ideas, knowledge, or patterns of behavior, although he transcends them at the same time. That is why the opposition between creativity and learning or routine production is only partial. Gradual change therefore means that the origination of an existing cultural item, like the design of a modern watch, cannot be explained by a single creative act. It has to be

---

[9] Some philosophers have gone even further: They deny that creativity can be explained at all, interpreting thereby creativity as demanding freedom or independence of the creative person from *any* causal influences. See Kronfeldner (2005) on this issue.

[10] Learning-by-experience in general and operant conditioning in particular have also been regarded as Darwinian. See for this version of Darwinian analogical reasoning from nature to culture: Skinner (1953), Jerne (1967), Blute (1979), Plotkin (1994) and Hull et al (2001), to name but a few.

explained by a gradual, cumulative process of multiple cycles of origination and diffusion.

Now, to explain the origination of an artifact with the help of an analogy from biological origination of novelty to cultural novelty can simply mean that cultural change is explained as a gradual change. If the thesis that cultural change is Darwinian were interpreted as merely saying that culture changes gradually, then the thesis about the Darwinian nature of culture would be almost trivial in two respects: First, as I illustrated in chapter 2, gradualism is not the distinctive feature of Darwinism. Gradualism as such is compatible with the idea of creation as well as with Lamarckian evolution. Gradualism is not the essential characteristic of the process of change that makes Darwinism so peculiar – compared with creationism and Lamarckism. Furthermore, graduality is not even necessary for a Darwinian picture. If mutations leaded not only to small but big changes in phenotypes, the pattern of change could still be a variational one, and could still contain undirectedness of variation, and hindsighted, opportunistic natural selection. Second, gradualism does not lead to any special thesis about creativity. The gradual nature of cultural change is no contentious issue anymore. Albeit there are still people who believe that certain grand human accomplishments are created solely by one individual, by one big genius, it is commonly accepted that every creative individual stands on the 'shoulder of giants.' Basalla (1988), for instance, could show with impressive evidence that technology usually changes gradually through the accumulation of the multiple creative activities of individuals.[11] It is then an empirical question how many steps were involved in the origination of, for instance, the watch, and how big the single steps were. The answers will vary from case to case. Furthermore, even within one creative individual, creativity is usually not an instantaneous act. Creativity takes time. Thus, to say that culture changes in a Darwinian manner because it changes through small, cumulative steps just reformulates an empirical fact and names it 'Darwinian,'

---

[11] Mesoudi, Whiten & Laland (2004) refer to further empirical evidence, for instance already in Pitt-Rivers (1875) on stone tools and spears, in Cavalli-Sforza & Feldman (1981) on lithic technology, and Wilder (1968) on the gradual emergence of the basic mathematical decimal system.

although a gradual change might as well be described and explained by one of the other two paradigms – creation or transformational evolution – as well.

Last but not least, even if a cultural item originated in a gradual way, it still holds that the individual steps are at the center of any explanation of the change. Gradual change makes the goal of explaining cultural change more complicated, since it divides the origination of a cultural product into many cycles of more or less small steps. At each step, novelty is introduced. Therefore, if we want to explain the origin of a cultural item, we have to look at each particular step. Each step contains either an error in social learning, individual learning, or the creative act of one person, standing at the same time on the 'shoulders of giants.' Consequently, the origination analogy has to be justified as a hypothesis about the *nature of each step* of the cumulative gradual change.

What has to be shown with respect to the steps is not that the steps are small, but that the nature of the steps leads to a Darwinian pattern of change that exhibits some undirected generation of novelty, the hindsighted selection that follows from this, and the kind of shortsightedness that is typical of natural selection – aspects that are essential and distinctive for a Darwinian pattern of change.

## Culture as a variational system

In contrast to graduality, a variational pattern of change is an essential and distinctive characteristic of Darwinism. Does culture change in a variational or in a transformational manner? That culture is indeed variational is not contentious, and often implicitly taken for granted by Darwinians as well as non-Darwinians. In addition, the claim that culture is variational has not been cited as a central aspect for the Darwinian approach to creativity. That is why I will not say much about it. I will take the variational nature of culture for granted, but let me nonetheless explain what it comes down to and refer to some evidence.

The concept that culture is variational rests on the assumption that a change in culture as a system is brought about through a change of the distribution of cultural items. First, there is variation: There are, for example, people in German culture who drive cars and there are some that do not. They ride bicycles. Second, a kind of sorting process – usually intentional selection by conscious or unconscious choice of individuals – is a necessary part of the explanation of a change in distribution of these habits. A transformational picture would have to deny that there is real evolutionarily significant variation in cultures. [12] But this is evidently false, as just indicated with the example of cars and bicycles. That such diversity usually exists is also taken as an empirical fact, for instance, in anthropology, with respect to many aspects of pre-modern societies (Barth 2002, Vayda 1994) or with respect to technological change in modern societies (Basalla 1988).[13]

A transformational pattern is also unlikely because of the following: As illustrated above, the origin of variation lies in the learning and creative activity of individuals. Individuals react differently to the influence of the natural and social environment. Although social transmission, in form of concerted or magistral transmission, and individual learning-by-experience of a common natural and social world can prevent variation to some degree, as shown in chapter 3, there is no overall unity of response, as assumed in Lamarck's concept of a typological and adaptive reaction of individuals to the respective environment. Variation comes about because not every person has exactly the same experiences, even if they are situated in the same selective environment. In addition, if there is some stimulus-freedom in the sense specified above, the associations we make between different experiences and

---

[12] The only evolutionary paradigm that includes a transformational pattern would be the paradigm of classical evolutionism in anthropology, mentioned in chapter 1. This tradition has been developed parallel to, but independently from Darwin's theory of evolution. It assumes that all cultures develop along a determined transformational axis of progress, i.e., civilization. Cultures develop along this axis since they are all essentially similar. For cultural relativists like Franz Boas (1911), there was no such commonality between cultures. Hence, according to him, there is no common destiny and no order of 'primitive' or 'higher developed.' This debate will not be of interest here. It takes whole cultures as the unit of analysis and addresses the overall path of culture. Relevant in this study, however, is the micro-evolutionary process of change that happens within cultures. On classical evolutionism and the concept of culture as transformational in this sense see Carneiro (2003), Sanderson (1990), Rindos (1985), and Fracchia & Lewontin (1999).

[13] See also Borofsky (1987), on the inhabitants of the Polynesian atoll of Pukapuka, and the dual-inheritance-theorists Durham (1991, 2002), Boyd & Richerson (1985, 2005), and Cavalli-Sforza & Feldman (1981). See also Rogers' (1995) diffusion studies, which also model empirical research on diffusion of innovations along variational lines.

conclusions we draw from them might well be diverse, even if they were build on exactly the same basic experiences. If these differences add up in a person, differences in the population result and lead to a variational pattern in culture. These individual differences exist. Culture can therefore be understood as exhibiting variation. Furthermore, the differences add up through multiple cycles of variation and selection, just as in biological evolution, leading thereby to a cumulative variational process of change.

Nonetheless, the variational nature of culture is often overlooked or not taken to be the object of study. The reason is that variation is related to a methodological problem in anthropology. In field observations, variation is often too easily lost by abstraction. A stereotype is falsely taken as the *'object* of observation' (Barth 2002: 28). But this methodological problem should not be confused with what the anthropologist Vayda (1994: 324ff) calls "essentialist bias" and Durham the "essentialist trap":

> "[C]reating the impression that each 'ethnolinguistic group,' for example, has its own distinctive, characteristic 'culture.' Such a move would be a step backwards, contributing to the false illusion that there is one uniform culture common to all people who, for instance, speak a given language or who occupy a particular geographic or political area" (Durham 2002: 194).

Although culture is often misguidedly understood as a homogenous, shared whole, culture is in fact more like an interbreeding population, consisting of many different cultural items that exhibit different frequencies in the culture-bearing individuals of a society. As the anthropologist Tim Ingold (1994: 330) writes, "[w]hat we do not find are neatly bounded and mutually exclusive bodies of thought and custom, perfectly shared by all who subscribe to them, and in which their lives and works are fully encapsulated." Hence, that culture is variational is evident in the fact that not every part of culture is shared by everybody who is a member of that culture. A person might be a member of a culture, although he does not share all the items that appear in this population of ideas. That culture is shared means that there is a close idea-transmitting interbreeding interaction between people, whereas the boundaries are only

> "relative barriers to social transmission for the specific cultural information under study: natural barriers (rivers, lakes, mountains, etc.), language barriers, social barriers (as may come with various forms of inequality, such as class or caste), and the like. […] A large and heterogeneous society can

> thus be subdivided into pertinent 'reference groups' within which individuals have similar sociocultural constraints and opportunities" (Durham 2002: 195)

That culture is shared does thus not mean that *everyone* thinks and does exactly the same. There certainly has to exist a homogenously shared basis. Without such a common ground, there would be no understanding and no transmission, hence no 'reference group,' as Durham calls it, and no culture as an interbreeding field of ideas. Homogeneity is ensured through enculturation, as well as imposition and conformity pressures. Homogeneity is essential for culture, but it is not total.

In short, culture is variational; it changes through a sorting process. However, even if culture is variational, cultural change can still be different with respect to the way novelty arises in individuals. As explained in section 2.3, a variational pattern can still be based on *guided variation*, due to *coupling.* Furthermore, the sorting process does not have to be natural selection.

### Intentional selection

The cumulative variational process of cultural change depends not only on the occurrence of novelty. For *cultural* change to occur novelty must be selected by the creative person, must be presented to the public, and must be accepted by some others. Selection of cultural items happens thus first in the learning or creative person, and then in other individuals who merely adopt the item and thereby determine the subsequent diffusion of the item. Through iteration of this dual process of originating steps and sorting, cumulative cultural change can occur.

The selective processes involved in such a cumulative evolution are evidently governed by plans and knowledge of the selecting individuals. Humans select by artificial, intentional selection as presented in section 2.3. The selection of cultural items by humans differs from natural selection in that it does not necessarily rely on the same kind of *opportunism* as natural selection: (i) Humans are able to have higher long-term goals and can thus globally maximize their decisions. They can make their decisions according to a greater outlook, according to a kind of master plan for a certain long-term project. (ii) Humans can thus also take into account positive effects that will only occur in the future – effects that have no positive effect at the moment of decision. They do this by predicting these future effects, something nature cannot do. They select by *predicting* outcomes, whereas nature selects by momentary *actual* outcomes; (iii) Humans are not necessarily bound to tinkering, when they create novelty. They can abstract from a given context and go back in the history of a development. They can do this, since, contrary to nature, they have a memory that enables them at the same time to build on already achieved cultural items and to ignore them. Furthermore, they can also take parts of certain complex cultural items and can ignore the rest, something nature cannot do either.

In chapter 5 I will say more about intentional selection as a cause of diffusion. Here we are only concerned with the selective activity of the creative person who brings about cultural novelty. The just mentioned differences between intentional and natural selection are usually uncontroversial. They are implicitly taken for granted by critics as well as defenders of Darwinian approaches to creativity, but not cited for or against the analogy. That intentional selection is as opportunistic as natural selection has never been part of the origination analogy at issue here and is thus not central for an evaluation of the analogy.

The following, however, is central and is highly controversial: As I will show in the next section 4.3, defenders of a Darwinian approach to creativity have claimed that the selective part of human creativity (i.e., the act of adopting your own ideas) rests on *hindsighted* selection, since creativity can, as natural evolution, produce novelty only by *blind* variation. There is no real foresight in human creativity with respect to the appropriateness of the products of our minds, even if we try to predict the appropriateness, and even if we respect expected long-term payoffs.

The latter claim – that creativity is based on *blind* variation – has been strongly criticized. As the deity introduces novelty in a directed way through wisdom, plan, and purpose, critics claim that human creativity is equally based on wisdom, plan and purpose, even if humans are not as perfect as an alleged creator deity. Whether creativity is blind or not is the most frequently reoccurring and hotly debated issue with respect to Darwinian thinking and the origination of novelty in culture. For this reason the issue about blind variation, to which I now turn, also stands at the center of my analysis.

## 4.2 BLIND VARIATION IN CREATIVITY

### Creativity as blind-variation-selective-retention

The debate about creativity as based on blind variation is at the center of the Darwinian approach to creativity. It had its starting and reference point in Donald T. Campbell's famous article *Blind Variation and Selective Retention in Creative Thought as in Other Knowledge Process* (Campbell 1987 [1960]). In this as well as in further articles, Campbell tried to defend that creativity is a "blind-variation-selective-retention process." If we apply Darwinian thinking to creativity, 'variation' stands for the *process* of generating novelty. Generating novelty can mean either the modification of existing, or the creation of new cultural items. 'Selective retention' stands for intentional selection of the generated items. The problematic part of the formula 'blind-variation-selective-retention' is the term 'blind.' I have explained the concept of blind variation for biological evolution in chapter 2. It means undirectedness, decoupling of the factors that cause novelty from the factors that select novelty.

The application of this concept of 'blind variation' to creativity has caused some confusion, partly because the meaning of blind variation for biological evolution has not been made clear either. As Hull complains, "[t]he characterization of the variation that functions in selection processes has been one of the most contentious topics in the literature – and the most frustrating. It seems that no adjective exists in the English language that accurately reflects the sort of variation that occurs in selection processes. Is this variation blind, chance, random, nonprescient, nondirected, nonteleological, unforesighted, what?" (Hull et al 2001: 513). Further candidate terms that are used now and then are 'unjustified,' 'wasteful,' or 'unconscious.' Campbell (1987 [1974a]: 56f; 1974b: 147-152) himself deplored the confusion. But Campbell is partly responsible, since he was not really clear about what blind variation means, as I already illustrated in section 2.3. To approach the issue what blind variation means for creativity, let me start with the central claim of Campbell's Darwinian approach to creativity.

Often quoted, Campbell stated:

> "[I]n going beyond what is already known, one cannot but go blindly. If one can go wisely, this indicates already achieved wisdom of some general sort" (Campbell 1987 [1974a]: 57).[14]

According to Campbell, although no blindness is involved in most cases of intellectual achievements, if we go "beyond what is already known," i.e., if we produce novelty, it must be generated blindly. *At the frontier of art, science or what so ever, creativity is necessary and creativity is necessarily Darwinian.* According to Campbell, that creativity is based on blind variation is thus an "analytical truth" (Campbell 1974b: 142).

We saw in section 4.1 that creativity in the narrow sense is characterized by psychological spontaneity. Psychological spontaneity consists of a partial independence of the cognitive process from the intentional control and plan of the individual who exerts his control through constructing a plan on the base of already acquired or created knowledge. This independence is required for creativity because of the gradual opposition between creativity and routine production. Furthermore, as shown above, such a partial independence entails the absence of *foreknowledge* of whether the trials that we emit are worthwhile. That is the reason why Campbell says that his claim is analytic.[15] If one merely reproduces in a routine production procedure what one knows already, as the potter, who makes the same pot over and over again, then one is

---

[14] Defenders of the Darwinian approach to creativity continuously repeated this 'analytical' claim, without much elaboration, especially when they get under attack about 'blindness'. See for instance the critical defense of Perkins (1998: 181), the strong and detailed defense of Cziko (1995: 289, 295; 1998: 194) or Simonton (2003: 316). Simonton rests content with merely paraphrasing Campbell's claim: "Something more must be added to take the creative mind beyond the limitations and constraints of that expertise, to generate truly original ideas that go beyond what has worked before."

[15] In a strict sense, it is no analytical truth, although the claim rests on the narrow concept of creativity. If we do not conceptually rule out that there is a possible world where muses or a omnipotent, all-knowing supernatural deity is reliably inspiring the creative person, then novelty can be brought about in a directed way by the inspiring source. See Gamble (1983: 359) for a similar point, but one that ignores psychological originality as an essential component of creativity in the narrow sense.

not creative and one has pretty good foreknowledge that the trials are appropriate, because one has seen the consequences of the 'trial' several times already. Indeed, what one does is not even a trial anymore; it is a mere reapplication of a recipe of which one knows that it is not a 'maladaptive' one. If one has such kind of foreknowledge, one is not creative. As Hausman (1984: 10) wrote, if an artist were to start with a "plan, then creation already would be complete in his mind," and he would not be creative. The alleged creation would be entailed in the plan, something already given. This is the reason why we exclude the craftsman from being creative. In making the kind of pots, which he made already several times, the actual making was entailed in his plan, although only as a plan or as a possibility.

This would be comparable to the perfect activity of a deity who creates, like a craftsman – by wisdom, plan and foreknowledge. Since the deity is perfect he knows everything in advance. Thus, in the sense of psychological creativity, defined as it is today, this deity cannot be creative in the strict sense. He is a mere uncreative craftsman. However, since humans do not have perfect foreknowledge, they have to rely on psychological spontaneity. Campbell thus seems to merely reaffirm that creativity involves psychological spontaneity, a partial independence from previously acquired knowledge.

What has, however, caused some of the above-mentioned confusion about blindness, is that Campbell sometimes illustrates his claim with respect to 'foresight' and 'prescience,' and sometimes with respect to 'wisdom,' 'existing knowledge,' and 'intelligent choice.' Compare the following variations on the argument: In his famous article from 1960, *Blind variation and selective retention in creative thought as in other knowledge process*, he writes:

> "Real gains must have been the products of explorations going beyond the limits of *foresight* or *prescience*, and in this sense blind" (Campbell 1987 [1960]: 92; Emph. added).
>
> "[I]nsofar as thought achieves innovation, the internal emitting of thought trials one by one is blind, lacking *prescience* or *foresight*" (*ibid.*: 96; Emph. added).

In *Unjustified Variation* (1974b) he writes:

> “The natural selection epistemology here offered has one special analytic feature: if one is expanding knowledge beyond what one knows, one has no choice but to explore *without benefit of wisdom* (gropingly, blindly, stupidly, haphazardly). This is an analytical truth central to all descriptive epistemologies of the natural selection variety” (Campbell 1974b: 142; Emph. added).

> “[I]ncreasing knowledge or adaptation of necessity involves exploring the unknown, going *beyond existing knowledge* and *adaptive recipes*. This of necessity involves unknowing, non-preadapted *fumbling in the dark*” (*ibid.*: 147; Emph. added).

That there is no foresight or prescience just says that we do not know whether the trials we emit are worthwhile or not. They are trials, guesswork, and psychological spontaneity is involved. But does exploring the unknown mean that we bring about the trials “without benefit of wisdom,” that we are “fumbling in the dark,” that we do not have plans and methods that help us find our way through the country of the unknown that lies beyond the frontier of already acquired knowledge? This is the essential question that will guide my subsequent analysis. It will lead me to a multi-layered critique of Campbell’s central claim.

What I want to show is that although psychological spontaneity can be interpreted as a certain form of blindness, this does not mean that we bring about ideas ‘without benefit of wisdom’ and in this sense blindly. To show this, I will distinguish between three different forms of blind variation in creativity. The first, blind variation as *random variation*, is clearly not the way novelty occurs in creativity; the second, blind variation as *unjustified variation*, is a kind of blindness that I think Campbell meant with his claim – and one that he should have meant, if the analogy between creativity and biological evolution should be *correct*; the third kind of blindness, blindness as *undirectedness*, is, however, what he should have meant, if the claim should lead to a *close analogy* between creativity and biological evolution. By clearly distinguishing between these three interpretations of the concept of blind variation in creativity, I hope to bring some order in the confusion about blind variation in creativity.

## Blind variation as randomness

Campbell wrote, for instance, that blindness of variation means:

> "[T]hat the occurrence of trials individually be uncorrelated with the solution, in that specific correct trials are no more likely to occur at any one point in a series of trials than another, nor than specific incorrect trials" (Campbell 1987 [1960]: 93).

In most cases we do not produce ideas in such a random way. There is no equiprobability of trials. If a chemist is engaged with a problem about chemistry, it is very unlikely that he comes up with some totally absurd idea that is totally unrelated to chemistry. A good idea is more likely. This holds at least as long as he is mentally sane. To take another example: We do not write a creative novel in the way the proverbial monkey from chapter 2 does. In the case of the monkey, each trial of the target 'ME THINKS IT IS LIKE A WEASEL' is equally likely. The monkey indeed produces randomly, "without benefit of wisdom (gropingly, blindly, stupidly, haphazardly," as Campbell (1974b: 142) would say. The monkey randomly produces trial after trial and reaches his goal by 'brute force.' It is usually beyond dispute that we are not producing our ideas blindly in this sense. The question whether we produce a new idea randomly in this sense might not even be applicable as Popper once said, even though he was a defender of Campbell: In cases of creativity, or in any other complex choice situation, the range of possible options is not given in advance and "since we do not know the elements of this range we cannot attribute probabilities to them, which we should have to do in order to speak of randomness in any clear sense" (Popper 1974a: 35).

Although the apparent non-randomness of creative ideas is often used as an argument against the Darwinian approach to creativity,[16] it would be unfair to criticize the approach because of that. This kind of blindness is not even required for organic evolution, as I illustrated in section 2.3. Although the 'ME THINKS IT IS A WEASEL'-typing monkey would be a perfect example of truly random variation, this randomness can only count as an ideal reconstruction of Darwinian evolution. This ideal is not even fulfilled by real

---

[16] See for instance Sternberg (1998: 163-165), Perkins (1999: 348), or Schooler & Dougal (1999: 351).

organic evolution. In actual organic evolution there is no equiprobability of variation. The probability of mutation and combination is determined by many factors; there are different mutation pressures; and the range of variation is developmentally constrained by the history of the species. What is, however, fulfilled by organic evolution is that the probability of adaptive trials is not increased *because* of its adaptivity: Novelty is undirected (i.e., decoupled, adaptively unbiased).

Campbell is, however, in part responsible for being criticized because of the non-randomness of creative ideas, since he indeed said that equiprobability is what blindness means. However, in the same paper from which the above quotation is from, he said that "equiprobability is not needed, and is definitely lacking in the mutations which lay the variation base for organic evolution" (Campbell 1987 [1960]: 92).[17] Dean K. Simonton (1999a: 27) wrote the following, in order to clarify what he and Campbell mean by "essentially blind":

> "By this qualifier Campbell did not insist that the variations be *absolutely random*, although they may be. He held only that the mind eventually reaches the point where it has *no a priori basis for knowing which ideational variations will prove most effective*. Neither prior experiences nor current environmental circumstances will provide sufficient clues about how to *restrict the range of choices*, nor does there exist any rationale for assigning useful priorities to the various alternatives." (Simonton 1999a: 27; Emph. added)

This says that the origination of novelty, leading to new variants, does not have to be random in order to count as blind. However, the meaning of blind variation is still ambiguous. Again, is it the same to say, first, that we do not have an "a priori basis for knowing which ideational variations will prove most effective," and, second, that actual trials are unrestricted by previously acquired knowledge, that the actually trials are reached 'without the benefit of wisdom,' as Campbell said? I will show that there is a clear difference between the first and the second.

---

[17] See also Campbell (1987 [1974a]: 56; 1974b: 148) or Blute (1979: 44-45) defending Campbell.

### Blind variation as unjustified variation

Campbell[18] claimed, as did Popper,[19] that science rests on unwarranted hypothesis formation and that it proceeds in this sense 'blindly.' As Popper puts it, "in the quest for new knowledge," we are in "the proverbial situation of a blind man who searches in a dark room for a black hat which – perhaps – is not there" (Popper 1974b: 1061). That is why Campbell said in 1974 that 'blind variation' first of all means "unjustified variation" (Campbell 1974b). This meaning was entailed already in 1960, where he wrote that blindness means that

> "variations are produced without prior knowledge of which ones, if any, will furnish a selectworthy encounter" (Campbell 1987 [1960]: 92).

For Popper and Campbell science crucially depends on hypotheses that are unjustified trials, or "conjectures" as Popper prefers to say. They are selected after they are proposed. Knowledge acquisition is a process of unjustified variation, followed by selection: It is assumed that the generated hypotheses have no warranty of being correct by their origin. They have to get their legitimation by something else. Popper thought that the trials of science get their legitimation not by their origin, but by surviving the method of rigorous falsification. For him, falsification was analogous to natural selection – a process of error elimination. The fact that beliefs are unjustified prior to selection is the reason why Campbell called his and Popper's program 'evolutionary epistemology.' How, and if at all, beliefs can get their justification, if not by their genesis, is an epistemological question that is irrelevant for this study.[20]

Here it is only relevant that Popper's and Campbell's approach contain the claim that the origin of theoretical knowledge is creative (i.e., involves psychological originality and psychological spontaneity). According to Popper, man is creative and creativity is opposed to what he has called

---

[18] See Campbell (1987 [1960], 1987 [1974b], 1974b).

[19] See Popper (1972, 1974b, 1984).

[20] See Nickles (2003) for an extensive survey on the impact of the claim about unjustified variation on 'generativism' in epistemology. Generativism claims that hypotheses receive their justification from their origin in a process of directed imprinting by the environment.

"observationism," "bucket theory of the mind" (Popper 1984) or "inductivism," which he judges to be parallel to "Lamarckism" (Popper 1974b: 1061). Recall from chapter 2, Lamarck explained local adaptation by pointing to the molding influence of the environment. Through this influence organisms were thought to directly adapt to their environment. As Lamarck conceived evolution, selection was not necessary for evolution, since each organism was thought to interact with its environment directly and to reach 'harmony' by this direct interaction. For Darwin, on the contrary, adaptive *and* non-adaptive features of organisms arise in a population. After repeated selection of the good ones, this indirect trial-and-error-process leads to adaptations. Popper transfers this contrast between Lamarckian and Darwinian evolution to science: Hypothesis formation cannot be reduced to gathering "'direct knowledge' of anything immediately 'given';" we are not "passive recipients of information impressed upon us from outside" (Popper 1974b: 1061). Instead, we *create* ideas and select them afterwards. Therefore, he concludes: "[W]e must abandon any approach which starts from sense data and the given, and replace it by the assumption that all human knowledge is fallible and conjectural. It is a product of the method of trial and error" (*ibid:* 1061).[21] In a nutshell, Popper states that psychological originality is essential for all knowledge acquisition processes and that they are creative in this sense.

Whether creativity should be defined as requiring psychological originality is one question. I claimed that to define it in this way is intuitive and mirrors the usage in psychology and philosophy. Whether all knowledge acquisition is creative in this sense is, however, a different question, which I will not answer here. I will thus not discuss whether and in which sense all knowledge in science is Darwinian in this sense (i.e., unjustified by its origin). I will only explain what this has to do with 'blind variation' and evaluate the claim that creativity is, because of this, analogous to natural selection. I will (i) show that if creativity in this sense exists, then it is indeed analogous to a

---

[21] Popper is not the first who stated this contrast. It goes back to William James (1979 [1880]) who accused Herbert Spencer of being a psychological Lamarckist.

Darwinian pattern of change, but I will also claim that (ii) this is a relatively trivial claim.

*(i). Creativity as a Darwinian selection process.* First, irrespective whether creativity occurs in science or elsewhere, ideas do fail and creativity thus involves selection. In contrast to Popper, Campbell laid more stress on psychological spontaneity. At the frontier of science, art or any other cultural domain, we cannot but go blindly, i.e., try ideas without foresight of which one will prove worthwhile. Ideas can fail, and since we often do not know in advance whether they will fail, creativity is *necessarily* selective and in this sense Darwinian. Whatever the origin, we indeed come up with ideas that are not appropriate for our diverse goals, be it in science, or in art, or in any other cultural domain. This is a bare fact we know from history. Consequently, in order to maintain only the 'adaptive' ones, we have to weed out the 'maladaptive' ones. "*Il faut cultiver notre jardin*," says Candide at the end of Voltaire's ironic persiflage on beliefs in the best of all possible worlds, guaranteed by God's wisdom and benevolence. Hence, as a matter of fact, there are sometimes ideas in our mind that are unjustified or 'maladaptive' by their origin for the purpose of, for instance, acquiring knowledge. They can arise and do arise over and over again. We do make mistakes. In this sense, it can be concluded that the origination of novelty in culture is parallel to the origination of novelty in organic evolution: It often demands for a hindsighted selection process. If the world, or something else, guaranteed the appropriate novelty to occur, we would not have to care about rigorous selection of the arising ideas. We would just have to produce them.

Thus the presence of false trials and hindsighted selection is in a way sufficient to exclude two kinds of accounts of creativity: First, it excludes that all novelties are produced in the sense of an automatically acting Lamarckian mechanism of directed generation of novelty. Second, it also excludes that we bring about new ideas like a perfect creator, who has foresight about the adaptedness of what he brings about. Such a foresight would make selection as superfluous as a Lamarckian mechanism of adaptation or any guarantee of appropriate ideas for a certain goal. But such a kind of 'sightedness' can safely be excluded for humans. If we had perfect knowledge and prescience, we would never produce false trials. We do produce false trials because of our lack of real foresight. The generation of novelty in creativity is in this sense blind. It is a selection process as Darwinian evolution is a selection process.

But note that the claim that creativity is a selection process should not be confused with the claim that creativity is a variational process, where certain ideas of a 'population of ideas' change their frequency in the mind of an individual. Nature always works in the manner of parallel processing, working with many organisms as trials, which are all at the same time subject to selection. Human minds are usually not engaged in such a kind of parallel processing, where massive diverse variants are simultaneously tested and where, as a consequence, the best variants increase in frequency. We do not breed ideas in our minds that then increase their frequency. In other words, ideas do not get 'babies.' This is an important difference to other Darwinian systems, like the immune system or the genetic algorithms of evolutionary computing. Both heavily rely on massive parallel processing and differential breeding. Certainly, in problem solving, we can compare different ideas that seem to be candidate solutions for one and the same given task. In this sense creativity can surely be variational. But sometimes we do not create this variety synchronically but only diachronically. Creativity is – as Nickles (2003: 63) says – an instance of "serial evolution," if it is Darwinian at all. We have an idea and try it out. If it fails, we often have to look for an alternative that is not at hand. Hence, we try another yet unwarranted idea. This procedure is evidently a selection process, although not a variational process at the individual level. The variational pattern arises only at the cultural level, where a new idea, if submitted to the public, can spread differentially in the public domain.

*(ii). To consider creativity as a selection process is relatively trivial.* My second claim about blind variation as unjustified variation is the following: The claim that creativity is a selection process is a *relatively trivial claim* that has to be distinguished from the stronger claim that this creative generation of novelty in mind is 'blind' in a more challenging sense: First, the existence of

false ideas is no news. That ideas can fail and that we do not have a foreknowledge of which one of our ideas will fail, is not an insight that becomes only visible from an analogy between creativity and natural selection. We know it from history. Hence, we do not need Darwinian analogical reasoning from nature to culture to know it. If the origination analogy is merely based on claiming that creativity involves selection, then it is in this sense heuristically trivial. It does not say anything new.

Second, such a claim is trivial in another sense. The concept of unjustified variation simply mirrors the definition of creativity as resting on psychological originality and spontaneity. Please recall that Campbell said that it is an analytical truth that in creativity, exploring the yet unknown, we can only proceed blindly, since we have to go beyond the knowledge we have reached so far. As explained above, this is the same as defining creativity in a narrow sense, as requiring psychological spontaneity. To see that creativity requires psychological spontaneity, we – again – do not need Darwinism. Furthermore, even if it is correct that humans are creative when they produce novelty without perfect foreknowledge, and that they are in this sense creating ideas blindly, we have still not learned anything about *how* novelty then occurs in our minds by pointing to the analogy between creativity and the evolution of organisms. The analogy cannot *explain* how we manage to be creative in the absence of foresight. The claim that novelty in mind is unjustified can only negatively exclude a crude externalist or creationist explanation of novelty in mind, since the former would contradict psychological originality and the latter psychological spontaneity. As I said in section 4.1, creativity in the narrow sense cannot fully be explained by reference to social learning, experience, and previously acquired knowledge alone. Since the Darwinian approach to creativity mirrors this concept, it can exclude a certain explanation of creativity, but it thus has not yet offered an explanation. If it does not offer itself such an explanation, the analogy has almost no explanatory force and is in the explanatory sense *trivial*, merely assuming what others assume for the concept of creativity. I will, however, show in section 4.5 how Simonton nonetheless tries to give such an explanation in Darwinian terms.

Last but not least, the claim that there are unjustified ideas is relatively trivial, because it is weaker than the claim that creativity is blind in a more fundamental sense, namely based on *undirected* variation, in the sense biological evolution is claimed to be undirected. Neo-Darwinism states that biological evolution is an instance of *decoupled* variational evolution. It does not only state that evolution is an instance of variational evolution, where selection is necessary for evolution to take place. It states that variation arises in an undirected way. Thus the concept of blind variation as 'unjustified variation' has to be distinguished from the stronger claim that creativity is based on truly undirected variation, in the sense of a decoupled setting of variation and selection, as explained in section 2.3. Only the latter would lead to a strong or close analogy between creativity and evolution.

### Blind variation as undirected variation

As just mentioned, blind variation in biological evolution means the absence of a statistical bias towards adaptivity. In blind, undirected variation, the factors bringing about novelty have to be *decoupled* from the selecting factors. Transferred to creativity this means: *Creative hypothesis formation is analogous to biological evolution only if it is based on decoupled change in the mind of an individual*. It is only analogous to biological evolution, if the occurrence of new ideas is *not* influenced by factors that determine the selection of these new ideas. Again, variants do not have to be produced at random in the statistical sense. This is not required for biological evolution either.

Whether Campbell has meant this kind of blindness is not easy to say. It is suggested that sometimes he did mean this kind of blindness, for instance, when he says that "if one is expanding knowledge beyond what one knows, one has no choice but to explore *without benefit of wisdom* (gropingly, blindly, stupidly, haphazardly)" (Campbell 1974b: 142; Emph. added). This sounds like demanding decoupled, undirected variation. However, demanding that "[r]eal gains must have been the products of explorations going beyond the limits of *foresight* or *prescience*, and in this sense blind" is more analogous to merely demanding unjustified variation (Campbell 1987 [1960]: 92). In 1974, Campbell wrote:

> "While most descriptions of discovery and creative processes recognize the need for variation, the present author's dogmatic insistence on the blindness of such variation seems generally unacceptable" (Campbell 1987 [1974a]: 57).

Simonton, defending Campbell, admits the same: "[a]lthough it is obvious that the creator must engage in selective retention, the notion that the creator also must generate 'blind' variations appears less so. This latter idea seems to run counter to the common assumption that creativity is a manifestation of intelligent behavior" (Simonton 2001a: 555). These latter two statements suggest that Campbell meant more than 'unjustified variation' when talking about blind variation.

Be it as it may, blind variation in the strong sense of undirected, decoupled evolution is at least what critics thought that he means, and, as I would like to add, what he *should* have meant, if creativity should be considered in *close* analogy to biological evolution. Undirected variation is, after all, what blind variation means for biological evolution and what critics have objected, while unjustified variation is usually taken for granted.

## 4.3 THE CRITIQUE OF GUIDED VARIATION

### Guided variation at a populational level

Critics claimed that our orientation towards certain problems makes variation in cultural novelty directed, be it creative novelty in the narrow sense or not. In culture, we have guided variation, not undirected variation. Since this basic objection exists in different variants, I will present two variants of it, in order to clarify what exactly provides a problem for Campbell's alleged claim that creativity is based on truly undirected variation.

Boyd & Richerson (1985: 81-98), for instance, state that culture is not dominated by blind, "random" variation. According to them, only misremembering or error-prone reconstruction in social transmission leads to random "cultural mutation." The dominant source of change, however, consists not in copying-errors, but in different kinds of problem-solving, individual learning processes, creative as well as non-creative ones. These are summarized as directed, "guided variation" (Boyd & Richerson 1985: 82). The processes underlying guided variation are not random, since humans "have objectives or guiding criteria, that allow them to rank possible outcomes of their behavior" (Boyd & Richerson 1985: 94). They assume that after an individual has solved a problem the output is "usually favorable" (*ibid.*: 82), precisely because of a decision process in the problem-solving individual that is guided by the guiding criteria. If such individually selected variants are then

> "culturally transmitted, the result is a force that increases the frequency from one generation to the next of the same variants whose frequency is increased within a generation by learning" (*ibid.*: 82).

This "force" is described further in the following way:

> "We call this the force of 'guided variation.' Early evolutionary theory placed great weight on this force, often attributed somewhat erroneously to Lamarck. Darwin […] stressed its importance under the rubric of 'the inherited effects of use and disuse.' Although students of genetics have all but ruled out 'Lamarckian' effects in that system of inheritance, it is likely that they are important in the case of culture" (*ibid.*: 82).

This can be reconstructed in the following sense: The 'usually favorable' cultural item, created and selected by one person, is fed into a cultural pool and is thus already biased towards 'adaptivity' when it enters this pool. This already directed novelty, guided by a process occurring in the creative individual, is then inherited by social transmission. The result is a 'force' that makes cumulative 'adaptive' change possible: the 'inheritance of *acquired* characteristics', i.e., the inheritance of new, '*usually favorable*' cultural items that have been introduced by this or that individual. What is learned in one generation of individuals of a society is transmitted to and maintained by the next generation, if adopted by the individuals of that generation. In this way, this force "increases the frequency from one generation to the next of the same variants whose frequency is increased within a generation by learning." However, the inheritance of newly created cultural items is not guided or Lamarckian because of the *inheritance* of these novelties, as Boyd & Richerson seem to suggest. It is guided, because of the *pre-selection* in the mind of the person who introduced the new item into the population. It is guided because at the moment when the innovation is fed into the cultural process, the novelty is *directed already*. In Lamarck's theory of evolution, what is fed into the transmission process is directed through instructive processes acting on phenotypic characteristics; it is directed *before* these characteristics are transmitted to the next generation. In Darwinian evolution, mutations are *not* directed at the moment when they are fed into the biological transmission process. Guided variation thus indeed provides a disanalogy between cultural change and biological evolution.

But guided variation in Boyd & Richerson's sense only means that the guiding criteria in the minds of people lead to a directedness *at the populational level of culture,* because of a guided *selection* at a *cognitive level*. It does not mean, and this is very important, that already the *occurrence* of new ideas is directed or guided in the sense Lamarck assumed for physiological characteristics. It can thus not be used to object Campbell's claim that creativity is based on undirected *occurrence of novelty* at the *cognitive level* – novelties that are then selected according to 'guiding criteria.' Guided variation in Boyd & Richerson's sense cannot be used against Campbell, since the argument does not show that the *occurrence of the variants at the cognitive level* is *made more likely by the presence of a selective environment*. This is, however, what is required for directed variation as defined for biological evolution (see section 2.3). Boyd & Richerson's argument merely shows that a selective process at the cognitive level leads to pre-selected ideas that enter the *cultural* selection process as pre-selected ones. Thus, defenders of a Darwinian account of creativity could answer that the *selection process at the cognitive level* is nonetheless a Darwinian process based on *undirected occurrence* of novelties – a process that only leads to 'guided variation' at the end of the cognitive selection process.

### Guided variation at the cognitive level

Michael Ruse writes that "cultural evolution must be necessarily different from biological evolution because the raw units of culture are introduced with a purpose in a way quite different from the random appearance through mutation of genes" (Ruse 1998b: 405). In 197 already, Ruse made the same point with the example of Semmelweis, who discovered that the epidemics of childbed fever in a Vienna Hospital could be stopped by rigorous hygienic measures of hand washing. Semmelweis' discovery looks like a serendipitous finding, but, as Ruse points out,

> "it cannot be said that the knowledge occurred 'randomly' – Semmelweis spent several years thinking and working hard, proposing and testing different hypothesis before he hit on the right one. Thus it would seem that this example strongly supports my claim that not all cultural elements – in particular, not all adaptive cultural elements – occur at random. Indeed, I suspect that most new cultural elements appear because they are needed" (Ruse 1974: 432).

Let me assume, for the sake of argument, that Ruse meant with randomness the undirectedness assumed for biological evolution, a form of what Boden called r-randomness (see section 2.3). Now, the long time of Semmelweis' engagement with his problem does not make the discovery less directed or more directed. The essential point is that Ruse claims that 'most new cultural elements appear *because* they are needed.' This is why Ruse mentions Fleming's discovery of penicillin as an example of random occurrence (*ibid.*: 432), in contrast to Semmelweis' discovery. The difference between Semmelweis and Fleming is that Fleming's discovery can count as truly serendipitous, while Semmelweis' discovery is an instance of trial-and-error: of the two, only the latter looked for a solution for the respective problem. Fleming did not look for penicillin.

Paul Thagard goes in the same direction, addressing the issue with respect to problem solving in a more general sense. He interprets Campbell as using a "false dichotomy" between 'blind' and 'prescience,' and goes on to say that "[t]ruly blind variation never occurs." In problem solving,

> "[t]here is no prescience […], since nothing guarantees that the structures activated [in problem solving, MK] will lead to a solution to the current or future problems. But variation is clearly not blind either, since formation of concepts and rules that may be useful in solving a problem is more likely to occur during the attempt to solve that problem" (Thagard 1988: 104).

Thagard's claim is not only that useful ideas are *more likely*. The claim is that useful ideas are more likely 'during the attempt to solve that problem.' They are thus more likely to occur *because we look for them*.[22]

In section 2.3, I explained that directed mutations are defined in an analogous way: Mutations are directed if they occur more likely *because* they would be useful in a given environmental context. Ruse and Thagard differ from Boyd & Richerson in addressing the issue at the cognitive level, claiming that creative problem solving is directed, since there is an *adaptive bias* involved: Our orientation towards certain problems makes the occurrence of appropriate solutions more likely.

However, both arguments still leave open *how* the need or orientation towards a certain problem can bias the occurrence of trials towards adaptivity. They do not show from where the bias comes from. In the following I want to make explicit in which sense orientation towards a problem leads to directed variation due to *coupling*, due to a '*Lamarckian correlation*' between producing and selecting factors. It is this kind of directedness that is excluded by the Neo-Darwinian concept of undirected variation.

### Guided variation due to coupling

When we are looking for a solution to a problem, then we project previously acquired knowledge onto the problem. Fleming's knowledge could not make the occurrence of penicillin more likely, since it was not 'projected onto,' not directed at that specific problem. If someone serendipitously finds a solution for a problem, like Fleming the penicillin, the probability that he will find it is thus not influenced by previously acquired knowledge, even though the recognition of the discovery as a solution is certainly influenced by previously acquired knowledge. The finding, i.e., the being aware of something as a candidate for solution, is a necessary condition for judging it as a solution. But the finding itself is a coincidence where the producing factors and the selecting

---

[22] See, for basically the same critique, also Richards (1977), Skagestad (1978), Amundson (1989), Sternberg (1998); Boden (1996, 2004), and also Popper (1974b: 1061), who points towards guided variation in this sense, although he defends Campbell with respect to unjustified variation.

factors are *decoupled*, in the way required for undirectedness in biological evolution. The invention of penicillin is not made more likely by the selecting factors. The invention is thus as undirected as biological evolution. True cases of serendipity are indeed parallel to what happens in organic evolution. Knowledge and intentions only have an influence on the selection of the occurrence of novelty, not on the occurrence itself.

In the case of trial-and-error-processes, however, such as in the case of Semmelweis, intentions also have an influence on the *occurrence* of novelty. First of all, the orientation towards a certain problem restricts the problem-space in which novelty should arise. Moreover, it might also influence the probability that the appropriate changes occur, since it influences which trials *occur in the first place*. Take Kekulé's discovery of the benzene molecule, which I discussed in section 4.1, as an example for creative trial-and-error that involves insight. Kekulé's general knowledge about chemistry and his knowledge that benzene has certain chemical features played a decisive role in the production of the idea that benzene builds a ring. The knowledge that is activated through the need for a solution structures the search-space and triggers certain ideas and not others. The knowledge prevented him from coming up with a totally bizarre hypothesis. It bootstrapped some ideas and not others and increased thus the likelihood that a solution occurs in his mind that he would select as appropriate. After producing the idea, he evaluated the idea. He evaluated it in the face of the *same* cluster of knowledge that influenced the production of the idea – namely, knowledge about chemistry and chemical features. The important thing to realize is that in such a case the *producing factors were not decoupled from the selecting factors*. Part of the producing factors were – at the same time – selective factors. Kekulé did not produce his idea in a truly undirected manner, since already acquired knowledge biased the occurrence of trials.

This is *how* the orientation towards a certain problem can lead to directed 'guided variation': *Certain trials are more likely to occur because the factors that are responsible for the generation of the novelty are coupled with those factors that determine the selection*. The influence of already acquired knowledge thus leads to a disanalogy between creative problem solving and biological evolution. Creative problem solving is an instance of a coupled selection process, while biological evolution is not.[23]

Nonetheless, the influence of already acquired knowledge is only "plastic," as Larry Briskman (1981: 147), defender of the Darwinian account of creativity, proposes. If the influence were too rigid, the occurrence of new ideas would be impossible. If previously acquired knowledge totally determined what we think, in being the only factor influencing the occurrence of new ideas, then no change of thinking would be possible. That is why the independence implied in psychological spontaneity (i.e., the independence from knowledge, methods, and the like) can only be a partial one. Generally speaking, creativity can only occur if the coupling in creativity is not 100 per cent. If it were 100 per cent, creativity would be impossible. Due to factors that are not part of previously acquired knowledge, creativity can occur. In the case of creativity, variation and selection are coupled, but only to a certain degree. The likelihood of a specific variant is increased since the individual, with his previously acquired knowledge, his skills and characteristics, makes up the actual 'selective environment.' This selective environment also includes some of the factors that produce novelty. Therefore, there is no decoupling between the factors that produce novelty and those that select novelty. Variation is thus biased towards adaptivity; it is *directed.* Organic evolution exhibits no coupling. Although creative trial-and-error problem solving involves no complete coupling, the partial coupling shows an important disanalogy with respect to the origination of novelty.

Let me recapitulate: The reason for the adaptive bias in cases where we look for a solution to a certain task, cases of trial-and-error, is a form of coupling. Thus although trial-and-error as well as serendipitous discovery may well be instances of Darwinian *selection with unjustified variation*, there is an important difference between the two: Only serendipity is Darwinian in a more *fundamental* sense. Only in serendipity can we have truly *undirected* novelty,

---

[23] This was also – often ignored – the point why Toulmin introduced the concept of coupling (Toulmin 1972: 337f).

since coupling is very unlikely to occur. Cases of serendipity are important in the history of science, technology, and certainly also in art – it would be whig-history to ignore them. However, they are not really the paradigm cases of creativity, as I mentioned in section 4.1. Hence, if we want to explain creativity, we also have to address standard non-serendipitous cases, such as Kekulé's case.

## Conclusion

Boyd & Richerson can only show that a disanalogy exists at the level of cultural diffusion, since novelty is already pre-selected when it enters the cultural pool. On the basis of a critique such as Ruse's or Thagard's, it can be concluded that in cases of trial-and-error, which often involve insight, novelty is *constrained by our previously acquired knowledge*. These cases are nonetheless creative, as the example of Kekulé illustrates. I have tried to show that the adaptive bias making variation in creativity directed is due to *coupling between the producing and selecting factors*. This contradicts one of the versions of the analytic argument quoted at the beginning of section 4.2. Campbell said that blindness means that, in the case of creativity, we proceed to novelty "without benefit of wisdom" (Campbell 1974b: 142). In creative problem solving, we go *beyond* previously acquired knowledge, but we are not doing this *without* benefit of wisdom. The fact that we indeed use our previously acquired wisdom in problem solving shows an important disanalogy between creativity and biological evolution, since the latter involves *decoupling between variation and selection*, while the former does not. Nonetheless, I consider serendipitous discoveries as cases that involve decoupling, as biological evolution. But since I do not regard them as the paradigm cases of creativity, guided variation provides a challenge for the origination analogy, if this analogy is understood as requiring really *undirected* origin of novelty in creative problem solving processes.

## 4.4 SELECTIONIST AND BIAS COMPATIBILITY

### Reaction to guided variation

Darwinians replied to the critique of guided variation that the apparent guidedness in creative problem solving does not show that the analogy between origination of novelty in creativity and in biological evolution is wrong. Indeed, it is not only important to ask whether there is an adaptive bias or not. A well-balanced evaluation of the analogy has to address whether such a bias really destroys the analogy or not.

Campbell actually acknowledged the influence of previously acquired knowledge. Yet he differs on the consequences of this influence for his claim that variation occurs in a blind manner. According to him, the above-described influence of previously acquired knowledge has an analogue in biological evolution: The increased likelihood of useful trials is reinterpreted as a standard effect of cumulativity. If mutations arise in an undirected manner, then there is no ‘Lamarckian’ or ‘adaptive correlation,’ as Campbell prefers to say, between selecting and producing factors. In such a case, there is no coupling that makes useful variants more likely to occur because of their usefulness. After conceding this, he writes:

> “But even were (and where) some degree of *adaptive correlation* to be found between a new environmental setting and the mutations which are concomitant with it, or, more likely between a new puzzle situation for an animal and the responses it emits, this *neither* violates the model nor provides an explanation of an eventual improvement of fit. For this adaptive bias in variations is itself an evidence of fit needing explaining. And the *only available explanation (other than preordained harmony) is through some past variation and selective retention process*” (Campbell 1974b: 151; Emph. added).

The argument that a Lamarckian correlation does not ‘violate the model’ can be interpreted in two ways, which I will call *selectionist compatibility argument* and *bias compatibility argument*.

### Selectionist compatibility argument

That guided variation does not violate the model can be interpreted as saying that it still holds that all problem solving processes involve selection. They all rely on unjustified variation, since the novelty that comes about in human creative minds is neither warranted to be useful by its origin nor warranted by foresight. There is no ‘preordained harmony’ neither by an adaptivity-guaranteeing molding force of the environment nor by foresight. Be it as it may, this would merely be restating the weaker claim that creativity involves selection of trials, of which we do not know in advance whether they are worthwhile. Answering in this way would show that the critique of guided variation is compatible with a weak analogy between creativity and biological evolution. This is what I would like to call the *selectionist compatibility argument*. Further above, I explained in which sense I regard the weak analogy between creativity and biological evolution, i.e., the claim that origin of novelty by creativity is a selectional process, as trivial: The selectional nature of creativity is correct but no news; it mirrors the narrow definition of creativity; and it does not lead to a restoring of a more close analogy between creativity and biological evolution, namely that both rest on *undirected* variation due to decoupling of variation and selection. It can thus not be used to restore a strong analogy; it cannot be used to counter the critique of guided variation.

### Bias compatibility argument

However, Campbell can also be interpreted as providing another compatibility argument. He can be interpreted as mainly saying that the kind of bias that follows from the influence of previously acquired knowledge is the same kind of bias we have in organic evolution, namely a bias that is the effect of previous selection.[24] This argument, which I call *bias compatibility argument*, can be separated into two claims: (i) the claim that there is a nested hierarchy of vicarious selection processes, and (ii) the claim that the effect of previously acquired knowledge is similar to the effect of cumulativity in organic evolution.

*(i). Campbell’s hierarchy of vicarious selection.* Campbell believed that there is a continuous “nested hierarchy of selective-retention processes”

---

[24] Without distinguishing clearly between the two different answers, defenders of Campbell have repeated this defense strategy. See, for instance, Cziko (1995: 289f) or Nickles (2003: 60-64).

(Campbell 1987 [1974a]: 54-62). Life began with biological evolution, a selection process leading to organisms that exhibit innate characteristics and some flexibility. Because of the flexibility, these organisms do not have to wait anymore until biological evolution comes up with further favorable mutations for new behavioral variants when they encounter an environmental problem. Flexible organisms can adapt during their life to their environment. They do this through diverse ontogenetic learning mechanisms. According to Campbell, these mechanisms are *vicarious* for biological evolution. They are themselves instances of blind-variation-selective-retention-processes and only replace biological evolution. In addition, these mechanisms can be ordered into a hierarchy of ontogenetic vicarious selection processes. This ontogenetic cascade begins with locomotion as a process of behavioral problem solving. The organisms try a way, hit, for instance, an obstacle, and try another one. In some organisms, locomotion has then been replaced by sensory systems, like echo location or vision. Organisms with such a sensory system do not have to hit a wall physically in order to 'learn' that this is not the right way to go. They can sense it before hitting it. These sensory systems also operate by an internal blind-variation-selective-retention process. The difference to locomotion is that the trials are internalized sensory trials. Sensory exploration in turn is replaced by imagination and 'thought trials,' as Campbell says. In sum, thought trials replace sensory exploration, while the latter has replaced overt locomotion. Although all these kinds of trials can lead to a direct selective interaction with the real environment, the selective interaction with the environment can become internalized as well: The thought trials then interact with an internal representation of the environment; they are thus not only generated internally but also selected internally. This is what Campbell called "mnemonically supported thought":

> "At this level the environment being searched is vicariously represented in memory or by 'knowledge', rather than visually, the blindly emitted vicarious thought trials being selected by a vicarious criterion substituting for an external state of affairs. The net result is the 'intelligent', 'creative', and 'foresightful' product of thought, our admiration of which makes us extremely reluctant to subsume it under the blind-variation-and-selective-retention model" (Campbell 1987 [1974a]: 62).

Daniel Dennett states a similar nested hierarchy in his “Tower of Generate-and-Test” (Dennett 1995: 373-381). His hierarchy mainly consists of four types of creatures: “Darwinian creatures,” which are “ ‘hard-wired’ phenotypes,” leading to “selection of one favored phenotype;” “Skinnerian creatures,” who blindly try different overt responses, leading to overt selection of these overt responses; “Popperian creatures,” which have “an inner selective environment that previews candidate acts” and selects them internally, i.e., before the favorite trial is tested in a real interaction with the outward environment; “Gregorian creatures,” who additionally import “mind-tools from the (cultural) environment,” like language or other symbolic systems, tools like photography or other devices that extend our knowledge-gaining abilities.

The important point about such a nested hierarchy of selective processes is, according to Campbell, that the nesting leads to the guided, i.e., adaptively biased variation: The products of previous selection stages restrict, i.e., guide further selection. The nested hierarchy thus forms the foundation for the second thesis (ii): Guided variation is compatible with the claim that novelty arises in close analogy to biological evolution, since guided variation can be reconstructed as a normal effect of cumulativity.

*(ii). Guided variation as effect of cumulativity.* The kind of cumulativity that is exhibited in biological evolution indeed leads to a biased pattern of variation. It does so because of the effect of *developmental constraints*, which I introduced in section 2.3. As I stated there, this effect has to be distinguished from ‘coupled’ evolution. This means that Campbell can restore the close analogy between creativity and biological evolution, only if the kind of coupling that has been stressed by critics can be reconstructed as analogous to developmental constraints. That this is possible was defended in detail by Stein & Lipton (1989), who described previously acquired knowledge as a kind of pre-adaptation. For Stein & Lipton, “in both cases [biological evolution and knowledge acquisition, MK], we must face the anomaly of apparently guided variation” (Stein & Lipton 1989: 54). Biological evolution is thus considered to be not “truly blind” either (*ibid.*: 43). Stein & Lipton are followed by others, for instance by Simonton (1999a: 62), who admits that there are only “degrees

of blindness" – in biological evolution as well as in creativity. The degree of blindness is dependent on how much the variation is restricted by constraints that are the effect of previous trials.[25]

The phenomenon of pre-adaptation in biological evolution, which I will explain in a minute, finds its cultural analogue in what they call heuristics. Heuristics are defined as "anything that restricts variation" (Stein & Lipton 1989: 40). They cite a hypothetical example of a chemist trying to understand the behavior of a chemical compound. The chemist will not make random conjectures, but use heuristics drawn from the explanations of similar compounds. These heuristics restrict the construction of Campbellian 'thought trials.' Now, their important move is to state that heuristics are in fact the 'pre-adaptations' of creativity, and that the pre-adaptations in biology and creativity explain the bias in the respective production of novelty. A pre-adaptation is best explained by an example. In the case of the evolution of a complex organ there must be cumulative selection, since it is almost impossible that a single mutation can lead to its evolution, as shown in section 2.3 with respect to the example of the Shakespeare-typing monkey. Stable intermediate structures that have an adaptive advantage on their own are thus essential for cumulative evolution of complex devices. Stein & Lipton refer to the half-wing structure of the ancestors of birds as such a stable intermediate pre-adaptation for the evolution of wings. As they suggest, the "half-wing may have been used for trapping insects," and was selected, even though the half-wing did not yet allow to fly (Stein & Lipton 1989: 37). Accordingly, heuristics are pre-adaptations since "[l]ike the half-wing of biological preadaptation, the epistemic preadaptation had to be good for things other than its current adaptive use (or, at least, not harmful)" (*ibid.*: 39).

---

[25] See also Simonton (1988: 4f, 1995: 473f, 1999b: 311, 2003: 316f); see also Nickles (2003), especially p. 65-66 on the effect of cumulative evolution on less blindness in creativity, or Buskes (1998: 115-123) for the same strategy, following Stein & Lipton. See also Herbert Simon: Although drawing an analogy between problem solving as trial-and-error and natural selection, he states that the trial is not 'blind' but 'selective,' and attributes the capacity to produce only promising trials to "cues signaling progress." These "play the same role in the problem-solving process the stable intermediate forms play in the biological evolutionary process" (Simon 1981: 205f).

But all this is only half way to the analogy between pre-adaptations and heuristics, since heuristics and pre-adaptations are brought in to show that both explain the statistical bias towards adaptive novelty in biological evolution and creativity. Both must be shown to *constrain* subsequent generations of novelty. According to Stein & Lipton, heuristics restrict the future generation of novelty in the same "way a pre-organ structure restricts the sort of final organ an organism will have" (*ibid.*: 47). They also point out that heuristics, as epistemic preadaptations, and biological pre-adaptations can be viewed in

> "two, quite compatible, ways: either as restrictions on future variations or as programs for the generation of new variants. […] In biological evolution, the appearance of the half-wing makes possible the appearance of some structures, most notably a wing. At the same time, the appearance of the half-wing prevents certain other structures from appearing – for example, an arm in place of the half-wing. In epistemic evolution, if a person adopts a certain heuristic, then various future conjectures are made possible, but others are made impossible or overwhelmingly unlikely. One possible disanalogy between biological and epistemic variation that seems to have appeared can be turned into another analogous feature between the two" (*ibid.*: 47f).

This is their main argument to resolve the "anomaly of apparently guided variation" (*ibid.*: 54). The difference between directed, i.e., guided variation in creativity and undirected variation in biology is an *illusion*. As said above, the essential claim is that guided variation can be explained as a perfectly Darwinian effect of previous variation-selection cycles.

In order to understand the core of this argument it is important to see that in biological evolution pre-adaptations only have a constraining effect because they represent *developmental constraints.* That an already existing half-wing prevents the appearance of an arm in place of a wing is due to developmental constraints. Developmental constraints do not restrict mutation; they only restrict which mutations will be supported developmentally and which mutations have a positive effect on the well-being, survival and reproduction of the organism. The existence of a pre-wing does not bias mutation. But if a pre-wing is already prevalent, a mutation for development of an arm might not lead to anything, since the developmental machinery for expressing this mutation might not be present. Furthermore, developmental constraints have this constraining effect, even if they are not connected to pre-adaptations that had a different positive effect before they became selected for something else. In short, Stein & Lipton's core argument rests on the following: Guided variation can be explained as the effect of developmental constraints.

### Differences between developmental constraints and coupling

I will now explain why I think they are wrong. Their argument rests on a simple error: They mix up a bias in variation that is caused by coupling with a bias that is caused by developmental constraints. Developmental constraints differ from coupling in two respects. First, as just and in section 2.3 explained, developmental constraints do not influence which mutations and recombinations arise *in the first place*; they only restrict which mutations or recombinations will have a positive effect on the organism, given the other traits of the organism. The existence of a pre-wing does not bias mutation. Coupling, on the contrary, would influence the occurrence of novelty from the onset. It would make adaptive mutations more likely to occur because they would be adaptive in the respective selective environment. Second, and much more important for our purpose here, developmental constraints are an effect of *past* selective environments. Coupling, however, would be a direct effect of the respective *current* selective environment. Only coupling leads to *directed* variation, as defined in section 2.3, even if developmental constraints lead to *restricted* variation.

To ignore these differences is confusing a bias with its cause. A bias can have many reasons. Developmental constraints is one such cause; coupling, leading to directed variation, is a different one. Furthermore, as I illustrated in section 2.3, although some Darwinians seem to oppose developmental constraints, they are in fact compatible with the Darwinian paradigm. As long as the bias in variation, the restriction of the range of variation and the consequent change in the probability of variation, is not caused by coupling, most Darwinians do not have any problems with such a bias. In contrast to developmental constraints, coupling is considered as incompatible with the Neo-Darwinian concept of Darwinian evolution and believed to be absent in biological evolution.

If we now go back to the question whether guided variation is a problem for the analogy between creativity and biological evolution, then the following holds. The bias compatibility argument says that – despite guided variation – creativity is closely analogous to biological evolution, since in both cases there is a bias in variation that can be explained as a normal effect of previous cycles of variation and selection. My claim is that, on the one hand, this is true and, on the other hand, it is not.

Is creativity an instance of coupling, as critics of the origination analogy claim? Or is it an instance of restrictions on variations that exert their influences due to the cumulative nature of evolution, such as developmental constraints, as defenders claim? Now, the problem is that creativity is *both at the same time*, since, in the case of creativity, the effects of *past* selections make up the *current* selective environment. Cognitive variation is directed since variants that are useful are more likely to occur because they are thought to be useful. This adaptive bias is caused by coupling, since it is the same kind of knowledge that influences the production of trial solutions as well as the selection of those trial solutions. At the same time, creativity is based on 'pre-adaptations,' since the knowledge that governs the selection and production of trials is the effect of previous selection stages, restricting future developments. There is thus a cultural analogue to developmental constraints: previously acquired knowledge, partly represented in the form of cognitive heuristics and certain standards of a domain (e.g., epistemic qualities like testability in science, aesthetic qualities in arts).[26]

Important for the question at issue here is that *the knowledge (selective environment) that causes coupling between variation and selection is that very knowledge that is the effect of previous cognitive selection.* This is not necessarily the case for biological evolution. The tree in the selective environment of a giraffe is not an effect of previous selections *of giraffes*, even if some phenotypic features of the giraffe itself can be regarded as a kind of inner selective environment for the undirected mutations. In contrast to the

---

[26] See Wimsatt (1999) for very detailed suggestions on how to understand such cultural developmental constraints in terms of 'generative entrenchment.'

tree, our previously acquired knowledge, which influences the selection and the production of thought trials, *is at the same time* the effect of previous selections of the body of belief (analogous to the giraffe) whose evolution is at issue. Because of this simultaneity of effects of previous selection and current selective environment in the case of creativity, developmental constraints and coupling are harder to distinguish in the case of creativity, although they are easy to distinguish for biological evolution.

Since creativity contains developmental constraints *and* coupling at the same time, critics and defenders are both correct: Creativity is an instance of coupling, and the coupling is an effect of cumulativity that is analogous to developmental constraints. The bias compatibility argument, however, still faces a problem. Stein & Lipton's bias compatibility argument is that coupling is no argument against the analogy, since the bias in creativity can be explained by cumulativity. Therefore, they state, "we do not need the coupling of variation and selection" (Stein & Lipton 1989: 53) to explain guided variation. On the one hand, this is correct, since in the case of creativity, coupling is explained by previously acquired knowledge. On the other hand, creativity evidently exhibits coupling, while biological evolution does not. Guided variation is an effect of cumulativity *and* of coupling at the same time.

## Conclusion

My concluding claim with respect to the bias compatibility argument is the following: We might not need coupling to explain the presence of knowledge that does bias cognitive variation. But we need coupling in order to explain how the knowledge that is the effect of previous selection *causes* the bias in variation. The bias is caused by coupling, a phenomenon that is still believed *not* to be present in biological evolution. For this reason, pointing to a cultural analogue of developmental constraints cannot restore a close analogy between creativity and biological evolution.

Stein & Lipton cannot see this, since they do not see the difference between developmental constraints and coupling. The *bias compatibility argument* fails because of this central error. It mistakenly explains a factual instance of coupling as a mere analogue to the effects of cumulativity in biological evolution. In creativity, previous selection leads to coupling that causes variation to be directed. In biological evolution, previous selection does not lead to coupling. In other words, while biological evolution exhibits the effects of cumulativity and developmental constraints, it does not exhibit the effects of coupling. Biological evolution is an instance of *decoupled cumulative evolution*, often affected by developmental constraints. Creativity is a special case of *coupled cumulative evolution*. Hence, the point of the critics of the Darwinian account of creativity – that creativity is coupled while biological evolution is not – can be defended. Since the decoupled structure of Darwinian change is an essential character of Darwinism, I regard this as an important disanalogy.

Although the bias compatibility argument does fail, the selectionist compatibility argument does not. But it does not restore a close analogy either, since, as I claimed, it refers to a much weaker claim, only implying that creative problem solving is a selection process, i.e., involving unjustified but not truly undirected variation. The Darwinian analogy holds only for blindness as 'having no guarantee of success.' This said, I now turn to a last compatibility argument that has been offered to restore the strong analogy between *undirected* variation in creativity and in biological evolution.

## 4.5 HIDDEN CHAOS COMPATIBILITY

### Creativity as unconscious blind variation

The *hidden chaos compatibility argument* refers to a 'hidden chaos' beneath the apparent guidedness of creative problem solving. The argument is implicit already in Campbell (1960, 1987 [1974a], 1974b) and has most prominently been defended by the creativity psychologist Dean K. Simonton in his "chance-configuration theory" (Simonton 1988, 1995).[27] The argument refers to a pre- or unconscious mechanism at the cognitive level that is itself truly blind, although being part of an overall process of guided variation at the conscious

---

[27] I took the term 'hidden chaos' from Stein & Lipton (1989: 39f), who also defend such a compatibility. I will not analyze their formulation of the argument and instead concentrate on the stronger and more detailed version of Simonton.

level. The move is similar to the one I mentioned as reaction to Boyd & Richerson's argument about guided variation at the populational level. The point is not that blind variation is opposed to guided variation due to expertise and wisdom. The point is that expertise works with a cognitive process that is blind at the cognitive pre- or unconscious level. With this move, blindness is restored by making it *hidden* or unconscious. Hence, implicit in such a defense against the critique of guided variation is a new connotation of blind variation that we have not considered so far, namely blind variation as *pre- or unconscious variation*.

Furthermore, Simonton appears to defend the Darwinian approach to creativity not only as an approach that excludes explanations that ignore the originality and spontaneity of creativity. He appears to defend an approach that offers an explanation of the cognitive processes involved in creativity, since he points to an alleged Darwinian mechanism at the cognitive level.

### Poincaré's explanation of creativity

Campbell (1987 [1960]: 99f) and Simonton[28] both refer to the mathematician Poincaré as a major forbearer of such a Darwinian explanation of creativity. I have already introduced Poincaré's four-stage model in section 4.1. But Poincaré did not only offer a description of the phenomenological datum of psychological spontaneity; he also offered an explanation for this phenomenological structure. He describes what happens in the mind between the incubation and inspiration stage in the following way. First, Poincaré reports his own introspective experience of how it is to be creative:

> „One evening, contrary to my custom, I drank black coffee and could not sleep. Ideas rose in crowds; I felt them collide until pairs interlocked, so to speak, making a stable combination" (Poincaré 1982 [1908]: 387)

Based on this and similar introspective reports, he speculated about the cognitive mechanism in order to interpret what he experienced. According to Poincaré, creativity relies on intuition: having an idea with a feeling of certainty that the idea is appropriate, without knowing that it is. This is similar

---

[28] See Simonton (1988: 27-33, 1995: 468-486, 1999a: 32-34).

to Campbell's claim that creativity relies on guesswork. More importantly, intuition is for Poincaré the outcome of an unconscious mechanism in the "subliminal self," as he says (Poincaré 1982 [1908]: 392). In the moment of insight, this subliminal self 'presents,' so to speak, to the conscious mind certain promising ideas, i.e., good guesses. However, for Poincaré it is not the case that the subliminal self has foresight itself or "delicate intuition" that would enable this unconscious part of the mind to produce only the good trials (Poincaré 1982 [1908]: 391). According to him, the subliminal self "*blindly*" produces various ideas. Only some of these ideas become conscious:

> "Among the great numbers of combinations blindly formed by the subliminal self, almost all are without interest and without utility; but just for that reason they are also without effect upon the esthetic sensibility. Consciousness will never know them; only certain ones are harmonious, and, consequently, at once useful and beautiful" (*ibid.*: 392).

Thus, which of the blindly formed ideas become conscious is not a matter of chance. Only those that "affect most profoundly our emotional sensibility" have this "privilege" (*ibid.*: 391). And "it is this special esthetic sensibility which plays the role of the delicate sieve" (*ibid.*: 392). The esthetic sensibility reacts to harmony of ideas, an esthetic quality that first guides the unconscious selection process, and then also the subsequent conscious selection process, which the individual imposes upon those ideas that arrive on the conscious screen. According to Poincaré, the esthetic sensibility is an important competence of a good mathematician. This leaves open whether the property of harmony is an objective or a subjective property of ideas.

Be it as it may, what is important for this study is that Poincaré believed that the mind *unconsciously produces a kind of waste*: Ideas that never have an impact on consciousness. This can be interpreted as showing a kind of 'chaos' that is hidden, but nonetheless part of the apparent guided generation of novelty. In one passage he speculates about whether the subliminal self truly produces ideas at random. He writes:

> "The conscious self is narrowly limited, and as for the subliminal self we know not its limitations, and this is why we are not too reluctant in supposing that it has been able in a short time to make more different combinations than the whole life of a conscious being could encompass. Yet these limitations exist. Is it likely that it is able to form all the possible combinations, whose number would frighten the imagination? Nevertheless that would seem

> necessary, because if it produces only a small part of these combinations, and if it makes them at random, there would be small chance that the *good*, the one we should choose, would be found among them" (Poincaré 1982 [1908]: 393f; Emph. in the orig.)

Given that the subliminal self produces trials at random, we face a serious problem. If the mind worked in the manner of the proverbial monkey, it would need much more time than it seems to need in order to randomly hit at the good ideas. To resolve the question how the subliminal self can nonetheless produce new and appropriate ideas, Poincaré opts for the same kind of guidance that is the basis for the critique of guided variation: Previous knowledge and engagement with a specific problem bias the overall process and make thus useful combinations more likely. The way this happens according to Poincaré is the following:

> "Permit me a rough comparison. Figure the future elements of our combinations as something like the hooked atoms of Epicurus. During the complete repose of the mind, these atoms are motionless, they are, so to speak, hooked to the wall; so this complete rest may be indefinitely prolonged without the atoms meeting, and consequently without any combination between them. On the other hand, during a period of apparent rest and unconscious work, certain of them are detached from the wall and put in motion. They flash in every direction through the space (I was about to say the room) where they are enclosed, as would, for example, a swarm of gnats or, if you prefer a more learned comparison, like the molecules of gas in the kinematic theory of gases. Then their mutual impacts may produce new combinations. What is the role of the preliminary conscious work? It is evidently to mobilize certain of these atoms, to unhook them from the wall and put them in swing. […] after this shaking up imposed upon them by our will, these atoms do not return to their primitive rest. They freely continue their dance. Now, our will did not choose them at random; it pursued a perfectly determined aim. The mobilized atoms are therefore not any atoms whatsoever; they are those from which we might reasonably expect the desired solution" (Poincaré 1982 [1908]: 394)

Previously acquired knowledge helps the mind to 'unhook' certain promising idea-atoms. Nonetheless, the unhooked ideas are presented as colliding by chance: they 'flash in every direction through the space,' 'like the molecules of gas in the kinematic theory of gases,' they 'freely continue their dance.'

### Simonton's chance configuration

Simonton (1988, 1995), who heavily relies on Poincaré, termed his Poincaréan-Campbellian theory "chance-configuration theory." According to Simonton,

> "creativity begins with the chance permutation of mental elements. The latter include ideas, concepts, recollections, emotions, sensations, or any other

> basic component of mental functioning. Most of these permutations are too unstable to enjoy anything more than an extremely ephemeral existence in the fancy. Nonetheless, from time to time, a specific combination of elements coalesces to form a cohesive whole, or conceptual Gestalt. This so-called chance configuration represents the insight that transfers to more deliberate and elaborate processing at later stages in the creative process" (Simonton 1995: 467).

This model has been explained to some detail in Simonton (1988: 1-23). The creative mind pre- or unconsciously forms permutations. These "chance permutations vary appreciably in stability" (Simonton 1988: 8). Highly stable permutations are termed 'configurations.' The stability is not an output of a conscious selection process but the input for further conscious information processing. Furthermore, for Simonton, stability seems to be a consequence of objective properties of ideas (*ibid.*: 13). I will not discuss the latter issue. Instead, I want to concentrate on the implied blindness of creativity and on whether the process postulated by Simonton is in fact the way humans produce novelty or not.

According to Simonton, chance does not mean equiprobability. He acknowledges the influence of prior knowledge on creative problem solving in about the same way as Poincaré did. He adds that even within the 'unhooked' mental atoms, combinations are not random in the sense of equiprobability. Whether the kind of randomness that is then implied is undirected variation is, however, not made clear. The only thing that he definitely requires is that myriad determinants influence the subconscious cognitive process and that a "large number of potential permutations exist, all with comparably low but nonzero probability" (*ibid.*: 7). I take this to imply in any case that there is blind variation in the sense that there exist unconscious trials that are a kind of 'waste,' since they never come to consciousness. And let me take for granted, for the sake of argument, that the kind of randomness is analogous to undirected variation, since Simonton evidently wants to defend Campbell's claim that creativity is analogous to a *Darwinian* kind of origination of novelty, despite the critique of guided variation.

At the basis of this, I take Simonton's chance-configuration theory to include at least two claims: (1) Creativity contains a hidden chaos, i.e., undirected, unconscious, false trials; (2) these trials are produced by a *special* cognitive process, the unconscious chance-configuration process, which accounts for creativity. The second claim has so far not appeared in this study and is less important for it; but the claim is not independent of the first claim, as a short examination of the evidence for the two claims will show.

### Evidence for a hidden chaos as an explanation of creativity

Is chance configuration the hidden mechanism by which creativity operates in the human mind? Poincaré himself admitted that his account of creativity is only a speculation. Campbell (1987 [1960]: 108f) conceded that his model of creativity as blind variation is not yet an explanatory psychological theory, since such a theory would require to specify and cite evidence for the exact cognitive mechanism of creativity. He treated Poincaré's speculations as a possible hypothesis about such a cognitive mechanism that would be compatible with his theory. This was in the 1960s. Psychology of creativity has come a long way since the 60s, and it is Simonton who claims that in the light of current developments of psychology of creativity, Poincaré's model turns out to be the correct cognitive explanation of creativity. The dispute between Simonton and other contemporary creativity theorists is, however, whether conclusive evidence exists for such a cognitive explanation.

*(i). Hidden chaos*. As mentioned in section 4.1, most creativity psychologists accept Poincaré's stage model. Within this model, the special process of unconscious chance-configuration would be the explanation for the incubation and subsequent inspiration stage. Incubation is the label for the phenomenon that insight often occurs during or after rests – during or after the time when people turn away from their engagement with a specific problem that demands a creative solution. Apart from (a) Simonton's "*subconscious random-recombination hypotheses,*" there are three other common hypotheses about what actually happens during incubation, which are debated in standard psychology of creativity: (b). The "*conscious-work hypothesis*" states that the cognitive processes are not really pre- or unconscious processes; the thought processes merely occur so rapidly so that they are hard to report by the creative person. (c). The "*fatigue-dissipation hypothesis"* states that during incubation people merely recover and are not engaged with the problem. (d). The

"*selective-forgetting hypothesis*" states that during a break people forget their false and inhibiting strategies and can thus more easily find a new way. Although it is a major problem to construct reliable and realistic experiments to test the different hypotheses, the evidence that was gathered by experimental psychology is not in favor of Poincaré's subconscious-random-recombination hypothesis (Seifert et al 1995).

*(ii) Special process view:* In addition, most theorists in creativity research have given up searching for a special process underlying incubation and insight. They point towards ordinary cognitive processes. According to the *ordinary process view*, creativity can be *demystified* as ordinary cognition that merely operates at its highest efficiency.[29] The ordinary processes that are assumed to make up creativity are for instance: *perception* and *visual imagery* in general; processes such as "*Janusian thinking*" (i.e., conceiving two or more opposite antithetical ideas or images simultaneously) or "*Homospatial thinking*" (i.e., superimposing different ideas or images);[30] diverse *cognitive heuristics*, like distorting, repeating, omitting and mixing parts of concepts; in particular, *associational linkage*, *conceptual combination*, *analogical reasoning*, *abstraction*, *use of metaphors*, *conceptual expansion* (i.e., the extension of the boundaries of concepts); in addition, *memory retrieval* in general and *spreading activation* (i.e. one remembered idea activates related memories and does this across many related memories); *defocused attention* as a cause for spreading activation.

In a nutshell, the ordinary process view claims that to explain creativity one does not require a special process of chance-configuration, or any particular special mental operation, which only creative people can perform. The diverse mechanisms that are very likely involved in creativity, which interact in complex ways, often cannot be reported or reconstructed by the creative agents themselves, if they are asked how they came up with their ideas, since these mechanism operate quite fast and some thoughts are rather

---

[29] See, in addition to Seifert et al (1995), Weisberg (1993: especially 42-50, 56-58 and the summary of his alternative model in ch. 8), or Ward et al (1999); see also Boden (2004: 260ff) and Mumford (1999).

[30] See Rothenberg (1986) for Janusian and Homospatial thinking.

fleeting. Nonetheless, they are not necessarily pre- or unconscious, even if they are not consciously monitored and reported. They also do not necessarily operate in a blind or random fashion. Jonathan Schooler & Sonya Dougal (1999: 352) for instance note that spreading activation is “by no means random in nature. Rather, the direction and extent of the spread of activation critically depends on (a) the specific items that were initially activated and (b) the underlying structure of an individual’s knowledge representation.”[31]

Furthermore, these processes can incorporate various clues from perception and self-generated (i.e., stimulus-free) perceptual imagination. These extraneous clues and imaginations, which can indeed be r-random with respect to the goals and knowledge of the individual and thus comparable to those events involved in truly serendipitous discovery, are then incorporated and associated with the other material that is used to tackle the problem or project at hand. This ordinary-cognition model of creativity thus builds on the traditional view that chance favors only the prepared mind: *Serendipitous clues* play a role even in non-serendipitous creative trial-and-error problem solving, but they can only play this role because of a complex and sophisticated network of cognitive mechanisms, which have little resemblance to the internal, random dance of Poincaré’s idea-atoms. Seifert et al (1995) coined a name for the incorporation of serendipitous clues. They call it “opportunistic assimilation”: opportunistic, i.e., serendipitous information is processed by normal cognitive mechanisms, leading to an assimilation of this information.[32]

Every creativity theorist accepts that creativity involves trials that are not warranted by their origin alone. Although they thus accept that creativity involves psychological spontaneity, they nonetheless deny that there is a special cognitive mechanism that is inherently Darwinian and that accounts for creativity. The consensus that emerged in recent years from creativity research is thus not only a change from a special-mechanism account to an account in terms of a single ordinary mechanism. The shift was one from a single

---

[31] See Mumford (1999: 345) for the same argument with respect to associational linking and conceptual combination.

[32] They also apply this approach to the example of Kekulé’s dream, see Seifert et al (1995: 115f).

mechanism account to an account that refers to many cognitive mechanisms that work together in complex and diverse ways, making up the myriads of idiosyncratic cases of creativity that are so hard to bring under a general model, precisely because of the sheer complexity of interaction of these mechanisms. The accounts of different creativity theorists certainly differ, but they all cite evidence that is not in favor of a special process of chance-configuration.

To give a full survey and explanation of these diverse approaches and the evidences they cite for their ordinary process view would much exceed the space available here. I shall rest content therefore with the following concluding claim with respect to the hidden chaos compatibility argument: As long as there is no definite empirical evidence for an unconscious production of a ‘hidden chaos,’ the claim that creativity relies on a process that is undirected, despite the apparent guidedness, cannot be defended. I will now present what Simonton himself has offered as definite evidence for his Darwinian approach to creativity and illustrate why this evidence also provides no basis for the hidden chaos compatibility argument.

### Simonton’s defense

It is essential to realize that in later papers Simonton implicitly withdraws from the claim that the cognitive mechanism that accounts for creativity is chance-configuration. Over the years, he seems to have moved back to a weaker position that does not defend a special cognitive mechanism that is treated as truly undirected and thus closely analogous to Darwinian evolution. Sometimes Simonton states that creativity is explained by a chance-configuration-mechanism (Simonton 1988, 1995: 467f). But sometimes he merely states that creativity is dominated by a general process of blind-variation-selective-retention. The chance-configuration-mechanism is then but one of many mechanisms accounting for this general process (Simonton 1999a, 1999b, 2003). This weaker version of his theory is in fact the weak claim I discussed in section 4.2: the claim that creativity is guesswork, i.e., lacking foresight. Thus Simonton says, for instance, that the stress of ‘blindness’ of generation of novelty in creativity denotes

> "the lack of foresight in the production of variations – the inability to generate purposively the most adaptive variations. […] The term blindness also has the advantage of not committing the theory to any particular variation mechanism" (Simonton 1999b: 310).

With this move, Simonton seems to settle for the ordinary process view of creativity. Nonetheless, he cites Poincaré in both cases. This is one of the reasons why it is sometimes hard to find out which claim he wants to defend.

The problem to which I want to point to is that although moving back to a weaker claim, he upholds that the Darwinian model *explains* creativity, i.e. that it is stronger than the claim that creativity involves guesswork. In section 4.2, I already pointed out that the thesis that creativity is guesswork does not have much explanatory force, since it merely re-describes the phenomenon of psychological creativity, as it is usually defined. Although this concept excludes certain explanations of creativity, it does not give a positive explanation itself.

I will now back up this critique of the Darwinian account of creativity by looking at the evidence Simonton himself cites for the Darwinian account. As it is often unclear which claim he wants to defend, it is unclear whether the evidence should be considered as evidence for the strong claim about an unconscious chance-configuration process, or whether it should be considered as evidence for the weak claim that creativity is guesswork.

In the *Precise* (Simonton 1999b: 312f) of his book *Origins of Genius* (Simonton 1999a), an article that has been commentated extensively by other creativity psychologists, Simonton (1999b) refers to the evidence for his Darwinian approach to creativity. The evidence he cites comes from three methodological domains: (i) evidence from the historiometric, (ii) evidence from the psychometric, and (iii) evidence from the experimental domain. By reviewing the debate about this evidence, I will show that even the evidence cited by Simonton himself does not provide *direct* evidence for (1) and (2), the claim that there is a special process of chance configuration that accounts for creativity.

*(a). Evidence from the historiometric domain.* I will start with the most indirect evidence, the historiometric evidence, which tends, as Simonton says, "to fall in line with what we would expect from a Darwinian model" (Simonton

1999b: 315). Over the years, Simonton gathered a lot of data about professional career development. This research provides evidence for what Simonton has called the “equal-odds rule” (Simonton 1997). The equal-odds rule says, for instance, that the expected *probability* of career success remains constant within one career, regardless of quantity of output, and regardless of the creator’s age and increase in accumulated knowledge. In other words, the *proportion* of hits, e.g. in terms of frequently cited publications, in relation to *total* attempts stays the same across time. It stays the same regardless of the age and increase in accumulated knowledge. A scholar might produce more works later in his life, but in relation to his total output he does not produce more good works; he also produces more waste. This is indeed an interesting finding.

However, the important question is in which sense this provides evidence for the claim that beneath guided variation lies a cognitive mechanism of blind, unconscious variation in the Darwinian sense. Simonton writes that careers of “creative output […] have several features that are most compatible with a Darwinian view of creativity. Probably the most remarkable feature is the consistent relation between quantity and quality” (Simonton 1999b: 316). As illustrated above, the equal-odds rule says that the relation between quantity and quality stays the same over time. This means that, within careers, the principle of the equal-odds shows that within a career more expert knowledge, acquired over the years of engagement with a domain, does not make it statistically more likely that a person produces a work that can count as ‘success,’ according to a certain standard. According to Simonton, this shows that the guided variation cited by critics is an illusion, since the individual has no chance to increase the likelihood of ‘hits’ by using already acquired knowledge. He writes: “The fascinating aspect of this principle is that it is what we would predict from the Darwinian viewpoint. If the variation process is truly blind, then good and bad ideas should appear more or less randomly across careers” (Simonton 199b: 316).

First, this interpretation of the data contradicts his repeated insistence that creativity normally is not ‘truly blind,’ but exhibits only ‘degrees of blindness,’ whereas the degree depends on the amount of knowledge accumulated over time.[33] Second, the data can also be interpreted as evidence for the weaker Darwinian analogy, namely that creativity involves guesswork, i.e., that the creative author of scientific or artistic output has no guarantee that he produces a ‘hit,’ according to this or that standard. Third, the historiometric method cannot provide direct evidence for a cognitive mechanism, since it measures only the socially ‘successful’ creative ideas. It measures how often we succeed according to a quantitative standard, for instance, according to the amount of citations. But the historiometric method does not provide direct evidence that allows to draw conclusions about *how* we come up with the ‘succeeding’ *and* the ‘non-succeeding’ ideas. In particular, it does not give evidence for an unconscious chance-configuration mechanism. From this perspective, the historiometric evidence does not allow any inference to any specific cognitive process. In direct reaction to Simonton, Michael D. Mumford refers to the same problem, in order to show the inadequacy of the historiometric evidence:

> “[C]reative thought is not a simple, uniform process. Instead, multiple processes, strategies, and mental operations may be involved, applied by different people, in different ways, at different points in a creative effort. When such complex causation exists, and we aggregate data over a variety of problems and settings, we can expect the resulting data to fit a random model. This point is of some importance because it suggests that caution should be exercised whenever aggregate historic data are being used to draw inferences about cognitive operations. More centrally, however, this point implies that inferences about the nature of creative thought, such as universal blind generation, must be made with reference to specific cognitive processes being examined under controlled conditions” (Mumford 1999: 344).

Because of these limitations, historiometric evidence cannot provide supporting evidence for the strong thesis of a chance-configuration mechanism.[34]

*(b). Evidence from the psychometric domain*. With respect to psychometric studies, Simonton cites the *Remote Associates Test* from Sarnoff Mednick (1962). It tests whether individuals are able to make rather remote

---

[33] See section 4.4.

[34] For further critique of inferences from the equal-odds rule to cognitive capacities and mechanism see Schooler & Dougal (1999: 354-355).

associations. He also cites Joy P. Guilford (1967)'s *Alternate Use Test* of divergent thinking, which tests how many different ideas of using a common object a person can come up with. According to Simonton, both tests "operate according to an implicitly variation-selection model of the creative process" (Simonton 1999b: 314). Indeed, these creativity tests measure the ability to produce novel, unusual ideas. They are compatible with the weak Darwinian thesis of creativity as guesswork, and are compatible even with the stronger postulate of an unconscious chance configuration process. Yet they are compatible with other cognitive processes as well. The tests thus do not provide evidence for a specific cognitive mechanism. They measure creativity according to a certain definition of creativity, which entails that novel ideas are produced. These tests 'operate according to' a certain *concept* of creativity and not according to a distinctive *Darwinian account of creativity*. They are compatible with almost any creativity theory resting on the narrow psychological concept of creativity. They are designed according to this concept.

The same holds for the findings about characteristics of creative personalities. And Simonton admits this, in writing that these findings are

> "quite compatible with what we would expect to be necessary from a Darwinian view of creativity […] That is creative personalities tend to possess those characteristics that would most favour the production of ideas both numerous and diverse" (Simonton 1999b: 315).

Although he cites this compatibility, he wants to defend a distinctive Darwinian explanation of creativity. In a commentary on a similar statement, the creativity psychologist Sternberg objects: "Really, is there any theory of creativity that would take issue with this statement? Almost any plausible theory could account for these and similar claims" (Sternberg 1999b: 358). Simonton just reformulates what creativity is: the production of interesting, novel ideas. As Sternberg states, the evidence does not "directly support the Darwinian theory of creativity. […] At best one can say that there are findings that are not wholly inconsistent with the evolutionary theory, but even this claim would be pushing things. And many other theories are at least as consistent with the findings presented" (Sternberg 1999b: 358). All the cognitive processes mentioned above, referred to by defenders of the ordinary process view, are consistent with the psychometric evidence cited by Simonton. Thus psychometric evidence does also not provide support for the claim that there is a hidden chaos, a chance-configuration mechanism, behind the apparent guidedness in creative problem solving.

*(c). Evidence from the experimental domain*. The last kind of evidence Simonton cites is experimental evidence from laboratory studies and computer simulations. They most directly bear on questions about cognitive processes. For instance, he cites that creativity seems to rely heavily on ‘opportunistic assimilation.’ It is important to realize that the latter was a process to which critics of a Poincaré-style chance-configuration process pointed to as well (Seifert et al 1995). And Seifert et al (1995) explicitly formulated their opportunistic-assimilation-hypothesis as an alternative to a chance-configuration process. The process of opportunistic assimilation is indeed a mechanism that allows the mind to incorporate and assimilate coincidental experiences and imaginations. But this mechanism is totally different from a chance-configuration-mechanism and “just because creativity can be fostered by random cues does not necessarily implicate randomness in the psychological process of creativity,” as Schooler & Dougal (1999: 352-353) pointed out, with reference to the approach of Seifert et al (1995) and in reply to Simonton (1999b).

Simonton also presents the so-called *Geneplore model* of creative cognition, formulated by Finke, Ward and Smith (1992), as supporting the Darwinian thesis, although he concedes that this model is “not explicitly formulated in Darwinian terms” (Simonton 1999b: 312). The problem is the same as with the other evidence cited by Simonton. The Geneplore model is consistent with the weak version of Simonton’s Darwinian model, i.e., consistent with the claim that creativity involves unjustified variation and selection. But the Geneplore model does *not* rely on the claim that the generation of novel ideas, so to speak, is ‘blind’ or dependent on a special process of unconscious chance-configuration. It only relies on the claim that creativity generates ideas that still have to be explored, elaborated, and tested in order to find out whether they are worthwhile. That is why the model is called *Geneplore* – generate and explore. Finke, Smith and Ward explicitly state that describing creativity as involving variation and exploration, which incorporates selection, is a mere "general framework," a "heuristic model" that does not yet explain creativity (Ward et al 1999: 191). In addition, they explicitly rely on the ordinary process view. Random generation through external coincidental clues or other mechanisms is judged by them to play some role, but, as described above, the influence of coincidental clues itself relies on diverse mechanisms that are rather structured and not 'blind' in any significant sense (Ward et al 1999: 209).

The same holds for computer models of creativity as evidence for a Darwinian analysis. Simonton cites genetic algorithms as "blatantly Darwinian" models of human creativity (Simonton 1999b: 313).[35] Genetic algorithms are analogous to Dawkins' computer monkey that randomly types the sentence from Shakespeare. Indeed, genetic algorithms can be regarded as perfectly Darwinian systems, relying not only on undirected, but on truly random generation of trials. But according to Boden, for instance, a defender of a connectionist model of computer and creativity, our mind does not work like genetic algorithms. Our flexibility in thinking is based on generative structures that put constraints on the generation of ideas even at the pre- or unconscious level (Boden 1999: 366-369). Our mind is not a structure-less Poincaré-like box in which nothing prevents the ideas' "random dancing from falling onto madness" (Boden 2004: 34). According to Boden, the presence of previously acquired knowledge does not only cause certain ideas to be part of the combinatorial mental 'dance.' It structures the Poincaréan 'space' of dancing idea-atoms, so that certain ideas are more likely to combine with others. If Boden is right, then genetic algorithms do not model our creativity. Now, Simonton admits that Boden's connectionist (and also classical AI-models) are not as Darwinian as genetic algorithms. Nonetheless, he considers them as

---

[35] See also Simonton (1999a: 55-60, 2003: 317).

Darwinian enough, since they all incorporate at least a ‘random’ number generator (Simonton 1999b: 313).

But all this taken together leads to a strange situation: If computer models *differ* with respect to the kind of cognitive mechanism for which they should be a model, as genetic algorithms differ from connectionist networks, how can they then count as supporting the *same* thesis about cognitive mechanisms modeled by these computer programs? The diverse computer models cannot provide evidence for a chance-configuration mechanism, only genetic algorithms as a viable model of our creativity could, if at all. Only genetic algorithms could serve as a computer model of truly undirected variation at the cognitive level, and could thus be cited in justification of the hidden chaos claim. Simonton’s claim that diverse currently existing computer models of creativity are evidence for his model shows that in the end ‘blindness’ means less than what is present in genetic algorithms. It only means that when we generate new ideas, we incorporate input that is r-random into a process that is strongly guided by expertise. The output we thereby generate is not guaranteed to be a solution for the problem at hand, but the output is not produced in an undirected way either. Simonton ends up with the weak version of the origination analogy applied to creativity: Creativity involves variation and subsequent selection, but not undirected variation.

### Compatibility with diverse cognitive mechanisms

This weak version is not only trivial for the reasons offered in section 4.2. It is trivial, since it is compatible with almost all kinds of cognitive mechanisms. What Simonton’s defense amounts to is that the empirical findings are *compatible* with the suggestion of a chance-configuration-mechanism. The problem is that although some of the evidence might indeed be compatible, it is compatible with other theoretical models as well, precisely because the evidence is not a direct evidence for a truly undirected chance-configuration mechanism. The Darwinian analogy therefore does not by itself explain creativity, since it has no evidence for the cognitive mechanism it suggests as explanation for creativity.

As quoted above, Simonton himself has talked about mere 'compatibility' of the evidence he refers to with his Darwinian model. A further example should make clear where this leads to. In an earlier paper, his chance-configuration theory is presented as most inclusive and therefore valuable. It is compatible with Freudian primary process thinking that is guided by the inner logic of the Freudian 'Id.' At the same time, computational models of creativity are presented as not being compatible with such Freudian theorizing (Simonton 1995: 488). However, the Darwinian theory is compatible with *both* of the opposing views. As I said, the problem is that such a Darwinian theory of creativity is compatible with almost all kinds of cognitive processes. But if we really want to understand and explain creativity, we have to look at the cognitive processes and whether they justify the Darwinian analogy despite the apparent guidedness. Psychology of creativity, as I illustrated, has started to do exactly this. Therefore, at the level of basic cognitive processes, such as those cited above as part of the ordinary process view, the Darwinian approach to creativity cannot provide an alternative explanation.

## Conclusion

If Simonton is understood as offering an explanatory claim about a special cognitive mechanism, then the Darwinian approach to creativity still has to bring in the evidence for this claim. With respect to the critique of guided variation it can thus be concluded that the hidden chaos argument cannot provide a way out of the critique of guided variation, as long as there is no direct evidence for a chance-configuration process. Thus, as long as there is no such evidence, the close analogy between undirected variation in creativity and biological evolution *cannot be justified.*

If Simonton is taken as not claiming the existence of a chance-configuration mechanism, then the Darwinian approach to creativity is *trivial* in explanatory terms: If the Darwinian account of creativity only refers to a general process of creating candidate ideas that may turn out to be false, then the claims that are derived from such a weak origination analogy are compatible with most findings in creativity theory, since the analogy is not offered at the level where explanations are suggested today, a level where, for sure, divergences in psychology of creativity arise. As long as the analogy does not specify a cognitive mechanism that is in itself Darwinian, like the chance-configuration-mechanism, the origination analogy does not provide an explanation of creativity, since the level at which psychology of creativity looks for an explanation is the level of basic cognitive processes. In addition such a weak analogy is also *descriptively trivial*, since it merely assumes the same concepts and facts other approaches to creativity assume as well, namely the contemporary folk-psychological and scientific concept of creativity, requiring psychological originality and spontaneity. As long as there is no direct evidence for a hidden chaos, the Darwinian account of creativity can only provide a general framework that is almost universally taken for granted by the psychology of creativity.

## 4.6 SUMMARY

Culture is a variational system, often changing in a gradual way, and creativity provides a major source of novelty in culture. However, the novelty in culture is not introduced in an undirected way. It is already pre-selected by the creative individual when it enters the cultural pool. Creativity itself is based on blind variation in the sense that there is no guarantee that the ideas that are produced are appropriate. Creativity is in this sense Darwinian. It is, however, not Darwinian in a strong sense, since it is based on a process that involves a guided origination of novelty that is analogous to a partial Lamarckian coupling of producing and selecting factors. This is an important disanalogy. In the case of creativity, variation is not *undirected in the Darwinian sense*.

The bias compatibility argument fails since it ignores the differences between coupling and developmental constraints. In addition, the disanalogy cannot be rescued by pointing to a ‘hidden’ Darwinian chaos, beneath the guided generation of novelty. The hidden chaos compatibility argument fails, as long as there is no direct evidence for such chaos. That the evidence does not seem to be in favor of such a hidden chaos has been defended by showing alternative approaches to creativity and by reviewing the evidence Simonton himself has cited for a hidden chance-configuration mechanism. The only way to restore the analogy is the selectionist compatibility argument, i.e., to insist that creativity involves guesswork, variation and subsequent selection. But this was never in doubt and is trivial, since it does not provide by itself an explanation of creativity, even though it excludes explanations that ignore originality and spontaneity.

To say that creativity is Darwinian beyond the almost trivial sense that it involves guesswork is extending the analogical game too far. Campbell saw this when he stated:

> "At the level here developed, one might better speak of an 'orientation to', or a 'perspective on' creative thought processes, rather than a 'theory of'" (Campbell 1987 [1960]: 108).

According to him, this constitutes a weakness of the model" (Campbell 1987 [1960]: 110). To quote a special mechanism, as Simonton did, for which there is no direct evidence has not helped to get rid of this weakness and I doubt that it ever will. Yet, to mix up the different levels of the Darwinian analogy from nature to culture is even worse. There are three levels of analysis where creativity has to be taken into account if cultural change is at issue: creativity as the source of variational change in culture at the populational level; creativity as a cognitive selective process itself; creativity as being based on undirected variation. These levels have so far not been distinguished in the detail developed here, neither by critics nor by defenders. Yet to distinguish these levels is required for a well-balanced evaluation of the origination analogy, and shows that the central error of the Darwinian approach to creativity lies in overextending the analogy: Although we can say that creativity leads to a variational pattern, that creativity is a selection process, it is not the case that creativity is based on undirected variation.

# 5 EXPLANATORY UNITS OF SELECTION ANALOGY: SELECTION OF MEMES

## 5.1 MEMES AS THE SELFISH UNITS OF CULTURAL SELECTION

### Diffusion from a traditional point of view

Evolutionary theory not only aspires to describe a pattern of change as a variational change, i.e., as a sorting process leading to a frequency change of cultural units. Evolutionary theory wants to *explain* the frequency change. Explaining the change involves two kinds of questions: How does an item originate and why does it spread and persist in a population. Whether the origination of cultural novelty is analogous to Darwinian evolution has been addressed in chapter 4. This chapter will be concerned with *why* certain cultural items (memes) *spread and persist* in a population and not others. At issue is the process of *diffusion* of cultural units. Memeticists point towards memes not only because they want to draw an ontological analogy as described in chapter 3. In analogy to gene selectionism, memeticists have claimed that the units that account for cultural selection are memes, the 'selfish replicators,' the ultimate beneficiaries and causal 'agents' of culture. Those memes which are 'selfish,' i.e., which have a high fitness of their own, will spread and others will not. This is what I call the *explanatory units of selection analogy*.

The traditional conception of culture entails that cultural items spread and persist because individuals select them. Humans determine which memes spread and persist – by adopting them or not. Selecting memes by adopting them is understood as intentional selection, 'sighted' and not 'blind', even if human selection does not rely on *foresight* of whether the selected items prove worthwhile. We do not have this kind of perfect knowledge. That is one of the reasons why our *rationality* in making selective decisions is restricted. Despite limited knowledge and the absence of foresight, we nonetheless guide the diffusion of cultural items in choosing cultural items according to certain beliefs, preferences, expected utility, and intentions. This is what philosophers call the *belief-value model of decision processes*, which is assumed by almost all social sciences and builds the foundation of folk-psychological explanations of human action. This model presupposes what Dennett has called the “intentional stance,” i.e., attributing beliefs and desires to humans (Dennett 1987).

In an empirical way, diffusion has been studied, for instance, in anthropology, by dual-inheritance-theorists, and in economics. I will take the tradition of diffusion studies in economics as an example, in order to show what this traditional conception of intentional selection processes amounts to. Over the last decades, Rogers (1995), with his paradigmatic research on diffusion, collected a lot of evidence for understanding diffusion: Diffusion of innovation could be shown to be dependent on characteristics of the respective newly introduced cultural item, but dependent only as these characteristics are perceived by individuals. Each individual interprets cultural items differently. The interpretations depend on many factors: the information available to the person, previously acquired beliefs and preferences, social status of the individual, or communication structure (i.e., who communicates with whom, use of mass media, etc.). Although the basic psychological model, assumed for the individual decision processes involved in adopting new cultural items, is the standard belief-value model, this traditional concept of culture does not assume that the people are perfectly *rational* in an *objective* sense.

Although I cannot go into the details of rational choice models, let me summarize some standard assumptions about rational choice. Rational choice can be described as referring to a *means-end-rationality*.[1] Individuals are assumed to rank options for actions according to subjective preference orderings. In economics it is further assumed that one can quantify these preference orderings as utility orderings by assigning numbers to the options. Higher numbers stand for more preferred options. According to this formalized model, developed by John von Neumann, Oskar Morgenstern, and Leonard Savage, rational choice is assumed to maximize expected utility. That a

---

[1] For detailed accounts see Hampton (1998), or Elster (1989).

decision-maker only maximizes *expected* utility means that it is not assumed that the decision maker has perfect knowledge. Many decisions are taken under risk, i.e., the decisions rely on more or less probable outcomes of choices. In addition, we often have to rely on *subjective* probability judgments about the respective probability distribution. This has been called decision under uncertainty.[2] Our beliefs, which are the basis for our calculation of the likelihood of a specific outcome, and the calculated expected utility of these outcomes might thus well be wrong. That we make errors because of this has already been an issue in the last chapter 4. In making such errors, we are thus not perfectly rational in an *objective* sense. But we nonetheless can be called rational in a *subjective* sense – rational at the basis of our limited knowledge. We made the best out of what we had available. Contrary to this subjectivist concept of rationality, philosophers sometimes assume that the beliefs and preferences that form the basis of our decisions have themselves to be rational, e.g., that the preferences confirm ethical standards, or that the beliefs are true or at least justified.[3] But most research in social science does not assume such normative standards for rational choice. However, there is another normative requirement that is indeed imposed on subjective rational choices. Even if our decisions are based on limited knowledge, the traditional concept of rationality demands that we are at least consistent in what we believe and desire, that we really maximize our utility, and that we satisfy certain axioms of expected utility theory.[4] If we conformed to these procedural standards, we would be perfectly rational, but in a subjective sense. Psychology, however, has gathered a lot of evidence that we are not even rational in this sense. Simon (1959), for instance, has claimed that we are not maximizing utility, we are merely 'satisficing,' since there are not only constraints on gathering information (e.g., limited access, high costs of obtaining information, and limited time), but also constraints on reasoning ability (e.g., limited memory, intellectual disabilities,

---

[2] See, for instance, Katz & Rosen (1998) as a purely economical treatment of rational choice as maximizing expected utility.

[3] See, for instance, Kutschera (1999) or Searle (2001). For a general overview see Hampton (1998) on practical rationality, and Adler (1998) on the rationality of belief.

[4] See Hampton (1998) for a brief description of the axioms of expected utility theory.

and too difficult, costly, or time-consuming calculations). According to Simon, our rationality is *bounded* in diverse ways. Furthermore, theorists have claimed that expected utility theory is empirically wrong, since we violate some of the axioms that are assumed in the expected utility theory for rational calculations.[5] As a reaction, the normative standards of consistency, of maximizing utility, and of expected utility theory are sometimes interpreted as a mere idealization of what we humans actually do and can do.[6] It is not important here whether we conclude in light of the just mentioned findings that humans are often just irrational or whether we reformulate the concept of rationality, since we still want to understand ourselves as rational.[7] What is important here is that the contemporary concept of human decision-making takes into account that we do not have perfect knowledge, that we make errors in our decisions and that our cognitive abilities are rather restricted, i.e., that we are often not able to make perfect and complicated calculations when we have to make a decision.

Let me go back to diffusion studies: The empirical research on diffusion has also observed again and again that *objectively* useful innovations fail to spread. For instance, boiling water failed to spread in villages in Peru. Rogers (1995) claims that the reasons for this (objectively) irrational resistance comprise the following ones: First, the people perceived the innovation as incompatible with local knowledge. Second, the so-called ‘change agent’ (i.e., the person introducing a cultural item into a population) was perceived as too different in social status and life style. Third, a prestige bias prevented adoption, i.e., the opinion leader in the respective population was opposed to the innovation. It is not important here what the actual reason for this behavior was. The important point is that these cases do not contradict the contemporary concept of a subjective rationality, since the case only shows that, at the basis of the preferences and beliefs of the villagers of Peru, boiling water was just

---

[5] See, for instance, Kahneman, Slovic & Tversky (1982).

[6] See Spohn (2002).

[7] Some philosophers claim that we should not abandon the concept of rationality, since rationality is an attribution to humans that is constitutive for attributing intentionality and language. See Dennett (1987), Stich (1990: 29-50), or Searle (2001).

not perceived as a good choice. The traditional belief-value model does not have to rely on strong standards of perfect objective rationality.

### The explanatory units of selection analogy

Meme theory wants to present an alternative to the just described traditional model: For meme theory, cultural items spread and persist, because the cultural items themselves have a context-independent survival value, *a fitness of their own* that is maximized, a fitness that plays an important explanatory role in answering why certain ideas spread. Memes are the selfish units of cultural selection, as genes are the selfish units of natural selection. In a section on the "the philosophical importance of memes," Dennett claims that the gene-meme-analogy is important, since memes are the units that explain *why* a certain unit spreads in a human population, even if memes are not replicators in the narrow sense, i.e., even if we cannot *identify* memes as easily as genes, and even if we do not know exactly *how* memes spread.

I ask you to recall from chapter 2 that Dawkins not only claimed that genes are replicators, he also suggested that they are the ultimate units of selection, since they are *active* replicators, difference-makers with a context-independent effect and a fitness of their own, building 'vehicles' that interact on their behalf with the environment. Genes are thus the ultimate 'agents' that get the 'credit' for adaptations – in terms of survival – and that are selected for these adaptations. For gene selectionism, fitness of genes is the only fitness that is maximized in all cases of evolution: The fitness of genes and their benefit in terms of survival does not only account for 'outlaw' genes (see section 2.5), but also for cases in which organisms could equally be regarded as units of selection, namely as reproducing interactors that benefit from their adaptations. Gene selectionism claims to account for these cases as well, since the causal power of these interactors can be reduced to the causal power of genes. For this reason, gene selectionism considers organisms in general as mere *consequences* of the 'selfish,' replicative, and organism-building power of the replicators.

In analogy to gene selectionism, memeticists have claimed that *memes are the selfish units of selection that explain cultural change*. Humans, carrying around these memes, are a mere consequence of their ‘selfish,’ replicative, and mind-building power. Gene selectionism is mirrored by *meme selectionism*. According to Dennett, the “crucial point” (Dennett 1995: 364) is the following claim:

> “[A] cultural trait may have evolved in the way it has simply because it is *advantageous to itself”* (Dawkins 1989 [1976]: 200; Emph. in the orig.).

Dennett considers this claim as providing a “striking new perspective” (Dennett 1995: 353). It “challenges one of the central axioms of the humanities,” namely the traditional view that we do what we do and think what we think because we believe that it is good for *us* (*ibid.*: 362). According to memetics, we do what is good for memes, what is good for their ‘fitness.’ Below I will explain the claims that are included in the explanatory units of selection analogy in a more systematic way.

Before that I want to say a little bit more how it relates to the preceding chapters. This explanatory units of selection analogy is independent of the origination analogy, and partly dependent on the ontological analogy. It can be true or false irrespective of whether cultural selection is based on undirected variation. But, it should be clear from chapter 3, that memes cannot have a fitness of their own, if they do not form a lineage. If the lineage requirement for replicators is not fulfilled, no memes (be it as type or lineage of token) survives. What survives is a blend or an average of many similar memes, but not a single meme. This argument would suffice to render the units of selection analogy pointless, at least for all cases of social learning that include averaging. However, the point that I will develop in this chapter, a point against the explanatory units of selection analogy, is stronger: Even if some memes form lineages, it can be doubted that these memes are active difference-makers that have any explanatory priority over their human carriers, as genes are claimed to have explanatory priority over their human carriers.

Although the explanatory units of selection analogy is dependent on the ontological analogy, it is not dependent on the truth or falsity of gene selectionism. Even if gene selectionism is correct, meme selectionism could still be false. As I said in section 2.5, I do not want to judge whether gene selectionism is correct. What I will do instead is the following: I have explained in section 2.5 that the controversial core of the units of selection debate lies in issues about causality: whether single genes-as-replicators are mere bookkeepers or have a special causal power – a fitness of their own that represents a causal influence that singles them out as replicators *for* a certain phenotypic consequence, i.e., a causal influence that allows to reduce the causal role of the organism in evolutionary processes as a mere effect of the causal power of single genes. This is the Achilles heel of gene selectionism. Genes can be the ultimate units of selection of biological evolution only if it can be shown that genes are not mere bookkeepers. The analogous claim about memes as the selfish units of cultural selection will be shown to have the same Achilles heel: Memeticists have to show that attributing to memes a fitness of their own is more than bookkeeping. They have to show that memes have a causal priority over other entities, in the case of memes, over human individuals.

### ‘Selfish memes’ from a systematic point of view

The explanatory units of selection analogy comprises two central claims: (1) The survival of the fittest meme explains cultural diffusion; (2) memes and not humans, with their beliefs and their preferences, determine and therefore explain diffusion. The latter claim entails that meme selectionism provides an alternative to the traditional explanation of diffusion introduced above. Apart from (1), this second claim is based on three arguments: (i) there is no connection between meme fitness and the utility of memes for us; (ii) irrationality can only be explained by ‘selfish memes;’ (iii) minds are built by memes. Let me explain these claims and arguments.

*(1). The survival of the fittest meme explains diffusion.* For memeticists, the survival of the fittest meme explains the pattern of diffusion we find in a culture. For Dennett, one of the main defenders of memes as an explanatory concept, a

> “[m]eme X spreads among the people because X was a good replicator” (Dennett 1991: 205).

I have already quoted Richard Dawkins, who became famous for introducing not only the ‘selfish gene,’ but also the ‘selfish meme,’ for writing:

> „What we have not previously considered is that a cultural trait may have evolved in the way it has simply because it is *advantageous to itself*" (Dawkins 1989 [1976]: 200; Emph. in the orig.).

Under the paragraph "Whose advantage," of her book *The Meme machine,* Susan Blackmore states:

> "The whole point of memetics is to treat the meme as a replicator in its own right, operating entirely for the benefit of its own selfish replication" (Blackmore 1999: 30).

At the end of her book she writes:

> "This is the power and beauty of memetics: it allows us to see how human lives, language, and creativity all come about through the same kind of replicator power as did design in the biological world. The replicators are different, but the process is the same. We once thought that biological design needed a creator, but we now know that natural selection can do all the designing on its own. Similarly, we once thought that human design required a conscious designer inside us, but we now know that memetic selection can do it on its own. […] If we take memetics seriously there is no room for anyone or anything to jump into the evolutionary process and stop it, direct it, or do anything to it. There is just the evolutionary process of genes and memes playing itself endlessly out – and no one watching." (Blackmore 1999: 242).

According to Dennett, Blackmore, and Dawkins, who are the three main theorists[8] in memetics, and according to other so-called memeticists, cultural evolution is about the struggle for existence of replicating memes. They do not only state that there are replicators or that humans are 'blind watchmakers,' just as natural selection; they state that there is 'no one watching.' All we need in order to explain culture is the fitness of the selfish memes. Memes are not only bookkeeping the change; they are the primary causal factor that we have to take into account in order to explain cultural change. As Sterelny has summarized memetics, "the crucial element of a meme-based theory is that the fitness of the memes themselves plays a crucial explanatory role" (Sterelny, forthcoming a).

In summary, meme selectionism tries to explain the retention and spread of ideas with 'the good of memes,' the differential fitness of memes

---

[8] I say theorists, since only Dennett and Blackmore can really count as defenders. We saw already at the end of chapter 3 that Dawkins restricted the force of the analogy after the first publication of *The Selfish Gene* because of the differences in ontology and transmission of memes and genes. Nonetheless, Dawkins was important in stressing the explanatory units of selection analogy by claiming that there are 'viruses of the mind' that invade us. These viruses survive despite their disadvantage for us (Dawkins 1993).

instead of the causal power and interests of individuals. This is accomplished by *invoking a property of ideas: their fitness as an explanatory concept*. Therefore the units of selection analogy is: As the survival of genes, and only their survival, can explain all cases of biological evolution, only memes and their survival can explain all cases of cultural evolution. The ontological analogy is extended into an *explanatory analogy*. As genes – being the units of selection, the ultimate 'beneficiaries' of biological evolution – explain biological evolution, memes – being the units of selection and beneficiaries of cultural change – explain cultural evolution.

*(2). Memetics as an alternative to the traditional explanation.* Part of this position is an opposition that is similar to the opposition between genes and organisms in the units of selection debate. Properties of memes and *not* the properties of individuals explain why a unit of culture spreads. Individuals are mere *vehicles*, hosts or resources for memes, driven by memes. They are secondary for the goal of explanation. As an organism appears from the gene's eye perspective as just a gene's way of a making another gene, from the 'meme's eye perspective,' "[a] scholar is just a library's way of making another library" (Dennett 1991: 202). This position has been justified by the following three claims that have to be distinguished carefully:

*(i). There is no connection between the fitness of memes and utility of memes for our goals.* According to Dennett,

> "The first rule of memes, as it is for genes, is that replication is not necessarily for the good of anything; replicators flourish that are good at …. replicating! – for whatever reason. […] The important point is that there is no necessary connection between a meme's replicative power, *its 'fitness'* from its point of view, and its *contribution to our fitness (by whatever standard we judge that)*" (Dennett 1991: 203; Emph. added).

Note that Dennett says '*by whatever standard we judge that.*' It is not only that memeticists want to show that we believe things that do not contribute to our biological fitness. For Dennett, and less rigorously for Dawkins, there are memes that do not even contribute to other goals that we might have: truth, aesthetic values, moral values, other more mundane general goals such as influence or money, or particular goals such as the desire to relax. Attributing such goals to individuals is what Dennett calls the "traditional model," since it "uses the intentional stance as its explanatory framework" (Dennett 2001a: 307). As illustrated above, if we take an intentional stance towards people, we describe them as having ideas, beliefs, and values that determine their decisions about what to believe, what to desire, what to do in order to fulfill the basic desires, and so on. Meme selectionism is thought to replace this traditional model, by explaining culture through reference to the fitness of memes. As explained above, this is parallel to the claim that genes are not only bookkeepers of biological evolution but have a fitness of their own that allows to single them out as the active units of selection that alone benefit, in the long run and in all cases, from the adaptive phenotypic consequences these memes cause in their 'vehicles.'

This is what I call the *general independence claim.* It can be understood as one argument for claiming that minds cannot explain cultural change, since the fitness of memes does not dependent on what individuals regard as their benefit, whatever that may be. But memeticists offer two further arguments for their second main claim that meme selectionism replaces the traditional perspective.

*(ii). Irrationality can only be explained by 'selfish memes' in the narrow sense*. One further argument is that the meme's eye perspective is *more inclusive* than the traditional intentional stance perspective. The traditional model of explaining diffusion is presented as failing to explain all cases of diffusion of cultural items (Dennett 2001a: 309). Dennett sometimes acknowledges that the traditional explanation can explain many cultural phenomena, but not all. There are cases – mainly cases of *irrationality* – where the pattern of diffusion of memes cannot be explained by referring to beliefs and intentions of individuals. It can only be explained by the fitness of memes.

Here is an example: Why do most of us use Microsoft and not Unix or Linux to run our personal computers? A hypothetical traditional explanation could be the following: Most of us have chosen Microsoft, because it was better advertised, easier to use, easier available, etc. Whatever the precise answer is, the answer would always include that we *judged* Microsoft to be better for certain goals, given the information at hand. After a while a further factor might have played a role, namely compatibility with the systems the majority uses. If people use not Microsoft system, file sharing might turn out to be difficult. Now, many computer programmers would agree on the fact that Linux is a better system. Be it as it may, Microsoft won the battle. It is costly and objectively irrational to use the system, if there is a cheaper and better one. But even if it were objectively the case that Linux is better, we can explain the fact of the widespread use of Microsoft systems through the fact that humans are not always perfectly rational and that given the situation the person is in – given their knowledge and their skills to work with computers, given what other people use – the choice for Microsoft can be considered as subjectively rational, given the constraints on the decision situation. – Do we need 'selfish memes' to explain such a case of irrationality? Meme selectionism states that we do: Even if we can explain cases of rational behavior by the traditional model, in order to explain irrationality, we need meme talk. Microsoft system spreads because its properties enhance or 'serve' the survival of the Microsoft-system-meme.

Note that this claim has to be distinguished from the general claims (1) and (2), which entail that the fitness of memes is *always* the ultimate causal source of their spread, independent of what we regard as useful spread. The claim about irrationality is used as a justification of (2), although it is more restricted than (2), since it does not provide a new alternative for *all* cases of diffusion. It only presents the fitness of selfish memes as an alternative for cases of irrationality. Recall from section 2.5, that one has to distinguish between 'selfish genes' in the narrow sense, cases where genes spread despite neutral or negative consequences for organisms, and 'selfish genes' in the broad sense, where genes are declared to be the ultimate source and 'beneficiaries' of *all* cases of evolutionary change. As with gene selectionism, if the general philosophical claim about genes as the ultimate units of selection in all cases of biological evolution is different from the more limited claim about 'selfish' outlaw genes, we also have to distinguish between the limited claim that memes have to be taken into account, since they can explain cases of irrationality, and the general claim that memes have a fitness of their own independent of our preferences in all cases of diffusion. That is why I call the claim about irrationality the *limited independence claim*.

*(iii). Minds are memes*. In addition, memeticists have offered a third argument in order to justify the second basic claim that memes provide an alternative to the traditional model: Meme selectionism does not only claim that memes have a fitness of their own that causes their spread independent of the preferences of human carriers; it claims that these carriers *are* nothing else than a conglomerate of memes. Persons are built by memes; they merely are the effects of memes. Dawkins, Dennett, and Blackmore describe human minds mainly as being either meme *vehicles*, as organisms are for genes, or as *hosts* for memes, as organisms are for viruses.[9] Human minds are built by them. Humans are thus understood as mere 'survival machines' or 'replicating machinery,' and considered as secondary for explaining diffusion of memes. This is employing the 'carrying'-geno-phenotype relation, which I introduced briefly in section 3.2. Anthropologists used to express the same in saying that humans are 'culture bearers.' Yet, it will prove decisive that humans are not just a bundle of memes, and that carrying memes is not the only role humans play in the diffusion of culture.

Let me add a note on the philosophical importance of the explanatory units of selection analogy, before I make clear in which sense I will approach this analogy. It is the second basic claim, namely that memetics challenges 'one of the central axioms of the humanities,' that makes the explanatory units of selection analogy so radical and philosophically important. At first sight, the units of selection analogy leads to a kind of Gestalt shift, like a radical move in our perception of culture: Before, there were individuals with a prominent role as creators and choosers of culture. They carry with them all kinds of cultural items, now called memes. Now, there are memes, living, surviving, being spread by 'individuals' that are driven by these memes to do so. Susan Blackmore for instance writes about the "power behind the idea of memes":

---

[9] The virus metaphor was part of the meme idea from the start (Dawkins 1989 [1976]: 192). See Dawkins (1993), exclusively on memes as viruses.

> "To start to think memetically we have to make a giant flip in our minds just as biologists had to do when taking on the idea of the selfish gene. Instead of thinking of our ideas as our own creations, and as working for us, we have to think of them as autonomous selfish memes, working only to get themselves copied" (Blackmore 1999: 7-8).

This is the central philosophical point of the explanatory units of selection analogy. Dennett himself does not shrink back from emphasizing the seeming importance of the question whether this analogy is correct.

> "I don't know about you, but I'm not initially attracted by the idea of my brain as a sort of dung heap in which the larvae of other people's ideas renew themselves, before sending out copies of themselves in an informational Diaspora. It does seem to rob my mind of its importance as both author and critic. Who's in charge, according to this vision – we or our memes?" (Dennett 1991: 202).

### Tautologies, dilemmas, a straw man, and minds as memes

Dennett does a lot to persuade the reader that memes are in charge and not minds, and that this is an important new insight. On the other hand, from time to time, he stresses that there is no conflict with our traditional idea of individuals as creators and critics of culture (e.g. Dennett 2001b). He then stresses that the traditional explanation is not opposed by, but included by meme selectionism. I will state that he indeed has to claim this, but for a reason he himself ignores. Dennett has to claim this *in order to give memes any explanatory role at all*: As any Darwinian explanation of change that refers to the 'survival of the fittest x,' the claim that cultural diffusion can be explained by pointing to the 'survival-of-the-fittest-memes,' is in danger of being tautological. This follows from what I illustrated with respect to the tautology problem in section 2.4. I will address in which sense meme selectionism faces a tautology problem in the next section 5.2.

Discussing the tautology problem for memes directly leads to my main critique of the explanatory units of selection analogy, which I will present in section 5.3: The tautology problem leads the explanatory units of selection analogy into an *explanatory dilemma*: The tautology charge cannot be solved for memetics *without giving up the second main claim*, namely that memetics provides an alternative to the traditional explanation. If meme explanations want to get out of the tautology problem, they inevitably end up with the traditional explanation of diffusion. There is no general independence of memes from characteristics of minds, be they preferences, beliefs, or other structural features of the mind. The meme's point of view is thus no *alternative* explanation: It does either not explain anything, or it is a redundant retelling of what we can explain, and can *only* explain, if we give a version of the traditional explanation: Individuals select memes and it depends totally on *them* which memes spread, i.e., have a higher fitness. The dilemma is: 'survival-of-the-fittest-meme'-explanations are either tautological or – if this is not the case – they are redundant, since they have to refer to the traditional explanation. In both cases the analogy is trivial, either because nothing is explained, or because the explanation is heuristically trivial, explaining diffusion in the same way we have always explained diffusion. I will also show where one of the central errors of the explanatory units of selection analogy lies: Memeticists tend to misconstruct the role that individuals play in culture by wrongly transferring the units of selection analogy to culture.

In section 5.4, I will then discuss whether the *limited independence claim* can give back some explanatory force to the units of selection analogy. There are indeed cases where memes are adopted by people, despite the fact that these memes are – as a matter of fact – not useful for them, or are at least not judged useful by them. I call these cases 'conflict cases' and will state that even for these cases, it is often possible to find a traditional explanation – a reason why the individual wanted to adopt the cultural item. I will also state that even in these cases, it is not memes that explain the patterns, but the *relation* between memes and the features of the mind that constitute the selective environment of these memes. The explanatory units of selection analogy still faces the before-mentioned explanatory dilemma. Discussing the conflict cases will make explicit where the second central error of the explanatory units of selection analogy lies: Meme selectionism fights against a *straw man*, namely a too rigid concept of rationality.

The last section 5.5 concentrates on the third claim that has been used to justify the second main claim of the explanatory units of selection analogy, namely that persons are build by memes. I will show that this argument does not provide a way out of the explanatory dilemma. Not everything in an individual – building up the main selective environment for a specific meme – is itself a meme. Furthermore, even if other memes – already in the mind of a selecting human person – make up an important part of the selective environment, the fitness of that meme, whose spread is at issue, is not independent of the individual who adopts the meme or not. The thesis that minds are nothing else than memes is wrong, and for those parts of the mind that are indeed build by memes the thesis is trivial, since we knew all the time that minds incorporate ideational units, now called memes. The explanatory dilemma still holds: In order to explain anything at all, the explanatory units of selection analogy has to refer to the traditional explanation of diffusion.

## 5.2 TAUTOLOGY PROBLEM OF MEMETICS

### The tautology problem applied

What explanatory work do 'survival-of-the-fittest-meme'-statements do? As we saw in section 2.4, explanations that make reference to 'survival of the fittest x' are in danger of being tautological. Meme selectionism has been criticized for facing exactly this problem. After rejecting the tautology charge for biological evolution, basically on the grounds explained in chapter 2, Kim Sterelny and Paul Griffiths, while adopting an argument from Elliott Sober (1993), claim that "a variant of the tautology objection" applies to memetics: "We can call a tune 'a meme with high replication potential' rather than 'catchy' if we like. But without source laws, this adds nothing to our understanding of musical trends" (Sterelny & Griffiths 1999: 334). Source laws are general statements about the relationship between genes, organisms or species, and their respective selective environment. As explained in section 2.4, this relationship between the units who are supposed to have a fitness and the respective selective environment builds the causal basis for fitness differences that provides a way out of the tautology problem.

David S. Wilson makes the same point in his review of Blackmore's *Meme Machine* (1999):

> "The oft-repeated accusation that natural selection is a tautology fails because fitness is not defined in terms of whatever evolves but in terms of the properties that enable organisms to survive and reproduce in their environments. […] For the meme concept to escape the same problem, we must define cultural fitness independently of what evolves. If the first four notes of Beethoven's fifth is a powerful meme only because it is common, we have achieved no insight" (Wilson 1999: 206).

As long as 'survival of the fittest x' statements do only refer to actual survival of memes, the statement that diffusion is explained by the survival-of-the-fittest-meme is tautological, since it explains survival by survival. Sober himself writes with respect to scientific theories as memes: "It seems harmless to agree that fitter theories spread, the question is what makes a theory fitter" (Sober 2000: 218). This tautology objection has often been stated, but not analyzed in detail.[10]

Talk of 'selfish' genes or memes and their 'interest' is considered by memeticists as a useful 'shorthand' for what the units of selection analogy actually is about:

> "We can say that memes are 'selfish', that they 'do not care', that they 'want' to propagate themselves, and so on, when all we mean is *that successful memes are the ones that get copied and spread, while unsuccessful ones do not*. This is the sense in which memes 'want' to get copied, 'want' you to pass them on and 'do not care' what that means to you or your genes." (Blackmore 1999: 7; Emph. added)

If this is what it says, the explanatory units of selection analogy is indeed stating a tautology: "that successful memes are the ones that get copied and spread, while unsuccessful ones do not" is one variant of the Darwinian principle 'survival of the fittest x.' Such a statement merely says that the successful ones are the successful ones (i.e., the ones that get copied and spread). Such a statement is tautological if successful is *defined*, *measured* or *explained* by 'being copied.' Hence, the relevant question is how memeticists define, measure and explain successfulness of memes, if looked at more closely?

They specify the fitness of memes via reference to longevity, fecundity, fidelity, and more concrete qualities such as attractiveness. As outlined in section 2.5, Dawkins defined genes as active replicators. Something is an

---

[10] See also Conte (2000) and Henrich, Boyd & Richerson (forthcoming).

active replicator, if it replicates in the narrow sense, and if it does this in an *active* way. It is active, if its "nature has some influence over its probability of being copied" (Dawkins 1982a: 83). According to Dawkins, memes are such active replicators. Furthermore, memes – like genes – direct evolution towards their 'interest,' since they are the units that ultimately are responsible for the features selection acts on. The meme with the highest survival value can persist over time and spread through the population of individuals. Thus, a good replicator is one that has a high "survival value" (Dawkins 1989 [1976]: 193). Note, that the term 'survival value' takes the place of the term 'fitness'. Contrary to Dennett (see quotations above), Dawkins uses the term fitness only for individuals, and uses 'survival value' for replicators instead (Dawkins 1989 [1976]: 136-37). And as fitness can be differentiated into viability and fertility, Dawkins differentiates survival value into three criteria that mark a good replicator: fidelity, fecundity, and longevity. For memes he writes:

> "But just as not all genes that can replicate do so successfully, so some memes are more successful in the meme-pool than others. This is the analogue of natural selection. I have mentioned particular examples of qualities that make for high survival value among memes. But in general they must be […]: longevity, fecundity, and copying-fidelity" (Dawkins 1989 [1976]: 194).

In the case of memes, the concrete qualities that account for high survival value are for instance 'attractiveness,' 'being catchy,' 'memorable,' or 'tempting,' etc. Dawkins writes for instance:

> "If the phenotypic effect of a meme is a tune, the catchier it is the more likely it is to be copied. If it is a scientific idea, its chances of spreading through the world's scientific brains will be influenced by its compatibility with the already established corpus of ideas. If it is a political or religious idea, it may assist its own survival if one of its phenotypic effects is to make its bodies violently intolerant of new and unfamiliar ideas" (Dawkins 1982a: 110).

Since his most favorite example is the "god meme", i.e., a religious belief in a supernatural being like the Christian deity, let me quote a passage from *The Selfish Gene*:

> "The survival value of a god meme in the meme pool results from its great psychological appeal. It provides a superficially plausible answer to deep and troubling questions about existence. […] God exists, if only in the form of a meme with high survival value, or infective power, in the environment provided by human culture" (Dawkins 1989 [1976]: 193).

This is how Dawkins specifies the ‘survival value,’ in other words, the fitness of memes. Now, do ‘being catchy,’ having ‘great psychological appeal,’ having ‘infective power, in the environment provided by human culture’ *explain* why certain beliefs spread and not others, or are these qualities mere re-descriptions of the high survival value of memes and thus not providing a way out of the tautology problem?

**Fitness of memes beyond actual survival**

The debate about the tautological character of Darwinian explanations in section 2.4 showed that a Darwinian explanation that refers to ‘survival of the fittest x’ can only escape the tautology objection, if (a) fitness is defined as a supervening propensity for survival and reproduction, if (b) it is possible to differentiate systematically and empirically between frequency changes that are caused by drift (frequency changes that are not due to fitness differences) and those that are caused by Darwinian selection according to differential fitness, and if (c) it is possible to give independent evidence of fitness by pointing to the causal *basis* of fitness differences, in order to really explain those selection processes that are due to differential fitness.

*(a). Fitness as supervening property.* Even if memeticists usually do not care about such details, it would be no problem to interpret the survival value of memes as a propensity for survival and reproduction. Dual-inheritance-theorists, using memes not as replicators in the wide sense have defined survival value as propensity: Durham (1991: 194), for instance, defines the fitness of memes as “replicability.”[11] He explicitly refers to the tautology problem and acknowledges that this does not solve the tautology problem, since replicability does not “reveal *why* variant 2 is the most readily transmitted and used” (Durham 1991: 195).

*(b). Memetic drift and memetic selection.* A systematic differentiation between drift and selection has no prominent place in memetics, but it should also be no real problem for them. What Dawkins says in the last quotation

---

[11] See also Cavalli-Sforza & Feldman (1981: 15), who define “cultural fitness” of a cultural trait also as propensity.

about the god-meme is that people who believe in God do this, because a belief in God has a high probability of spread, given the environment of human minds. He ignores that the persistence might be due to factors similar to drift. He assumes that it is due to fitness differences.[12] Henrich, Boyd & Richerson (forthcoming) linked their claims about biases in the transmission of memes to the tautology issue. Conformist biases (i.e., biases to follow the majority) or prestige biases (i.e., biases to follow the prestigious) are factors that affect the probability of spread of a meme, but *irrespective of the content of a meme*. According to Henrich, Boyd & Richerson, the units of selection analogy, singling out memes as the causal agents whose properties explain why they spread, ignores that such non-content biases affect spread of memes in a way that is analogous to *drift*. The memes spread although no quality of the meme explains the increase of the spread. Some diffusion of memes is thus 'survival of the lucky meme' and not the survival of the fittest meme. The lucky meme spreads, because it happens to be in the mind of a prestigious person and not because of its content. It does not spread because of an intrinsic property of the meme itself that would make it more attractive for humans. An example is the spread of the Western business four-in-hand tie. According to Henrich, Boyd & Richerson, this cultural item did not spread because "the four-in-hand tie is intrinsically more attractive than its many alternatives, but because it happens to be associated with the economic and military prowess of the West" (Henrich, Boyd & Richerson, forthcoming).[13] Be it as it may, let me assume, for the sake of argument, that meme selectionism could differentiate between drift and selection.

*(c). Basis of fitness differences.* Given that memes are defined as propensity and that drift and selection can be distinguished, the important question is the following: Does meme selectionism *explain the supposed fitness differences* by providing an account of the causal basis of the different survival value, for those cases where drift can be excluded? As said, without giving an account of the causal basis of fitness differences any explanation based on the

---

[12] But there are other places, e.g. Dawkins (1999: xiv), where he considers drift as a force.

[13] Gil-White (forthcoming) makes the same point about 'lucky' memes.

principle ‘survival of the fittest x’ is deemed to be vacuous. Where do these fitness differences come from? In other words, can we identify the fittest beforehand, independently of actual survival, as Wilson asks for in his critique of Blackmore? These are the questions that need to be focused on.

In biological evolution the causal basis for fitness of organisms lies in the adaptedness of the organism, given a certain environmental setting. As I said in section 2.5: Adaptedness is a property of an organism that emerges from the relationship between an organism and its selective environment. To be ‘fit’ in the qualitative sense means to ‘match’ the selective environment. This property of the organism is independent of actual quantitative fitness, in terms of survival and reproductive output, but it is *not independent* of the relationship of the organism to the environment, since the relationship is the cause of the adaptedness of the organism, and adaptedness, in turn, is the causal basis for the fitness of the organism. As with genes, the *ultimate cause of fitness differences* lies in the *relationship* between memes and *their* selective environment. This relationship *always* determines the fitness of memes. That is the reason why Dawkins writes about “infective power, *in the environment provided by human culture.*”

In order to escape the tautology problem, meme selectionism has to point to the selective environment of memes. Now, the decisive selective environment of *cultural* items is the human mind. That is consensus within and beyond memetics.[14] The human mind has a certain structure, incorporates certain abilities and not others, houses already acquired beliefs, decision heuristics, and values. On the one hand, memeticists acknowledge that. But on the other hand, when memeticists give an account of what determines attractiveness (meme fitness at the concrete level), they give a slightly different account of the role of the human mind – and this difference is important.

---

[14] Certainly, some memes can spread in computers without a human mind ever noticing it. But these computer memes cannot be counted as ‘cultural,’ since they would in no way be part of human culture, i.e., shared between people.

## Memeticists' factors – influencing the fitness of memes

Balkin (1998: 74-88), defender of memetics as a new theory of ideology, correctly differentiates between three kinds of factors that influence the fitness of memes: First, "substantive factors" – the content of memes; second, "psychological factors – the cognitive structure of human minds and their comparative susceptibilities;" third, "ecological factors" that, according to him, include other memes, as well as the "nature of social institutions, methods of storing information, and technologies of communications."

According to Balkin, psychological factors mainly include ease of memorization, ease of comprehension and ease of communication. Ecological factors, such as social institutions, involve skills and standards that change the selective environment for memes by changing the minds of individuals. Science and universities are the best examples for such an ecological factor: These institutions are responsible for keeping alive the skills and standards that are used in these institutions to decide what to consider as a reasonable belief or action. Models and authorities, as said above, also influence the fate of memes, since we might adopt an idea not because of the content, but because of the prestige of the model. This is how drift enters the picture. Media influences the pattern as well: Writing, for instance, made possible that ideas that are less easy to memorize can spread more easily. Let me add a note on the importance of social structure: Human mind has to be brought in contact with a meme, as Laland & Odling-Smee (2000: 134) stress for instance. Contact is determined by social structure, institutions, media technology, and so on. If a person is studying at a university, the likelihood that it adopts the Darwinism-meme is increased since the likelihood of contact is increased. Thus, the fitness of memes also depends on this ecological factor. Other memes, however, do only belong to an ecological factor if they influence the distribution of memes via these ecological factors. But they mainly influence the diffusion of memes via the psychological factors, since they are normally represented in the minds of individuals. Although ecological factors are important, I would like to focus on the relationship between Balkin's substantive factors and his psychological factors, since what is at issue in this chapter is the alleged independence of memes from their human carriers.

Blackmore, for instance, concentrates on these two factors. At one point, she explicitly asks which factors influence the survival value of memes. She mentions various factors: memory capacity and "fidelity of the brain," imitative capacities, properties of the meme itself (e.g. the property of being easy to imitate), and, last but not least, human preferences (Blackmore 1999: 58). On a different occasion, she is more systematic and says:

> "There are many reasons why some memes succeed and others fail. These reasons fall roughly into two categories. First, there is the *nature of human beings as imitators and selectors*. From the memetic point of view the human being (with its clever thinking brain) acts both as the replicating machinery, and as the selective environment for the memes. Psychology can help us understand why and how this operates. There are the properties of our sensory systems that make some memes obvious and others not, the mechanisms of attention that allow some memes to grab the available processing capacity, the nature of human memory that determines which memes will be successfully remembered, and the limitations of our capacity to imitate. We can, and will, apply this to understanding the fate of memes but it is more properly the domain of psychology and physiology than memetics. The other kinds of reasons concern the *nature of the memes themselves*, the tricks they exploit, the ways they group together and the general processes of memetic evolution that favour some memes over others. These have previously been studied by psychology and are an important aspect of memetics" (Blackmore 1999: 15-16; Emph. added).

In a nutshell, according to Blackmore, there are two categories of reasons for the fitness of memes: properties of the mind (i.e., Balkin's psychological factors) and properties of the meme (i.e., Balkin's substantive factors). This is correct but confusing at the same time.

### The essential relation between memes and minds

The important point is to see that these two categories are not alternative or complementary in the sense that they are two kinds of *independent* influences. The two categories specify the above-mentioned *essential relation between the meme and its environment that accounts for the 'survival value' of memes*. Blackmore does not see this. Certainly, in analogy to organisms and their fitness, certain 'adaptive' intrinsic properties of the meme are the basis of the fitness of memes. However, these adaptive properties are adaptive only relative to a selective environment. And that means that the properties of a meme are adaptive, leading to a high fitness of memes, only because of properties of the human mind, which build the main part of the selective environment of memes. The two factors build the essential relation between memes and minds – that very relation that accounts for adaptedness, which builds the causal basis for fitness differences of memes.

Blackmore, however, presents the two factors as two *independent* factors, while acknowledging at the same time that humans are the "selective environment for the memes." Balkin seems to acknowledge that the factors are "linked in practice." He writes, for instance: "[T]he kinds of substantive content that make memes more attractive or more often discussed may depend on structural features of the human mind and existing religious or educational institutions" (Balkin 1998: 74). But he ignores, that the fitness of memes not only *may* depend on structural features of the human mind and on already acquired beliefs. The fitness of memes definitely *has* to depend on these features, since *any* fitness of any entity *depends on the properties of the selective environment of this entity*. And the selective environment of memes is mainly determined by properties of the human mind. Although Balkin regards the substantive and the psychological factors as 'linked in practice,' he also presented them in the way Blackmore did, namely as two independent factors.

Both do not take into account that the substantive factors do not constitute an own factor. They *are factors determining the fitness of memes in the way they do only because of their relation to the psychological factors*. They ignore that the content of a meme stands in the same relation to certain properties of the mind, as an opposable thumb stands in relation to properties of the environment of the organism that exhibits this opposable thumb. That memes have the property of 'attractiveness' is thus the *outcome of the relation between content (substantial factor) and properties of the mind (psychological factor)*, given a number of other conditions like social structure, institutions and media. This is the first essential point that must be taken into account in order to get an answer to how memetics can get rid of the tautology charge.

The consequence of this point is the following: Attractiveness is indeed just another word for 'fitness' as a propensity for replication of a specific meme. We can only *explain* this attractiveness, when we explain *why* the individuals find the respective meme attractive. As Sterelny says with respect to the god-meme and Dawkins argument that its spread is due to the high replication potential of the god-meme:

> "Suppose we agree with Dawkins that religious ideas are both fundamentally irrational and costly to their hosts. Even so, the argument would go, we explain nothing by labelling religious ideas as cognoviruses or memes. To explain the prevalence of religion in human life we need to know why humans in so many cultures find religious ideas salient, credible, memorable. Religion would not be part of human social life if people found religious ideas absurd or unintelligible. The crucial problem is one of human psychology: explaining why we find occult-force explanations credible. Once we find out why humans find credible explanations of their environment in terms of occult forces, what else is there to explain?" (Sterelny, forthcoming a).

Despite the fact that other factors, like the kind of media or institutions, also influence the probability of spread, this is the cornerstone of the problem the explanatory units of selection analogy has to face. Memes cannot have a fitness that is *independent* of properties of the human mind, since the human mind is the selective environment, without which we cannot explain why memes are attractive. Consequently, there is *no general independence* of meme spread from humans – with their preferences, beliefs, and other structural features of their minds.

Let me explain this point with a further example. Imagine two memes, one is easy to remember, another less easy to remember. And assume that this is the only thing that makes them different. Which one will spread more likely? The one that is easier to remember. Why? Because easy-to-remember memes have a higher survival value than less-easy-to-remember ones. Despite the fact that this is a property of the meme, this property of a high survival value arises only from the relationship between the precise content of the meme and the minds that remember these memes and thereby select them. Memes are not *easy to remember on their own*. They are easy to remember *for somebody or something*. If we were, for instance, like a computer that can easily remember long chains of numbers, memes with long chains of numbers would have a higher survival value than they have in their human environment. They don't have such a high survival value, because *we* have problems with remembering long chains of numbers. That something is easy to remember can depend on an innate bias towards certain contents, as researchers like Sperber 1996 and Atran 2001 would claim. It might well be that the reason why we cannot easily remember long chains of numbers is such an innate bias. However, the probability of memorizing a meme is likely to depend as well on general structural features of the mind (e.g., that we can learn from others at all). And most importantly, the probability of meme spread depends on what we incorporated into our web of belief before. A sentence from Kant's *Critique of Pure Reason* might be easy to memorize for a philosopher, but hard to memorize for a layman. Furthermore, already acquired beliefs not only determine what we can easily memorize or understand at all. They also determine our goals in acquiring memes. Only if I want to find the truth of something, is the property of 'seem to be true' a factor that influences my adoption of the respective meme. If we value certain aspects of life, then a religious idea might be more attractive than a competing non-religious one. If we value others, it might not be attractive at all.

Let us go back to Dawkins' example of the god-meme. The god-meme has a high survival value only because the god-meme provides an answer to "deep and troubling questions about existence" of humans (Dawkins quote above). As Sterelny writes, if we changed the respective cognitive environment, then these religious ideas would no longer spread socially, though others would (Sterelny, forthcoming a). Present the god-meme to Martians and it might have no chance at all. The fitness of memes is *always* dependent on the selective forces the individual imposes on it.

The problem is not that Dawkins, Dennett, and Blackmore do not know that the environment is *the* factor, or at least *a* factor, that determines or influences the spread of memes. The problem is that they do not realize that *this constitutes a problem for their claims* about memes and diffusion. If they acknowledge that the fitness of memes is determined by their selective environment, that the selective environment is the human mind, and that the relation between mind and meme is constitutive for the causal basis of fitness differences of memes, then they end up with the following: What explains the fitness of memes and hence diffusion is not the fitness, but the causal basis for fitness. Reference to 'fitness' of memes is just a placeholder for the actual explanation of fitness differences by pointing to the relation between the evolving unit and its selective environment. The fitness of memes is thus never *independent* of 'psychological factors.' In addition, fitness of memes is not even a factor. It merely is a supervening propensity that follows from other properties of the memes itself that are enhancing fitness only because of the essential relation between the substantive properties of the meme and its selective environment.

The essential thing to see is that there can thus be *no explanatory primacy of one factor over the other*. From this it follows that if the claim that the fitness of memes explains diffusion wants to circumvent the tautology problem, it cannot – at the same time – claim any causal priority of memes over humans, since they are the selective environment of memes. This is the central point I want to draw from the critique that meme selectionism is subject to a tautology charge.

As just mentioned, Dawkins, Dennett, and Blackmore know this, but they do not seem to see the consequences for their claims. One of the reasons, why they cannot see it lies in the following: They simply define what I call the selective environment in human individuals as 'other memes': memes are thus selected by other memes, which we acquired before. The human mind is thus crossed out of the picture. But, first of all, this is true only if *all* things that constitute the selective environment can be reduced to memes. That this is not the case will be defended in section 5.5. But even if that were the case, my essential point against memeticists would still hold: What is at issue in explaining an instance of cultural change is the fitness of a *specific* meme, given its selective environment. It is important not to forget that the selected meme is different from the selecting memes, the latter making up the selective environment of the selected meme, together with other factors. The fitness of each individual meme is thus never independent of what makes up the mind – its structure and already acquired beliefs and preferences. The general independence claim is wrong. Not the meme whose diffusion is at issue determines its own fitness. The relationship between this meme and its environment determines its fitness. This holds for all cases, since otherwise ‘survival-of-the-fittest-meme’-statements would be tautological.

### Conclusion

The point I want to make against memetics with respect to the tautology problem is that simply referring to the fitness of memes, to a high survival value, or to memes’ attractiveness is not giving any answer to the question why certain ideas spread or not. Although memeticists cite the human mind as one factor that determines, apart from intrinsic properties of the meme, the fitness of memes, they ignore that memes and minds are not two different factors. They ignore that it is the relation between minds and memes that constitutes the fitness of memes. Fitness of memes is a consequence of the relation of memes to their selective environment, which is dominated by human minds. Fitness of memes is thus not a ‘cause’ that explains diffusion on its own. Fitness of memes can only explain diffusion, if the fitness is explained by the causal basis of the differential propensity for the survival of memes. This causal basis, a property of memes, arises from the relationship of the meme to human minds. There is thus no survival value of memes that is independent of the human mind as selective environment.

## 5.3 THE EXPLANATORY DILEMMA

### Caught in a dilemma

It follows from the conclusion of the last section that memetics needs an ‘ecology of memes’ that investigates the adaptivity and selective environment of memes. But it also follows, and this is the important consequence, that the second main claim of meme selectionism cannot be defended. Memetics does not provide an alternative to the traditional explanation via memes that have a fitness of their own that is independent of our interests. For if memetics wants to get out of the tautology problem, it gives a *redundant* retelling of the traditional explanation by pointing to the mind as the selective environment of memes. In other words, if the first basic claim that the fitness of memes explains diffusion is true and non-tautological, then the claim that this explanation provides an alternative to the traditional model of explaining diffusion is wrong. This leads memeticists into a *dilemma*: The explanatory units of selection analogy either rests on a tautological statement of the 'survival-of-the-fittest-meme' and is thus *explanatorily trivial*, or, if memetics wants to escape the tautology problem, it ends up with the traditional explanation and is thus *heuristically trivial*, re-phrasing the traditional explanation. It cannot defend the two basic claims (1) and (2) at the same time. This is what I call the *explanatory dilemma* of the explanatory units of selection analogy. Although this conclusion should be evident as a consequence from the last section, let me illustrate in this section: first, the dilemma itself in more detail; second, why memetics ends up with this dilemma; and, third, what follows from it.

If we have a traditional explanation, citing all the psychological factors that determine the selection of certain memes, we get the following kind of explanation of diffusion: There are human individuals, with minds that can easily remember or learn certain things and not others, minds that make their decisions about adopting a meme or not according to previously acquired skills, beliefs and values. Humans then select memes, because they understand these memes, can remember them, and because they appreciate the properties they see in them. Now, is there some explanatory work left to do for memetics? I think there is nothing left to do at this level of explanation, a level of explanation that is typical for social science and folk-psychology. The explanatory units of selection analogy does not say more than that we adopt memes because of the way we are. So it is *not* enough to acknowledge, as Aaron Lynch does, that memetics

> "is hardly offered as a replacement for all existing social science. Historiography, psychology, economics, sociology, anthropology, political science, and other fields remain as vital as ever. It is simply inappropriate to demand that memetics explain everything about a social phenomenon" (Lynch 1996: ix)

It is not enough to acknowledge this, since at the same time Lynch presents memetics as "an important and long overdue addition to social science" (Lynch 1996: xi). Nobody expects memetics to explain everything, but their two basic claims should at least be valid. Given the explanatory dilemma, memetics is not even an important *addition* to social science. It is a redundant and therefore trivial rephrasing of the traditional social science and folk-psychological explanation of diffusion – a rephrasing in Darwinian terminology, i.e., in terms of the fitness of memes.

While a traditional explanation does not need this rephrasing in Darwinian terms, the rephrasing in Darwinian terms does indeed need the traditional explanation, in order to give the first basic claim, that fitness of memes explains diffusion, some explanatory force. Explanations in terms of 'actual survival,' i.e., explanations in terms 'being catchy,' 'having great psychological appeal,' or 'infective power' *explain* why we believe what we believe and not something else, only if these explanations refer at the same time to the essential relation between the properties of the meme and the respective selective environment in human minds, which is part of a traditional explanation.

### Misleading analogies

The latter point can also be illustrated by simply playing out the analogy-game a little further. This will also show why meme selectionism ends up with the above-mentioned dilemma. *Saying that memes replicate independently of the individuals who set the selective criteria – consciously or unconsciously – would be analogous to saying that genes replicate independently of their selective environment.* That this does not make sense should be obvious. Formulating the issue in this way shows that the answer to the 'who's in charge' and 'cui bono' question should be the following in the case of cultural diffusion: The entity that plays the role of the environment; the entity that 'does' the selecting. In the case of culture this is the individual. 'Mother Nature' does not have preexisting preferences and she does not replicate, albeit she does persist. We do have preferences, but – as 'Mother Nature' – we do not replicate like memes or genes. Nevertheless both – the selective environment of biological evolution and the selective environment of cultural evolution – select and in this sense *direct* the evolution of the replicators. And both do this always; it cannot be other. The selective environment of a meme is thus *a* *primary causal factor in* the replication of ideas, a factor that is not reducible to properties of the ideas whose selection is at issue. This is what ‘selecting’ means. Therefore, it is wrong that ‘no one is watching,’ as Blackmore says.

It just does not make sense to speak of memes replicating irrespective of the interests of individuals, as it would not make sense to speak of genes replicating irrespective of properties of their selective environment. Memeticists almost directly transmit the units of selection debate to cultural change. For gene they insert meme. For phenotype or vehicle they insert person or mind. For genes versus organisms as units of selection, they insert memes versus person or mind. And since we learned to talk about ‘selfish genes,’ we now talk about ‘selfish memes’ and their ‘interests’ in replication. For the fitness of the organism, and the corresponding interest of the organism in its survival and reproduction, memeticists insert interests, values, or beliefs of the person. They ask ‘cui bono’ and ‘who’s in charge’ and answer: the unit of selection of culture, which is the meme. But the whole units of selection issue, the whole talk of ‘interests’ of memes versus individuals, of memes as being the more fundamental causal agent, is misleading, since meme selectionism tends to misconstrue the role of the individual. They tend to stress the role of humans as culture bearers. Yet the most important causal role of a human person in diffusion is its role as selective environment.

The latter point directly leads to another reason why the explanatory units of selection analogy ends up in the described dilemma. Memeticists tend to ignore that the relationship between individuals and memes is a different one than the relationship between organisms and their genes. The ‘vehicles’ of genes do not play the same role in biological evolution as people play in cultural evolution. It would thus be better not to regard people as ‘vehicles’ of memes. As illustrated in chapter 3, when the ontology of memes is at issue, memeticists tend to regard artifacts or patterns of behavior as the ‘vehicles’ or ‘phenotype’ of memes. These entities interact with the selective environment (humans) and make the replication of memes differential in their selective environment. For instance, the ‘organism’ (i.e., the phenotypic expression) of the idea of a hammer is a hammer and not the person who uses the hammer.

Furthermore, the original debate about units of selection refers to the question which unit is the ultimate unit that survives over the long run and is selected because of certain properties. Memeticists wrongly transfer the debate about units of selection to the cultural realm not only because they tend to misconstruct the role of individuals. The debate is also wrongly applied to cultural selection, since it was never at question whether individuals *or* memes are selected because of certain effects. Although individuals change or *evolve* through culture, since they incorporate memes, identify with some of them, and can be said to change through this process, individuals just *are not selected by cultural diffusion*. Hence, if we want to use Darwinian thinking for culture at all, then the only real *evolutionary* role humans can take in this game is the role of the environment.

All this shows that, first, for culture as well as for biological evolution, it is ridiculous to state a conflict between a gene or a meme and its selective environment, even if the environment includes other genes or memes. Second, for culture it would be ridiculous to state a conflict between the meme and the cultural analog of the organism, which is the hammer. These are the deeper reasons why the explanatory units of selection analogy states a false contrast between fitness of memes and individuals.

Memeticists swing back and forth between two perspectives: from describing the individual as a 'host' or a 'vehicle' of memes, to describing the role of individual as selective environment. Take the two following statements from Dennett:

> "The likelihood of a recipe getting *any* of its physical copies replicated depends (mainly) on how successful the cake is. […] at getting a host to make another cake […] at getting the host to make another copy of the recipe and passing it on" (Dennett 2002: E-88).
>
> "The fate of memes – whether copies and copies of copies of them persist and multiply – depends on the selective forces that act directly on the physical vehicles that embody them" (Dennett 1991: 204).

It is not that Dennett changed his mind between making these two statements, the first in 2002 and the latter in 1991. He constantly changes between these descriptions. This creates confusion and shows the central tension in the explanatory units of selection analogy: The passive vehicle of memes is *at the* *same time* the active, selective force that determines the fitness. Whether the memes can get the host as *passive* vehicle to replicate them depends totally on the *same* host as the main *active selective environment*.

I suggested above that it is better to leave out the vehicle-metaphor and the host-metaphor altogether, since the real evolutionary role individuals take in cultural diffusion is the role of the selective environment. The vehicle- and host-metaphor do no explanatory work, since for the goal of explaining the fitness of memes this role of humans is irrelevant. What is relevant for this goal is their role as selective environment. In addition, to call humans 'vehicles' is heuristically trivial for another reason: It adds nothing to the traditional conception of humans as culture bearers. The metaphorical language of humans as passive vehicles or hosts is only motivated by a misguided application of the units of selection debate to culture – an application of Darwinism to culture that leads nowhere.

### Bookkeeping

In culture, all the active work is done by humans: They generate memes; they generate the phenotypic expressions of memes (artifacts etc.); they learn memes through inferential reconstruction from others; and – most importantly – they select them. Humans are the ultimate agents of cultural change. Recall that critics objected to gene selectionism that gene selectionism ignores the causal role of organisms as interactors. Whether that is right has not been addressed in this study. The only thing I do claim is an analogous critique against meme selectionism: Meme selectionism ignores the causal role of humans as creators and selectors of memes. If we take these roles into account, memes are mere bookkeepers of cultural change.

Meme selectionism, as a description of diffusion, does indeed work always, as gene selectionism as a description of biological evolution does work always. We can describe diffusion as the differential spread of memes due to their fitness differences. However, to state that this shows that memes and not minds are the primary causal factors, or 'selfish,' i.e., the sole beneficiaries of the diffusion process, is distorting the causal picture. Just as the meme's eye perspective, the traditional perspective works equally well. Every case can be described at the same time from the meme's point of view and from the individual's point of view. Saying that there are cases that cannot be explained by reference to facts about the individual would be similar to saying that there are selective processes that cannot be explained by reference to a selecting environment. The traditional view can explain *every* instance of meme selection. And *only* the traditional perspective *explains* such instances of diffusion. The individual's point of view is not yet a full explanation. It certainly is a 'placeholder' for a deeper explanation – a placeholder for basic cognitive processes, and a million of different neurological, sociological, or ecological factors. Nonetheless, the meme's point of view alone does not give *any* explanation, except when it is combined with the traditional explanation.

In addition, since memes are also created by individuals, meme origination and meme survival are, all things considered, the effects of the causal power of individuals. Origination and diffusion of memes is caused by humans. Memes exist and spread because of the meme-creating, meme-transmitting and meme-selecting power of humans.

## The role of content

In order to prevent misunderstanding, I would like to add a last clarifying point. The content of memes certainly plays a role in determining the diffusion of memes, even though memes are created, transmitted, and selected by humans. The meme's eye perspective and the traditional perspective of explaining diffusion are no alternatives because the answer is not 'either – or.'[15] In reaction to meme selectionism some might say that the selective environment itself is the cause of the adaptedness and hence of the diffusion of memes. This is true in one sense, but not in another one. It is true in the sense that memes cannot have a fitness that is not dependent on their selective environment. If we, however, then say that the selective environment has a total priority, merely by being the selective environment, we end up with

---

[15] Midgley (2002: 130-133), for instance, makes this mistake in claiming that it is the complexity of human motivation that explains why memes spread.

another too radical claim. Contents of memes certainly play a role in determining the fitness of memes.

Again, a clear comparison with biological evolution helps to see the point: The selective environment of organisms is not alone determining the replication of genes and reproduction of organisms in biological evolution. There are two reasons for this: First, the organism is not only a passive thing that faces an environment. The organism selects his selective environment.[16] If I am correct, this cannot be transferred to culture, albeit, in culture, memes take the role of genes. Memes and their 'organisms' (i.e., ideational units and their outward expression) are passive in the sense that they are mere consequences of the agencies creating them, making decisions and conducting actions. In the case of culture the only active thing, making decisions, having desires etc., is the selective environment, human beings.

But, secondly, for biological evolution as well as for cultural diffusion, it is imprecise to say that it is the environment that determines fitness and thereby *causes* evolution. As I said in section 5.2, what determines fitness is the *relation* between a selective environment and the respective entity whose evolution is at issue. What explains the frequency changes and the corresponding evolution of organisms? – Neither the environment, nor the traits of the organism. Asking 'what explains the trait, the environment or the trait' is ridiculous, since the opposition is ridiculous. Why do giraffes have such long necks? – Because they lived in an environment with high trees? Because there were some giraffes with long necks? In both cases the answer is 'No.' Giraffes with long necks have spread and persist until today, since long necks in an environment with high trees leads to a higher propensity for survival and reproduction of the individuals with these necks, and hence to a higher replication of the genes that are connected with this trait.

The same holds for memes: The *relation between memes and minds* – and not minds *or* memes – explains the fitness, as the relation between the long neck of the giraffe and the respective environment explains why the giraffes

---

[16] Well-known for his critique of adaptationism that interprets the organism as totally passive is Richard Lewontin (e.g. 1985 [1983]).

with long necks have survived. Content plays a role. But since memes are not only selected by humans, but also created by humans, they are explanatorily secondary, mere consequences of the activity of humans. The ultimate agents of culture who are to charge for cultural change are humans, not memes.

### Conclusion

Because of the tautology problem, the explanatory units of selection problem, faces an explanatory dilemma: If meme selectionism wants to get out of the tautology problem, it ends up with the refutation of its second basic claim, namely that memetics is an alternative to the traditional explanation. The explanatory units of selection analogy is either tautological or heuristically trivial. Memeticists end up with this dilemma, since they misconstruct the role of individuals in diffusion. A transfer of the units of selection debate to culture does not make sense, since the relation between memes and minds is significantly different from genes and their organisms: Minds are not the analogue to phenotypic effects of memes. They are – first and foremost – the selective environment of memes. To explain diffusion as the replication of 'selfish genes' is thus distorting the causal picture, because the selective environment is a primary causal factor in any evolutionary change. In addition, since memes are not only selected, but created by humans, their fitness is a mere consequence of the power of humans. Memes 'keep the book' of diffusion, but they are not ultimate causal agents that replicate irrespective of what we are, think and want, even if their content certainly plays a role in our selective choices.

## 5.4 MEME FITNESS AND IRRATIONALITY

### Limited independence

What I have stated so far provides a general argument that applies to any cultural item that is remembered, talked about, or adopted by individuals. But as indicated in section 5.1, meme selectionism is sometimes restricted to cases of irrationality. This claim is used to provide justification for the claim that meme selectionism provides an alternative to the traditional explanation: At least for cases of irrationality, meme selectionism provides a striking new perspective. Dennett (1995, 2001a, 2001b) and Dawkins (1976, 1993) frequently cite examples of ideas that spread in spite of "our judging them useless" (Dennett 1991: 203). These are the "viruses of the mind" (Dawkins 1993). According to Dennett and Blackmore,[17] these are the cases that need a *special* explanation that can *only* be given, if we invoke memes as the selfish units of selection of culture. In those cases memes can spread, even if they are useless or judged useless. On this basis, memeticists could answer that the critique I developed above only applies to memes that are useful for us, but not to those memes that spread independently of our "fitness by whatever standard we judge that" Dennett (1991: 203).

What is the fitness of humans, if not judged by the standard of their biological survival and reproduction? 'Fitness by whatever standard we judge' is what we value as useful for various goals. Fitness in this context can only mean benefit. Whatever benefit individuals have from their mental and behavioral operations can be considered as contributing to their 'fitness' in an abstract sense, as they judge it. On this interpretation, the core of the limited independence claim is that there are instances where the individual selects certain memes, where the relation between properties of the individual and properties of the meme determines thus the fitness of memes, but where, at the same time, the resulting fitness of memes *does not correspond with a benefit for the individual*. The 'interests' of memes and the interests of individuals stand in conflict. Recall that gene selectionism also claimed that there are outlaw genes, i.e., selfish genes in the narrow sense that circumvent the Mendelian fair lottery of reproduction. They can spread, although they do not provide an advantage or even a disadvantage for the organism. The 'interests' of these genes are thus in conflict with the interests of the organism. Analogously, meme selectionism states that there are memes that spread, although they are not useful for us. Now, if this should give the explanatory units of selection analogy a heuristic value, it must hold that these conflict cases can be predicted and explained *only* from the meme perspective. Only

---

[17] See Dennett (1991: 205), or Dennett (2001a, 2001b). See also Blackmore (1999: 176).

then is Dennett correct in stating that the gene-meme-analogy provides a striking new perspective.

### Conflict cases

In order to show that the conflict cases can also be predicted from a perspective that does not involve any reference to an analogy between genes and memes, I will look at three kinds of cases, where memes spread, although they do not provide benefits for the person who adopts the respective meme. The three categories are: (a) cases, where people *think* it is advantageous for them to adopt a meme, but where the *outcome* of the selective process comprises no benefit for the individuals – from a perspective that is *not the perspective of the individual adopting a meme*; (b) cases, where people *think* it is advantageous for them to adopt a meme, but where the *outcome* of the selective process comprises no benefit for the individual from *the perspective of the individual*; (c) cases, where the people *do not even* think that adopting the meme is advantageous and do it nonetheless.

*(a). False or costly but subjectively appreciated memes*. Cases of the first category can easily be reconstructed from a traditional perspective. The whole issue depends on what it means that something provides a benefit for somebody. Given the goal of having scientifically justified beliefs, a belief in God might not provide a benefit for the person and might thus be an irrational belief – from a scientific perspective. But the person believing in God might just not value scientifically justified beliefs as high as other people do. The person might well have other basic preferences and believing in God might provide indeed a benefit for the person, given these preferences. From the perspective of the person, it might thus be rational in a subjective sense to believe in God. Adopting the meme would thus not be a case where a meme spreads despite the fact that it is useless for that person. The person might well *think* that he has a benefit from believing in God and so adopts the meme for that reason. And the person might think so, even if the belief has led to

considerable costs, e.g., to a life that others would judge negatively.[18] In a nutshell, adopting the god-meme might be adopting a false belief from the scientific perspective, or might lead to a costly life, and nonetheless be subjectively rational – given certain goals and preferences. Take an example: A person that wants to circumvent its existential anxiety about death might have a benefit from believing in life after death and the supernatural deity guaranteeing this eternal life. The person might simply not care so much about scientifically justified beliefs or worldly costs that follow from the belief in a supernatural deity and life after death. The highest value of the person is to circumvent the anxiety. Hence, the belief serves a major goal in the person's belief-value system. Given what the person tries to achieve, given which memes are available in a certain culture, given what the person knows already, the belief of the person in such a deity can be reconstructed as 'rational' from the subjective perspective of this person, as long as the person is rational in the sense that it calculates its utility function according to the procedural standards mentioned in section 5.1. Whether the memes that are adopted according to the belief-value system of that person are objectively justified or not is a totally different question. But as I explained, the traditional perspective, regarding humans as the creators and selectors of culture, does not refer to such an objective standard of rationality. We can regard what the person does as objectively irrational or as rational, in both cases the traditional perspective could explain and predict such a case, since the person does what he does because he thinks that it is useful for him.

Now, does meme selectionism provide us with something that gives an alternative explanation of such a case? No, since it would have to claim that the individual had *no reason at all* for choosing the meme, i.e., that we cannot give a story in terms of subjective reasons of the individual. But a religious person surely has a reason for selecting the god-meme. This reason constitutes the

---

[18] That Dawkins describes religion as a virus seems to be based on a pure evaluative basis, assuming – in addition – scientifically justified knowledge as the only legitimate basis of any kind of belief: According to Dawkins (1993), science is not a 'virus,' even though it spreads in the same way as religion. It is not a virus since it is subject to more stringent selection criteria, subject to standards that would guarantee that science is objectively rational.

relation between mind and meme that accounts for the fitness of the god-meme. As long as we assume that religious people adopt the god-meme for a subjective reason, whatever that reason is, and as long as they appreciate what follows from believing in God, then adopting the god-meme is in fact not a 'conflict case' where a person adopts a meme, although it does not have a benefit from it. The case is just a standard case where a person does what he does, because he wants to do it.

*(b). Unintended consequences.* Cases of the latter kind have to be distinguished from cases that involve *unintended* consequences of adopting a certain meme. These are cases where somebody thinks that it is beneficial to adopt a meme, but it turns out that it is not. The person has to admit that adopting the meme has led to outcomes that are judged negatively from the perspective of the person adopting the meme. Such cases of unintended consequences are not unusual. We constantly make errors in our decisions. If we assumed a kind of perfect rationality, with perfect foreknowledge, then such unintended consequences of adopting ideas or patterns of behavior would indeed be irrational. But we just are not perfectly rational and we know that. Being subject to a chain letter is a simple and perfect example that memeticists like to use as an example of a meme that behaves like a 'virus' – invading our mind despite the fact that it is useless for us. Imagine that somebody comes into contact with a chain letter for the first time in his life. The person believes that adopting a chain-letter-meme and acting accordingly provides a benefit for him, given his belief-value system. It turns out that this is wrong. Can we explain such cases with the traditional intentional stance perspective, which assumes that people do what they do because they think it is useful? We can, since such cases of irrationality can be explained by acknowledging that humans often do not have an objectively justified basis for evaluating certain memes. It might also turn out that the person was inconsistent in his preference system. But first, as I have illustrated in section 5.1, this is not part of the contemporary concept of rationality used in social sciences. Our rationality is bounded. Second, individuals select memes for a reason, or a couple of reasons, and these reasons, even if sub-optimal reasons, build the selective

criteria according to which these people select memes. The meme spreads according to these reasons and not despite these reasons. Thus, these cases of irrationality can also be reconstructed as normal cases of the traditional belief-preference model of choice, where a meme spreads because somebody thinks that it provides a benefit. The traditional explanation does not fail in these cases and the claim that memes spread despite our disapproval of them can simply not be applied.

*(c). Weakness of will.* Memeticists might answer that this is correct and restrict their claim to cases where we do not even think that a meme provides a benefit, when we adopt it. These cases would be cases, where we do not want to believe or think something, where we do not want to do something, and think or do it nonetheless. This is *weakness of will.* There are sometimes conflicting interests in a person and one interest gains predominance, though the person does not want that this interest gains predominance. The person has higher-order evaluations that structure the list of his first-order values or interests. There might be motivations that drive the person towards a certain behavior, although he – from a reflective perspective of his higher-order values – does not approve these first-order motivations. In addition, the driving motivation might even be an unconscious motivation. Unconscious motivation leads to unconscious selection. Take the example of computer games. Memeticists would say that ‘playing computer games’ spreads because this meme is a good replicator, and not because we select it according to our beliefs and values. Indeed, ‘playing computer games’ seems to be a pretty good replicator, at least in the wide sense, and with respect to some humans. If a person does not want to play a computer game and does it nonetheless, the person is definitely irrational, since the person is not *consistent*. But even if somebody falls victim to that meme, the person still plays computer games *for a reason in the wide sense*: unconscious motivation or a first-order motivation that gains predominance despite higher-order conflicting values. Imagine that somebody wants to do his work at the computer. Eventually the person ends up playing computer games. Why? Maybe there was a need in the person for relaxation or an irresistible desire for this particular game. The game is played,

because it fulfills these needs or desires, even if the person – in principle – dissociates from these desires. The person does it, although the person does not identify with these actions. The person does not want that he wants to play the game. But the computer playing is not explained by a fitness of the computer-play-meme that is disconnected from the person's reasons. The fitness arises from a relation of this meme to certain reasons, in the wide sense, that the person has for playing the game. The computer playing is explained by weakness of will or unconscious selection. That means that the traditional belief-value explanation of the behavior does not fail for such cases of irrationality.

I would like to draw a first conclusion about the limited independence claim. The important point is that memetics can only be an alternative to the traditional explanation, if we exclude weakness of will and unconscious motivation from the traditional explanation. Yet a traditional explanation does not have to exclude it. The traditional explanation does fail, only if we construct it in the form of a *straw man*. If the traditional explanation is considered to refer to a too rationalistic picture of the human mind that assumes that humans have perfect knowledge and foresight, that humans are always consistent and never subject to weakness of will, then this traditional intentional stance perspective does indeed fail for the conflict cases. But as I illustrated in section 5.1, neither empirical research, such as diffusion studies, nor the contemporary concept of rational choice assume such rigid standards as a realistic model for decisions processes.

Let me add a note on cases that most directly remind one of 'viruses of the mind,' penetrating our mind, despite the fact that we do not want them to do so. We all know ear candies – tunes, sentences, or thoughts that just do not want to get out of our mind. We hear something and hum it over and over again. We want to concentrate on something else, but the melody comes back. They are so easy to memorize and catchy so that we cannot do otherwise. We are attracted by it, as we are attracted by other things, for instance, by another person. Certainly, the attractive person or the attractive meme plays a role in the explanation why I am attracted to this person or meme. It is because of certain properties that I am attracted, but my predisposition to be attracted by such a person or meme plays an equally important role. As I said already, in accordance with Sterelny, change the selective environment and other ideas or persons are attractive. Take again the memes with long chains of numbers. They are unlikely to become an ear candy, since we are not built to memorize them easily. That is why they do not have a chance to become a meme with a high replication potential. In computers or in Martian creatures, they might have a good chance to find a good selective environment, and a tune from Mozart might in turn have no chance at all to find a good selective environment in computers or Martian creatures.

These cases point to the fact that there are factors that determine the selective environment of memes, factors that are not reconstructable as part of the belief-value system of a person: properties of the mind that determine what we can easily comprehend, or what we can easily memorize. These properties of the mind are not connected to any kind of benefit for the individual, except the benefit one gets from the good feeling of being able to understand or memorize something. The fact that we cannot remember long chains of numbers is such a property of human minds. However, as said already in section 5.2, that easy-to-remember memes spread more likely than others does not follow simply from an intrinsic property of the memes. It follows from their relation to our minds. This relation makes them easy to remember for *us*. Therefore, these cases can also not provide an explanation that is not given by a traditional explanation, since this traditional explanation does not have to exclude such properties of the mind.

### Explanatory dilemma still holds

I have discussed examples that might be considered as cases where a meme spreads despite the fact that it does not provide a benefit for the person adopting the meme. For all cases, however, it could be shown that there is an explanation that refers to certain pre-existing values, beliefs, and features of the human mind that are not reconstructable as values or beliefs. In all cases, we can refer to these factors in order to explain why the respective people do what they do. Hence, it can be concluded that, first, there is not even a limited independence of memes from properties of the mind that holds for those cases of irrationality.

Second, we *have to* refer to these properties of the mind, since the explanatory dilemma also holds for these cases: If meme selectionism wants to get out of the tautology charge, it ends up with the refutation of their second main claim that the spread of memes is due to the fitness of memes and not due to individuals – due to their 'interests' and causal role. Meme selectionism can thus not provide an alternative to this traditional perspective. The limited independence claim is therefore as *explanatorily trivial* as the general independence claim.

Third, we do not need an analogy between 'selfish genes' and 'selfish memes' in order to explain these cases of supposed irrationality. We do not need the analogy to show that we are rational, and we do not need it to show that we are irrational. It is correct that we are not always perfectly and objectively rational, since we are subject to limited knowledge, to weakness of will, and certainly to a myriad of other shortcomings. It is also true that we humans sometimes like to think of ourselves as the rational controllers of what we think and do. But although the idea of a perfect rational mind exists as an ideal, it is at the same time clear that we actually are not always rational in an objective sense, and often not even in a subjective sense. Furthermore, our conscious preferences and conscious beliefs do not explain everything, we have to take other properties of the mind into account as well, for instance that we cannot easily remember long chains of numbers. In short, there is no perfect rational subject that is the ultimate master of its mind.

A rational choice model that assumes perfect subjective or objective rationality could thus not explain all cases of human decisions. But today social sciences, psychology, economy, as well as philosophy do take this into account. To say that the traditional perspective, which has the human mind at the center, cannot explain cases of irrationality is thus fighting against a *straw man of a theory that relies on the idea of a perfectly rational human mind.* If such a rigid rational choice model were all the traditional explanation comes down to, memetics would indeed have a point. But since the traditional belief-value model does not assume a perfectly rational human mind, as I illustrated in section 5.1, we do not need the explanatory units of selection analogy in order to see that we are not perfectly rational. That humans are not perfect is a fact we know from history; and it is a fact that is easily visible from theoretical perspectives that do not involve any Darwinian analogical reasoning from nature to culture. The limited independence claim is also in this sense *heuristically trivial*.

Last but not least, a model of the human mind that takes into account that our rational capacity is limited, that our will is sometimes weak, that our rationality is bound by numerous structural facts of our human mind, can indeed offer an explanation of irrationality, at least at the level of the intentional stance perspective, which is typical for social sciences and folk-psychology. Certainly, this is not a full explanation. The difficult task – from within the traditional model of explaining diffusion – is to find out more about human minds as selective environment of memes. For a deeper explanation, one would have to find out *which facts and which reasons* influence the understanding, memorizing, and adopting of memes; one would have to find out *why* the selective environment is the way it is, how this relates to basic cognitive processes, how this relates to the material level of the brain etc. For instance, if we explain a certain behavior by saying that the person made a mistake, or that the person did not confirm to his second-order beliefs, standards or values, these explanations still have to explain *why* a person makes mistakes, why the person does things or believes something that he – in his reflective identification with himself – does not want to do or believe etc. Or consider an explanation why we cannot remember long chains of numbers. Or why can't we just ignore ads? Why can't smoker not just stop smoking? Why aren't we more rational? What are the basic cognitive processes accounting for our decision processes? Each of these questions needs careful investigation, covering many research areas in social sciences, psychology, and philosophy. In short, the entire social sciences and the humanities are dealing with these questions and the answers will vary with the kinds of irrationality at question. In principle, these disciplines can give an answer. They would provide the ‘ecology of memes’ needed to find out why certain memes are fitter than others, given a certain selective environment.

### Conclusion

I have explained that even in cases of irrationality, ‘survival-of-the-fittest-meme’-explanations have to refer to diverse properties of the mind and to the respective reasons that individuals have for adopting a meme. They have to do so in order to escape the traditional tautology problem of evolutionary theory. A traditional explanation that is not based on a too rigid concept of rationality can account for these cases and does not need the explanatory units of selection analogy in order to explain these cases. What such a traditional explanation needs to make progress is not the explanatory units of selection analogy, but further investigations about decisions on the basis of limited knowledge, about structural features of the mind, about weakness of will etc. The explanatory units of selection analogy cannot offer any new insights because of the explanatory dilemma. If meme selectionism wants to give any explanation of diffusion, it merely gives a re-telling of the traditional story in Darwinian terms.

## 5.5 MINDS AS BUILT BY MEMES

### Reducing the selective environment to memes

My critique of the explanatory units of selection analogy rests on the following claim: The tautology problem shows that the replicative power of memes is merely the consequence of the relation between these memes and the respective individuals that are their selective environment. Since the content of ideas is created and selected by individuals, the fitness of memes is only a consequence of the activity of individuals. This is tantamount to what Dennett has called the traditional explanation. Dennett has offered an argument that would make this critique pointless. According to him, a human mind is nothing else than a conglomerate of memes. Therefore, the contrast or conflict between us with our interests and memes with their ‘interests’ in their survival, is resolved through reducing us to memes. In Dennett’s own words:

> "The haven all memes depend on reaching is the human mind, but a human mind is itself an artifact created when memes restructure a human brain in order to make it a better habitat for memes. […] But if it is true that human minds are themselves to a very great degree the creations of memes, then we cannot sustain the polarity of vision with which we started; it cannot be 'memes versus us,' because earlier infestations of memes have already played a major role in determining *who or what we are*. The 'independent' mind struggling to protect itself from alien and dangerous memes is a myth. […] Our existence as us, as what we are as thinkers are – not as what we as organisms are – is not independent of these memes" (Dennett 1991: 207f; Emph. in the orig.).

First of all, it is important to be precise with respect to *what kind of thing* is in fact created by memes and is, therefore, only a consequence of them. Blackmore says that memes show that there is "no one watching," meaning a '*conscious designer*' (Blackmore 1999: 242). In other places, she talks about the *self* as being a "memeplex" (Blackmore 1999: 219) or a "pack of memes" (*ibid.*: 235). Dennett talks about '*mind'* that is *partly* build by memes and memes playing a "major role" in determining 'what we as *thinkers* are.' In other places, he talks about the "*self*" or "*ego,*" the "*I*" as the "captain of my vessel" (Dennett 1995: 366-368). According to the latter, it is not the *mind as such*, but the *conscious self* that is made out of memes. And this is also why Dennett included memes in his naturalist explanation of consciousness (Dennett 1991). Memes install a 'virtual machine' in our brains. This virtual machine only *seems* to do all the meaning, believing, decision-making etc.

I will first of all explain in more detail which role memes play in Dennett's naturalistic philosophy of mind and consciousness. I will then show that the strategy to reduce minds to memes fails to provide a way out of the explanatory dilemma and the critique I have built on this dilemma.

### Dennett's naturalistic theory of consciousness

According to Dennett's naturalistic theory of mind and consciousness (Dennett 1991), there is no conscious Cartesian ego that is free, that does all the thinking, decision-making, meaning etc.[19] The folk-psychological impression that there is such a central I or self is considered by him to be a mere useful illusion. What he means by this can best be explained the way he does, by

---

[19] See also Dennett (1995, 1996, or 2003) on various aspects of his naturalist theory of mind.

using an analogy to computers: The only two things that exist in our minds are a serial working “virtual machine,” analogous to a computer program, and a parallel processing brain, on which this ‘virtual machine’ is installed. Despite the existence of this virtual “user interface,” which is as real as a computer program, all the work that is actually done – at a causal level – is done by the brain, as the computing in a computer is at the basic level done by the hardware of the computer (Dennett 1991).

The important issue for this study is how the brain acquired such a virtual machine of consciousness. According to Dennett, it evolved. Conscious minds are “cranes” – created by evolution and speeding it up, since they are useful devices. They are not “skyhooks,” falling out of heaven, i.e., emerging out of nothing like a miracle that is unexplainable by a naturalistic perspective (Dennett 1995: 73ff). Recall from section 4.4 that Dennett believes that there is a “tower of generate-and-test” (Dennett 1995: 373-381): The trial-and-error of biological evolution, where each trial is selected by death of the organism, is replaced by other trial-and-error processes. This is how we are “losing our minds to Darwin” (Dennett 1995: 370). Evolution went from genes, simple replicators, over Skinnerian creatures, up to Gregorian creatures. Humans are Gregorian creatures. They have an inner representation of the external world and internalized mind-tools, such as words and numbers. On the basis of this, we produce Campbellian ‘thought trials’ and select them.

Such internalized selection processes presuppose consciousness. Evolution has led to a consciousness that makes conscious internal selection processes possible. This includes that we can describe the evolved ‘virtual machine’ of consciousness as a Cartesian ego that consciously communicates its own processes. Mind has intentional states and makes decisions about what to believe and what to do. What the virtual machine does is as real as what my word processor does, even though all what it does is due to basic electronic processes at the level of the hardware. The ascription of consciousness and intentionality is made from what Dennett has called the intentional stance (Dennett 1987), which detects certain ‘real patterns’ that are not observable at the level of the brain.

In *Consciousness explained* (Dennett 1991), Dennett explains *how* evolution has led to a plastic brain with ‘autosimulation’ of a conscious self. The details of how the plastic and conscious human mind evolved are not important here. What is important is how memes enter the picture. According to Dennett, memes are important in two senses: phylogenetically and ontogenetically.

Some habits that – back in the times of the early *homo sapiens* – were socially transmitted, have become innate. In the language of the computer-analogy: They became hardwired. One of the so-called ‘just-so-stories’ Dennett tells about what happened in the evolutionary past is, for instance, the following: Creatures with proto-language started to give information to others, when they were asked to do so. This has then led to the ‘invention’ of auto-simulation: Talking to oneself. Later, some of these habits or abilities got hard-wired. In this way, consciousness and language co-evolved. A little bit of mind leads to proto-language, proto-language leads to a more efficient mind, to more language, to a more efficient mind, and so on, until we reach a human mind with consciousness and all the abilities that we have as a matter of fact. What is important for this investigation is that ‘habits of minds,’ such as the habit of talking to each other, are regarded as memes. These memes have furthered the phylogenetic evolution of a conscious self that is able to talk to itself silently. This is the phylogenetic role that memes play in Dennett’s philosophy of mind.

Ontogenetically memes still play the same role. Given the product of evolution so far (i.e., the already hard-wired basic machinery of autosimulation), further abilities or ideas will be invented somehow. When a human being is born, its mind, which is still plastic to some degree, will be filled with further memes. This happens either by individual or by social learning. These memes are thus like ‘software’ and ‘data’ that are ontogenetically acquired. Starting with hardware and ‘basic software,’ each human creates or ‘downloads’ further memes – further ‘software’ and ‘data.’ Humans learn, for instance, the habit-like meme of using symbolic notations (e.g., musical notations), and they learn data-like memes (e.g., a tune from Beethoven). Another example for such an ontogenetically acquired meme is language, in Dennett's terminology, a meme-like 'habit of mind.' On the basis of a general innate language ability, we learn to speak English or Japanese, which would be a habit-meme, and we learn the specific words with their meaning, which are different kinds of memes.

Much more could be said on Dennett's philosophy of mind. I only briefly presented those aspects that are important for an evaluation of the explanatory units of selection analogy. I will not say much more, and I will also not take sides whether his account of consciousness and his account about its evolutionary origin make sense or not.[20] I will, however, justify why I will not take sides. Philosophy of mind deals with all these questions in detailed and complex ways. Some of the most important questions are: What is a conscious self? What is the exact relation between mind and brain? What about qualia and intentionality? Dennett's claim that there is no Cartesian ego, no 'subject' that eludes a naturalistic explanation, can be criticized from various philosophical points of view. To elaborate on the arguments against a naturalistic perspective on mind as such, would lead much to far away from the present concern. Furthermore, even from within a naturalistic perspective, critical issues arise: Hypothesis about the evolution of mind and language should be taken as what they are, namely speculations. This holds, for any phylogenetic hypothesis that has almost no empirical evidence from fossils, even if one is very well disposed towards a naturalistic explanation of mind. We will never find out how it really was. Therefore, whether Dennett's evolutionary suggestions on the evolution of mind, consciousness, language, are good speculations, would demand careful analysis, which cannot be given here and need not be given here.[21]

---

[20] See Elton (2003) instead.

[21] That is also the reason why I totally ignore Blackmore's just-so-stories about the evolution of language, conscious selves, and altruism (Blackmore 1999). She uses the same strategy as Dennett: She speculates how they could have evolved. See Dennett's (1995) strong defense of adaptationist explanations in general and for some of his hypotheses in particular, e.g., the evolution of mind, language, meaning, or morality. See Gould's (1997a, 1997b) equally forceful critique of what has called 'Darwinian fundamentalism.' For a good introduction to debates about adaptationist explanations see Godfrey-Smith (2001).

The only thing that is important for this study is what he says about the phylogenetic and ontogenetic role of memes. But it is not whether memes indeed have the phylogenetic and ontogenetic role he states what is important here. Even if Dennett were correct about the role of memes, it would not help him out of the explanatory dilemma. This is my central critique against the claim that minds are built by memes, be it phylogenetically with respect to the evolution of consciousness, or ontogenetically with respect to the mental contents humans acquire during life. My critique rests on two arguments: Given all that Dennett says about the evolutionary role of memes, it still holds that, first, not everything that makes up the selective environment of a specific meme is itself a meme. Second, even already acquired memes in our minds do not provide a way out of the explanatory dilemma.

### Not everything is a meme

My claim that not everything is a meme can be defended from outside of a naturalistic frame as well as from within. Let us assume, for the sake of argument, that Dennett's naturalistic frame is wrong. There is a conscious self that cannot be explained as a mere 'virtual machine' of a brain. This Cartesian ego does the 'believing,' 'meaning,' 'adopting,' 'deciding' etc. If this is the case, then there is 'something' that (a) is itself not a meme, that (b) is not created by memes, and that (c) is more then the mere locus where memes interact, as Clark (1993: 13-14), for instance, has objected to Dennett's claim that minds are built by memes. If such a Cartesian ego exists, minds are not built by memes.

Be it as it may, even if we do not assume such a non-naturalistic Cartesian ego, it stands to question whether everything in the mind is a meme. Although Clark referred to a non-naturalist Cartesian ego, he makes an important point:

> "For genes to work […] there must be cells. For ideas to infect the mind, there must be minds to infect. Cells are not constructed out of genes, but by earlier cells, partly in accordance with the genetic 'instructions'. Minds, by analogy, are not made up of memes, even if they are often influenced or infected by them" (Clark 1993: 12).

Even if we regard, for the sake of argument, minds as being nothing but the vehicles of memes, memes *ontologically* need minds, as genes need cells.[22] Whatever theory of mind we assume, almost everybody would admit that the mind of an individual stands in an 'intimate' relation with (a) the brain of this individual; (b) there are certain capacities, abilities, and competences, such as consciousness, intelligence or rationality. The latter rest on competences such as language production, understanding, memory, inference making, social cognition, etc. These in turn, rest on basic cognitive mechanism, such as the basic cognitive processes mentioned with respect to creativity in section 4.5. Each book of cognitive psychology is full of descriptions of these basic processes or mechanisms. Finally, there are (c) cognitive contents (i.e., beliefs, values etc.).

Even from a naturalistic standpoint, the processes involved in (a) or (b) are *not* memes as defined in chapter 3. These processes work with memes and they are influenced by memes, but they *are* not these memes. Memes are the contents the mind is fed with, but memes are not the mind as such – as genes are not the cells that 'house' these genes, even if these genes help to build these cells. Memes certainly played some role in shaping the brain and its abilities, such as to think consciously, to memorize, to learn on one's own, to learn from others. These abilities and processes are also constantly filled with further memes. Nonetheless, these abilities *are not memes*. I can neither transmit my brain to somebody else, nor can I transmit my ability for conscious thinking. I cannot even transmit my ability to speak Japanese.

Recall what I have said in chapter 3 on memes and the triggering problem. Applied to the question whether mind is made out of memes the following holds: If our language ability is innate, then the evolution of the capacity to use and understand Japanese might well have been influenced in the evolutionary past by the social transmission of certain communication habits, which were transmitted from person to person. But the human language ability is not a meme anymore and the ability to speak Japanese involves triggering of this innate ability. I might transmit 'memes,' in this case words, to somebody

---

[22] Jeffreys (2000: 229f) made a similar point.

else. This transmission helps this person to learn Japanese. But these memes are not the ability itself. The ability is a consequence that follows from the innate ability and the transmission of particulate memes. At each point in time, there are not only memes that have been acquired before; there is always ‘something’ that is not a meme itself. This ‘something’ is part of the category (a) or (b) or both.

Dennett seems to see this, when writing, for instance, that

> “the evolution of memes could not get started until the evolution of animals had paved the way by creating a species – *Homo sapiens* – with brains that could provide shelter, and habits of communication that could provide transmission media for memes” (Dennett 1991: 202).

The evolution of memes could not get started and it will not get on even today, if a selective environment is not already present – a mind that can learn, think about, and adopt these memes. Dennett does not seem to see, or does not want to see the consequence of this for the claim that memes provide an alternative to the traditional explanation of diffusion, which always refers – at the same time – to properties of the ‘shelters’ *and* to properties of the memes, in order to explain why certain memes spread.

It follows from what I have said on the tautology problem that we explain why a certain meme has a certain propensity for survival, only if we understand in which way the brain and those basic competences of the mind involved in acquiring a meme determine the selective environment of the memes whose spread is at issue. To use again an example we have met already: That the meme ‘1096z4i5p094 zh8t46 8re9puzh’ is not easy to memorize – that it has a low fitness – is not explained by merely pointing to the properties of this meme. The low survival value is only explained, if we look at the fact that our mind has problems with memorizing such chains of symbols, because memory in humans works in certain ways. And, as I illustrated in this section, the *properties of the mind that account for our memory or language abilities are not memes themselves*.

## Autoselection is selection nonetheless

All this said, it is still true that the selective environment of memes does not only comprise the structure of our brain and basic abilities. Memes are also part of the selective environment. The fitness of memes is therefore also determined by the kinds of things summarized above under category (c): those ideas, heuristics, rules, and values that govern belief-acquisition and decisions for actions. Yet these mental contents are memes in the narrow sense, only if they are socially acquired. They need not be socially acquired, since they may also be innately specified or mental contents that are not transmittable to others. If we understand mind as comprising only those memes that are acquired and transmittable, then Dennett has a point in claiming that memes are selected by other memes. These memes, previously acquired knowledge, make up an important part of the selective environment of memes. In this sense diffusion is *autoselection* of memes: memes selecting memes.

But it is of utmost importance that the fitness of each meme is not independent of its selective environment. In each case some memes *partly* make up the selective environment, and these *other* memes are relevant for the fitness of the meme whose spread is at issue. These other memes are not the meme whose spread is at issue. Therefore, any singular meme cannot be a 'selfish' replicator, since its fitness is dependent on something else. This 'something else' has always been, and should still be called an important part of 'mind.' As I already said at the end of section 5.3, the tautology problem exists even if part of the selective environment is made out of other memes. This is so because the selected meme has to be distinguished from the selecting memes. The fitness of each individual meme is never independent of what makes up the mind at a given point in time: its structure and content. The relationship between a certain meme and *its* environment determines its fitness and not the meme itself, in all cases.

That, in culture, much of the current selective environment of memes is made out of previously acquired memes, is why creativity exhibits coupling, as I explained in chapter 4. Human cultural evolution in this sense relies on a kind of autoselection. This shows that cultural change is different from biological evolution, since the presence of previously acquired memes in the selective environment of newly occurring memes leads to coupling of variation and selection. The origination analogy fails because of this kind of autoselection. With respect to the explanatory units of selection analogy, autoselection has been used to justify that minds are built by memes. That this is not a plausible view of mind and that it would not help solve the explanatory dilemma, even if it were plausible, was the purpose of this last sub-section.

**Conclusion**

Memeticists claim that memes are the ‘selfish’ causal agents of cultural diffusion, the units of selection of culture, since the human mind is built by these memes in two senses: phylogenetically, memes are claimed to have furthered the evolution of mind and consciousness; ontogenetically, memes are claimed to fill mind and consciousness with content. There is nobody watching. There are only memes, evolving in our mind, which is a mere vehicle built by them. I have shown that this is a crude overstatement that does not help meme selectionism out of its explanatory dilemma. First, a human mind is more than its contents. Second, even if memes are involved in the generation of things like consciousness, intelligence, rationality, or basic competences of minds, it still holds that these things are not memes. Third, even for those parts of minds that are memes, the claim mixes up those memes whose selection is at issue with memes who are part of the selecting environment. No given meme has a fitness of its own – a fitness that explains its diffusion and is – at the same time – not dependent on what a human person is and regards as his interests, beliefs, values, etc.

## 5.6 SUMMARY

The original two basic claims of the explanatory units of selection analogy were (1) that the survival-of-the-fittest-memes and therefore their ‘selfish’ interest in their survival explain diffusion. Such a meme selectionist explanation has been presented (2) as a striking new alternative to a traditional perspective, which explains diffusion by pointing to human beings that have certain abilities and belief-value systems according to which these humans select cultural units. This claim was justified by saying that humans are not the ‘beneficiaries’ of cultural evolution, since the only benefit, which is fostered in all cases of diffusion of memes, is the fitness of memes. If one takes the tautology problem into account, this view cannot be sustained, since the fitness of memes is determined mainly by memes’ relation to *us* – persons with minds, building a major part of the selective environment of memes. There is thus no general independence of memes from the interest of the individuals invoked, since these interests are part of the selective environment. There is also no limited independence: Even if there are irrational cases that show an independence from certain values of a person, there are some beliefs or preferences in the mind of the respective person that do account for the behavior. The meme is always dependent on beliefs or preferences of the person. That humans are irrational is no new insight that can be seen and explained only with the help of a Darwinian analogy from genes to memes. The defense strategy that human minds are built by these memes fails, since not every part of mind is itself a meme, and since a specific meme does not build its own selective environment. Thus there cannot be a selfish meme that can spread irrespective of its selective environment.

Given that there is no independence of meme diffusion from human individuals, the explanatory units of selection analogy ends up in an explanatory dilemma: *Either* the analogy is *heuristically trivial*, because it loses its main claim, namely that memetics presents an alternative to the traditional explanation, which is given in terms of properties and interests of humans, *or* the explanatory units of selection analogy is *trivial in explanatory terms*, because it is tautological – it does not explain anything, since it merely states that those memes that have a high actual survival are those memes that have a high propensity for survival, without explaining where this high fitness emerges from.

# EPILOGUE

## Descriptive and explanatory force of the analogies

Darwin's theory has proved to be a very successful theory in explaining biological evolution. Darwinian analogical reasoning from this natural domain to other scientific domains has proved to have its merits as well. It played an important heuristic role in immunology and computer programming, leading to powerful theories and applications in these domains, as briefly mentioned in section 1.1. However, the same does not hold for memetics and the Darwinian approach to creativity. The three basic analogies, on which these Darwinian approaches to culture are based on, are deficient in descriptive as well as explanatory terms. The three basic analogies are either wrong or trivial. All things considered, these Darwinian approaches to not provide a fourth Freudian insult for mankind, a 'second Darwinian revolution,' by which not only God's creating power, but also our authorship of culture is demystified as illusory. Although we do not have a God-like creativity, we are nonetheless the authors of culture, creating, transmitting, and selecting ideational units that are not replicators in a narrow sense and have no existence and causal force that is independent from their authors.

In particular, I have illustrated that memes – the ideational units of culture – have a dubious ontological status, that they are not easy to observe if taken as generative units of culture, and that they do not replicate in a narrow sense. The ontological analogy, claiming that memes are gene-like replicators fails. With respect to the origination analogy, I looked at creativity as one source of novelty in culture. I claimed that creativity is not based on undirected variation and that this provides a major argument against a close analogy between origination of novelty in culture and origination of novelty in nature. This holds even if culture itself is a variational system and even if creativity comprises a cognitive selection process. With respect to the units of selection analogy, I explained in which sense memes fail to be 'selfish' replicators whose fitness is independent of their human carriers. Since humans are necessarily the selective environment of these memes, these memes cannot spread independently of their human carriers. Without taking the relation between memes and their selective environment into account, the claim that diffusion can be explained by the differential survival-of-the-fittest-memes becomes tautological. A transfer of the units of selection debate to culture does not make sense, since the role that individuals play in culture in relation to memes is not comparable to the role organisms play in biological evolution in relation to genes. Since humans create, transmit, and select memes, they are the primary causal agents of cultural change. Memes are necessarily a mere consequence of their causal influence, even though memes keep the book of cultural change, since they are the units that are selected by humans. All three basic analogies lack descriptive adequacy. Central claims of the analogies were thus shown to be wrong.

Nonetheless, I also showed that there are other statements about creativity and culture that are part of the ontological, the origination, and the explanatory units of selection analogy and that are indeed correct. But for these statements, all three analogies rely on insights from other research schools, to which the Darwinian approaches do not contribute new insights. The analogies are thus trivial – in heuristic as well as explanatory terms. With respect to the ontological analogy, I admitted that culture relies on social transmission of ideational units. Social transmission can be regarded as replication in the wide sense: Ideational units are transmitted over time. However, this is no new insight; on the contrary, it is a mere restatement of the contemporary anthropological concept of culture, which assumes social transmission of ideational units independently of a Darwinian point of view. Thus, if the ontological analogy is taken to refer to replication in the wide sense, the analogy is heuristically trivial, a mere reformulation of what is visible from any perspective on culture that assumes the contemporary concept of culture. In addition, if one wants to explain social transmission, one has to acknowledge all the findings of the psychology of social learning. This research does not need an analogy between genes and memes, and the analogy does not provide new insights for this research. The ontological analogy thus does not contribute to an explanation of social learning. On the contrary, in order to give any explanation of the transmission of memes, memetics has to use all the insights carefully investigated by the social sciences and philosophy. I have also shown that the same holds for the origination analogy. The analogy relies and reinvents the psychology of creativity, which assumes a concept of creativity that excludes that we have foresight of which of our ideas prove worthwhile. In addition, psychology of creativity tries to explain creativity by pointing to basic cognitive processes. If the analogy is taken as pointing to a close analogy between creativity and biological evolution, demanding that creativity is based on undirected variation, and if the analogy is offered as a claim about a cognitive process involved in creativity, then the claim is unjustified so far. It is unjustified, since the evidence cited for such a blind Darwinian cognitive process is not convincing. If, however, the analogy is understood as not requiring undirected variation, as merely pointing to guesswork, and interpreted as a mere perspective about creativity, then the analogy is trivial in descriptive as well as explanatory terms. The analogy is trivial, since it mere assumes the narrow concept of creativity that does not need the Darwinian approach, and since it does not provide an explanation for the processes involved. This version of the analogy merely re-describes the findings of the research in psychology of creativity in Darwinian terms, not offering any new insights that demand an analogy between creativity and biological evolution, since the assumed concept of creativity and the findings are independent of the analogy. The origination analogy is thus trivial in heuristic as well as explanatory terms. Last but not least, the explanatory units of selection analogy is tautological and in this sense trivial in terms of its explanatory force, as long as it does not refer to the traditional explanation of cultural diffusion, which states that diffusion is explained by the decisions of humans. If it does, however, offer a non-tautological explanation, the explanatory units of selection analogy ends up with that traditional explanation. It can thus not provide an alternative to this standard social-science explanation of diffusion and does not provide any new insights for the study of diffusion.

## A sheep in wolf's clothing

All things considered, Darwin's theory is not a "universal acid" – a "dangerous idea" or a "wolf in sheep's clothing," as Daniel Dennett (1995: 521) called it, an idea that provides frightening and striking new insights for studying creativity and culture. On the contrary, it is a 'sheep in wolf's clothing,' so to speak, that pretends to be dangerous, whereas, if looked at closely, it turns out not to be dangerous but trivial. The individual only disappears in the Darwinian approaches to creativity and culture because the causal role of the individual is divided up into *many* roles, which appear disconnected to each other on different places of the overall evolutionary picture of culture, leading to the impression that evolutionary theory destroys our image of ourselves as the primary agents of culture, as it destroyed the image of God as the creator of the world. The human being is then described as a mere 'vehicle' of memes in which these memes somehow arise when humans are creative and which then invade the minds of others. Although memes certainly exist, I tried to show that nobody has to be afraid of them.

The idea of 'selfish memes' is currently very popular. The reasons for this may be various. I think that one reason lies in a new trend towards anti-individualism. The modern Western world is characterized by radical freedom of choice. But for some people this freedom seems to be too much: The more options some people have, the more confused and anxious they get: One could make the wrong choices. One can have too many options and having options also means that one is responsible for choosing one of the options. Those religions that claim that our fate is predestinated have always served as a relief from the burden of choice and responsibility. Genes and memes can do the same. Once it was the devil, now it is the gene or the meme that made you do it. And – as in a curious reiteration – memeticists could even say that they are not responsible for all the analogical reasoning, since they are themselves victims of a selfish replicator, namely victims of the idea of memes. Analogical reasoning can be a fruitful source of insight, but we should be careful in using it, otherwise it would indeed be a 'dangerous idea' – preventing a very important habit of mind: critical thinking.

## The Wittgensteinian ladder

At the end of chapter 3, I conceded that there is one positive role Darwinian analogical reasoning can play. It can serve as a link between diverse disciplines. It can connect such different disciplines as philosophy, psychology, anthropology, economical diffusion studies, and evolutionary theory. It can bring these disciplines together – to study mankind in its totality. Memetics and the Darwinian approach to culture can therefore provide a general starting-point for an interdisciplinary perspective on culture and creativity. Nonetheless, the three basic analogies, which lie at the foundation of these two Darwinian approaches to culture, are still mere Wittgensteinian ladders. We can and should throw them away as soon as we come to the details.

# REFERENCE LIST

Adler, J. E. (1998). Rationality of Belief. In E. Craig (ed.). *Routledge Encyclopedia of Philosophy, Version 1.0*, (pp. online). London and New York: Routledge.

Allen, C., M. Bekoff & G. Lauder (ed.) (1998). *Nature's Purposes*: *Analyses of Function and Design in Biology*. Cambridge, MA: MIT Press.

Amundson, R. A. (1989). The trials and tribulations of selectionist explanations. In K. Hahlweg & Hooker, C. (ed.). *Issues in Evolutionary Epistemology*, (pp. 413-432). Albany: SUNY Press.

Ariew, A., R. Cummins & M. Perlman (ed.) (2002). *Functions*: *New Essays in the Philosophy of Biology*. Oxford: Oxford University Press.

Ariew, A. & R. C. Lewontin (2004). The confusions of fitness. *British Journal of Philosophy of Science 55*, 347-363.

Atran, S. (2001). The trouble with memes: Inferences versus imitation in cultural creation. *Human Nature 12*, 351-381.

Aunger, R. (ed.) (2000). *Darwinizing Culture*: *The Status of Memetics as a Science*. Oxford: Oxford University Press.

Aunger, R. (2002). *The Electric Meme*: *A New Theory of How We Think*. New York: Free Press.

Avital, E. & E. Jablonka (2000). *Animal Traditions*: *Behavioural Inheritance in Evolution*. Cambridge: Cambridge University Press.

Ayala, F. J. (1970). Teleological explanations in evolutionary biology. *Philosophy of science 37*, 1-15.

Ayala, F. J. (1977). Natural and artificial teleology. In T. Dobzhansky, Ayala, F.J., Stebbins, G.L. & Valentine, J. W. (ed.). *Evolution*, (pp. 500-504). San Francisco: Freeman.

Ayala, F. J. (2001). Chance and necessity: Adaptation and novelty in evolution. In J. B. Miller (ed.). *An Evolving Dialogue: Theological and Scientific Perspectives on Evolution*, (pp. 231-261). Harrisburg, Pennsylvania: Trinity Press.

Balkin, J. M. (1998). *Cultural Software*: *A Theory of Ideology*. New Haven: Yale University Press.

Barkow, J. H., L. Cosmides & J. Tooby (ed.) (1992). *The Adapted Mind*: *Evolutionary Psychology and the Generation of Culture*. Oxford: Oxford University Press.

Barth, F. (2002). Toward a richer description and analysis of cultural phenomena. In R. G. Fox & King, B. J. (ed.). *Anthropology Beyond Culture*, (pp. 23-36). Oxford: Berg.

Basalla, G. (1988). *The Evolution of Technology*. Cambridge: Cambridge University Press.

Bateson, P. (1978). Book review of 'The Selfish Gene'. *Animal Behaviour 26*, 316-318.

Beatty, J. (1992). Fitness: Theoretical contexts. In E. F. Keller & Lloyd, E. A. (ed.). *Keywords in Evolutionary Biology*, (pp. 115-119). Cambridge, MA: Harvard University Press.

Behe, M. (1996). *Darwin's Black Box*: *The Biochemical Challenge to Evolution*. New York: Free Press.

Benzon, W. (1996). Culture as an evolutionary arena. *Journal of Social and Evolutionary Systems 19*, 321-362.

Bethell, T. (1976). Darwin's Mistake. *Harper's Magazine*, 70-75.

Beurton, P. J., R. Falk & H.-J. Rheinberger (ed.) (2000). *The Concept of the Gene in Development and Evolution*: *Historical and Epistemological Perspectives*. Cambridge: Cambridge University Press.

Black, M. (1962). *Models and Metaphors*: *Studies in Language and Philosophy*. Ithaca: Cornell University Press.

Blackmore, S. (1999). *The Meme Machine*. Oxford: Oxford University Press.

Blackmore, S. (2000). The power of memes. *Scientific American 283*, 64-73.

Blackmore, S. (2002). Meme. In M. Pagel (ed.). *Encyclopedia of Evolution*, (Vol. 2, pp. 713-716). Oxford: Oxford University Press.

Bloch, M. (2000). A well-disposed social anthropologist's problems with memes. In R. Aunger (ed.). *Darwinizing Culture: The Status of Memetics as a Science*, (pp. 189-203). Oxford: Oxford University Press.

Blute, M. (1979). Learning theory and the evolutionary analogy. *Cogprints http://cogprints.ecs.soton.ac.uk*.

Blute, M. (2005). Memetics and evolutionary social sciences. *Journal of Memetics 6*, http://jom-emit.cfpm.org/2005/vol9/blute_m.html.

Boas, F. (1911). *The Mind of Primitive Man*. New York: Macmillan.

Boden, M. A. (1996). The constraints of knowledge. In A. E. Andersson & Sahlin, N. (ed.). *The Complexity of Creativity*, (pp. 1-4). Dordrecht: Kluwer.

Boden, M. A. (1999). Computer models of creativity. In R. J. Sternberg (ed.). *Handbook of Creativity*, (pp. 351-391). Cambridge: Cambridge University Press.

Boden, M. A. (2004). *The Creative Mind*: *Myths and Mechanism*. 2. ed. London and New York: Routledge.

Borofsky, R. (1987). *Making History*: *Pukapukan and Anthropological Constructions of Knowledge*. Cambridge: Cambridge University Press.

Bowler, P. J. (1988). *The Non-Darwinian Revolution*: *Reinterpreting a Historical Myth*. Baltimore: John Hopkins University Press.

Bowler, P. J. (2003). *Evolution*: *The History of an Idea*. 2. ed. Berkeley: University of California Press.

Boyd, R. & P. J. Richerson (1985). *Culture and the Evolutionary Process*. Chicago: University of Chicago Press.

Boyd, R. & P. J. Richerson (2000). Memes: Universal acid or a better mousetrap?. In R. Aunger (ed.). *Darwinizing Culture: The Status of Memetics as a Science*, (pp. 143-162). Oxford: Oxford University Press.

Bradie, M. (1994). Epistemology from an evolutionary point of view. In E. Sober (ed.). *Conceptual Issues in Evolutionary Biology*, (pp. 453-476). Cambridge, MA: MIT Press.

Bradie, M. (1998). Models and metaphors in science: The metaphorical turn. *ProtoSociology 12*.

Brandon, R. N. (1982). The levels of selection. *PSA - Proceedings of the Philosophy of Science Association 1*, 315-322. (repr. in Brandon & Burian 1984: 133-141).

Brandon, R. N. & R. M. Burian (ed.) (1984). *Genes, Organisms, Populations*: *Controversies over the Units of Selection*. Cambridge, MA: MIT Press.

Briskman, L. (1981). Creative Product and creative process in science and art. In D. Dutton & Krausz, M. (ed.). *The concept of creativity in science and art*, (pp. 129-155). Dordrecht: Martinus Nijhoff Publ.

Brodie, R. (1995). *Virus of the Mind*: *The New Science of the Meme*. Seattle: Integral Press.

Buller, D. J. (ed.) (1999). *Function, Selection and Design*. Albany, NY: State University of New York Press.

Burckhardt, R. J. (1995). *The Spirit of System*: *Lamarck and Evolutionary Biology*. 2. ed. Cambridge, MA: Harvard University Press.

Burnet, F. M. S. (1957). A modification of Jerne's theory of antibody production: Using the concept of clonal selection. *Australian Journal of Science 20*, 67-69.

Burt, A. (2002). Selfish Gene. In M. Pagel (ed.). *Encyclopedia of Evolution*, (Vol. 2, pp. 1018-1020). Oxford: Oxford University Press.

Buskes, C. (1998). *The Genealogy of Knowledge*: *A Darwinian Approach to Epistemology and Philosophy of Science*. Tilburg: Tilburg University Press.

Buss, D. M. (2004). *Evolutionary Psychology*: *The New Science of the Mind*. 2. ed. Boston: Pearson.

Campbell, D. T. (1965). Variation and selective retention in socio-cultural evolution. In H. Barringer, Blanksten, G. & Mack, R. (ed.). *Social Change in Developing Areas: A Reinterpretation of Evolutionary Theory*, (pp. 19-49). Cambridge, MA: Schenkman.

Campbell, D. T. (1974b). Unjustified variation and selective retention in scientific discovery. In F. Ayala & Dobszhansky, T. (ed.). *Studies in the Philosophy of Biology: Reduction and Related Problems*, (pp. 139-161). London: Macmillan.

Campbell, D. T. (1987 [1960]). Blind variation and selective retention in creative thought as in other knowledge processes. In G. Radnitzky & Bartley, W. (ed.). *Evolutionary Epistemology: Rationality and the Sociology of Knowledge*, (Vol. 67, pp. 91-114). LaSalle: Open Court. (org. in 1960. The Psychological Review 67, 380-400).

Campbell, D. T. (1987 [1974a]). Evolutionary epistemology. In G. Radnitzky & Bartley, W. (ed.). *Evolutionary Epistemology, Rationality, and the Sociology of Knowledge*, (Vol. 1, pp. 47-89). La Salle: Open Court Publ. (org. in 1974. P. A. Schilpp (ed.). The Philosophy of Karl Popper, (Vol. 1, pp. 413-463). La Salle: Open Court Publ).

Campbell, D. T. (1997). From evolutionary epistemology via selection theory to a sociology of scientific validity. *Evolution and Cognition 3*, 5-38. (posthum, ed. by C. Heyes and B. Frankel).

Cannon, W. B. (1940). The role of chance in discovery. *Scientific Monthly 50*, 204-209.

Cannon, H. (1957). What Lamarck really said. *Proceedings of the Linnean Society of London 168*, 70-87.

Carneiro, R. L. (2003). *Evolutionism in Cultural Anthropology*: *A Critical History*. Boulder: Westview.

Cavalli-Sforza, L. L. & M. Feldman (1981). *Cultural Transmission and Evolution*: *A Quantitative Approach*. Princeton: Princeton University Press.

Cavalli-Sforza, L. L. (2000). *Genes, Peoples, and Languages*. New York: North Point Press. (transl. by M. Seielstad, orig. 1996. Gènes, peuples et langues. Paris: Jacob).

Charlesworth, B. (2002). Fitness. In M. Pagel (ed.). *Encyclopedia of Evolution,* (Vol. 1, pp. 359-362). Oxford: Oxford University Press.

Chomsky, N. (1966). *Cartesian Linguistics*: *A Chapter in the History of Rationalist Thought*. New York: Harper & Row.

Clark, S. R. L. (1993). Minds, Memes, and Rhetoric. *Inquiry 36*, 3-16.

Clark, A. (2002). That special something: Dennett on the making of minds and selves. In A. Brook & Ross, D. (ed.). *Daniel Dennett*, (pp. 187-205). Cambridge: Cambridge University Press.

Conte, R. (2000). Memes through (social) minds. In R. Aunger (ed.). *Darwinizing Culture*, (pp. 83-119). Oxford: Oxford University Press.

Cosmides, L. & J. Tooby (1992). The Psychological foundations of culture. In J. H. Barkow, Cosmides, L. & Tooby, J. (ed.). *The Adapted Mind: Evolutionary*

*Psychology and the Generation of Culture*, (pp. 19-136). Oxford: Oxford University Press.

Crick, F. (1958). On protein synthesis. *Symposia of the Society for Experimental Biology 12*, 138-163.

Cronk, L. (1991). Human behavioural ecology. *Annual Review of Anthropology 20*, 25-53.

Csikszentmihalyi, M. (1996). *Creativity*: *Flow and the Psychology of Discovery and Invention*. New York: Harper Collins.

Cummins, R. (1975). Functional Analysis. *Journal of Philosophy 72*, 741-65.

Cziko, G. A. (1995). *Without Miracles*: *Universal Selection Theory and the Second Darwinian Revolution*. Cambridge, MA: MIT Press.

Cziko, G. A. (1998). From blind to creative: In defense of Donald Campbell's selectionist theory of human creativity. *Journal of Creative Behavior 32*, 192-209.

Darwin, C. (1859). *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*. 1. ed. London: Murray.

Darwin, C. (1862). *On the Various Contrivances by which British and Foreign Orchids are Feritlized by Insects, and the Good Effects of Intercrossing*. London: Freeman.

Darwin, C. (1868). *The Variation of Plants and Animals under Domestication, 2 vols*. 1. ed. New York: Orange Judd & Co.

Darwin, C. (1875). *The Variation of Plants and Animals under Domestication, 2 vols*. 2. ed. New York: Orange Judd & Co.

Darwin, C. (1876). *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*. 6. ed. London: Murray.

Davis, L. (1991). *The Handbook of Genetic Algorithms*. New York: Van Nostrand Reinhold.

Dawkins, R. C. (1976). *The Selfish Gene*. 1. ed. Oxford: Oxford University Press.

Dawkins, R. (1978). Replicator selection and the extended phenotype. *Zeitschrift für Tierpsychologie 47*, 61-76. (repr. in Sober, Elliott (1984). Conceptual Issues in Evolutionary Biology: An Anthology, (pp. 125-141), Cambridge, MA: MIT Press. ).

Dawkins, R. C. (1982a). *The Extended Phenotype*: *The Gene as the Unit of Selection*. Oxford: Oxford University Press.

Dawkins, R. (1982b). Replicators and vehicles. In King's College Sociobiology Group (ed.). *Current Problems in Sociobiology*, (pp. 45-64). Cambridge: Cambridge University Press. (repr. in Brandon/ Burian 1984: 161-180. ).

Dawkins, R. (1983). Universal Darwinism. In D. S. Bendall (ed.). *Evolution from Molecules to Man*, (pp. 403-425). Cambridge: Cambridge University Press.

Dawkins, R. (1986). *The Blind Watchmaker*. London: Penguin.

Dawkins, R. C. (1989 [1976]). *The Selfish Gene*. 2. ed. Oxford: Oxford University Press.

Dawkins, R. (1993). Viruses of the mind. In B. Dahlbom (ed.). *Dennett and his Critics*, (pp. 13-27). Cambridge: Blackwell.

Dawkins, R. (1999). Foreword to Blackmore. In S. Blackmore (ed.). *The Meme Machine*, (pp. vii-xvii). Oxford: Oxford University Press.

Dawkins, R. (2003). *A Devil's Chaplain*: *Reflections on Hope, Lies, Science, and Love*. Boston, New York: Houghton Mifflin Company.

Dawkins, R. (2004). Extended phenotype, but not too extended: A reply to Laland, Turner and Jablonka. *Biology and Philosophy 19*, 377-396.

Deacon, T. (1999). Memes as signs: The trouble with memes (and what to do about it). *The Semiotic Review of Books 10*, 1-3.

Delius, J. D. (1991). The Nature of Culture. In M. S. Dawkins, Halliday, T. R. & Dawkins, R. (ed.). *The Tinbergen Legacy*, (pp. 75-99). London: Chapman & Hall.

Dembski, W. A. (1998). *The Design Inference*: *Eliminating Chance through Small Probabilities*. Cambridge: Cambridge University Press.

Dennett, D. C. (1975). Why the law of effect will not go away. *Journal for the Theory of Social Behavior 5*, 169-187. (repr. in Dennett, D. C. (1978) Brainstorms (pp. 70-89). Cambridge, MA: MIT Press. ).

Dennett, D. C. (1987). *The Intentional Stance*. Cambridge, MA: MIT Press.

Dennett, D. C. (1990). Memes and the exploitation of imagination. *The Journal of Aesthetics and Art Criticism 48*, 127-135.

Dennett, D. C. (1991). *Consciousness Explained*. Boston: Little, Brown & Co.

Dennett, D. C. (1995). *Darwin's Dangerous Idea*: *Evolution and the Meanings of Life*. New York: Simon and Schuster.

Dennett, D. C. (1996). *Kinds of Minds*. New York: Basic Books.

Dennett, D. C. (2000). Foreword to 'Darwinizing Culture'. In R. Aunger (ed.). *Darwinizing Culture: The Status of Memetics as a Science*, (pp. v-ix). Oxford: Oxford University Press.

Dennett, D. C. (2001a). The Evolution of Culture. *Monist 84*, 305-324.

Dennett, D. C. (2001b). The Evolution of Evaluators. In A. Nicita & Pagano, U. (ed.). *The Evolution of Economic Diversity*, (pp. 66-81). London: Routledge.

Dennett, D. C. (2002). The New Replicators. In M. Pagel (ed.). *Encyclopedia of Evolution*, (Vol. 1, pp. E-83 - E92). Oxford: Oxford University Press.

Dennett, D. C. (2003). *Freedom Evolves*. New York: Viking.

Dennett, D. C. (2004). Could there be a Darwinian account of human creativity? In A. Moya & Font, E. (ed.). *Evolution: From Molecules to Ecosystems*, (pp. 273-279). Oxford: Oxford University Press.

Dennett, D. C. (forthcoming). From typo to thinko: When evolution graduated to semantic norms. In S. Levinson & Jaisson, P. (ed.). *Culture and evolution*. Cambridge, MA: MIT Press.

Diaz de Chumaceiro, C. L. (1999). Serendipity. In M. A. Runco & Pritzker, S. R. (ed.). *Encyclopedia of Creativity*, (Vol. 2, pp. 543-549). San Diego, CA: Academic Press.

Dobzhansky, T. (1973). Nothing in biology makes sense except in the light of evolution. *American Biology Teacher 35*, 125-129.

Dobzhansky, T. (1974). Chance and creativity in evolution. In F. Ayala & Dobszhansky, T. (ed.). *Studies in the Philosophy of Biology: Reduction and Related Problems*, (pp. 307-338). London: Macmillan.

Durham, W. H. (1991). *Coevolution*: *Genes, Culture, and Human Diversity*. Stanford: Stanford University Press.

Durham, W. H. (2002). Cultural variation in time and space: The case for a populational theory of culture. In R. G. Fox & King, B. J. (ed.). *Anthropology Beyond Culture*, (pp. 193-206). Oxford: Berg.

Elster, J. (1979). *Ulysses and the Sirens*: *Studies in Rationality and Irrationality*. New York: Columbia University Press.

Elster, J. (1989). *Nuts and Bolts for the Social Sciences*. Cambridge: Cambridge University Press.

Elton, M. (2003). *Dennett*: *Reconciling Science and Our Self-Conception*. Oxford: Blackwell.

Findlay, A. (1968 [1965]). *A Hundred Years of Chemistry*. 3. rev. ed. London: Duckworth.

Finke, R. A., T. B. Ward & S. M. Smith (1992). *Creative Cognition*: *Theory, Research and Applications*. Cambridge, MA: MIT Press.

Fisher, R. A. (1930). *The Genetical Theory of Natural Selection*. Oxford: Clarendon.

Fogel, L. J., A. J. Owens & M. J. Walsh (1966). *Artificial Intelligence Through Simulated Evolution*. New York: Wiley.

Fogel, L. J. (1999). *Intelligence Through Simulated Evolution*: *Forty Years of Evolutionary Programming*. New York: Wiley.

Fox, R. G. & B. J. King (ed.) (2002). *Anthropology Beyond Culture*. Oxford: Berg.

Fracchia, J. & R. C. Lewontin (1999). Does culture evolve. *History and Theory 38*, 52-78.

Futuyma, D. J. (1998). *Evolutionary Biology*. 3. ed. Sunderland, MA: Sinauer Associates.

Gamble, T. J. (1983). The natural selection model of knowledge generation: Campbell's dictum and its critics. *Cognition and Brain Theory 6*, 353-363.

Gatherer, D. (1998). Why the thought contagion metaphor is retarding the progress of memetics. *Journal of Memetics 2*, http://jom-emit.cfpm.org/1998/vol2/gatherer_d.html.

Gentner, D. & M. Jeziorski (1993). The shift from metaphor to analogy in Western science. In A. Ortony (ed.). *Metaphor and Thought*, (pp. 447-480). Cambridge: Cambridge University Press.

Gil-White, F. J. (forthcoming). Common misunderstandings of memes (and genes): The promise and the limits of the genetic analogy to cultural transmission processes. *http://www.psych.upenn.edu/~fjgil/.*

Gillispie, C. C. (1956). The formation of Lamarck's evolutionary theory. *Archives Internationales d'Histoire des Sciences 9*, 323-338.

Godfrey-Smith, P. (2000). The replicator in retrospect. *Biology and Philosophy 15*, 403-423.

Godfrey-Smith, P. (2001). Three kinds of adaptationism. In S. Orzack & Sober, E. (ed.). *Adaptationism and Optimality*, (pp. 335-357). Cambridge: Cambridge University Press.

Goldberg, D. E. (1989). *Genetic Algorithms in Search, Optimization, and Machine Learning*. Reading: Addison-Wesley.

Gould, S. J. (1979). Shades of Lamarck. *Natural History 88*, 22-28.

Gould, S. J. (1982). *The Panda's Thumb: More Reflections in Natural History*. New York: Norton.

Gould, S. J. (1991). *Bully for Brontosaurus*: *Reflections in Natural History*. New York: Norton.

Gould, S. J. (1997a). Darwinian fundamentalism. *New York Review of Books 44*, 34-37.

Gould, S. J. (1997b). Evolution: The pleasures of pluralism. *New York Review of Books 44* , 47-52.

Gould, S. J. (1998 [1976]). Darwin's Untimely Burial. In M. Ruse (ed.). *Philosphy of Biology*, (pp. 93-98). Amherst, NY: Prometheus Books. (org. in 1976. Natural History 85, 24-32).

Gould, S. J. (2001). The evolutionary definition of selective agency: Validation of the theory of hierarchical selection, and the fallacy of the selfish gene. In R. Sing, Krimbas, C., Paul, D. & Beatty, J. (ed.). *Thinking about Evolution: Historical, Philosophical, and Political Perspectives*, (Vol. 2, pp. 208-234). Cambridge: Cambridge University Press.

Griffiths, P. & R. Gray (1994). Developmental systems and evolutionary explanations. *Journal of Philosophy 91*, 277-304.

Griffiths, P. & R. Gray (1997). Replicator II: Judgment day. *Biology and Philosophy 12*, 471-492.

Griffiths, P. (2001). Genetic information: A metaphor in search of a theory. *Philosophy of Science 68*, 394-412.

Guilford, J. P. (1967). *The Nature of Human Intelligence*. New York: McGraw-Hill.

Hadamard, J. (1954 [1945]). *The Psychology of Invention in the Mathematical Field.* New York: Dover.

Hampton, J. (1998). Practical Rationality. In E. Craig (ed.). *Routledge Encyclopedia of Philosophy, Version 1.0*, (pp. online). London and New York: Routledge.

Harms, W. (1996). Cultural evolution and the variable phenotype. *Biology and Philosophy 11*, 357-375.

Hausman, C. R. (1984). *A Discourse on Novelty and Creation*. Albany: State University of New York Press.

Hausman, C. R. (1993). *Charles S. Peirce's Evolutionary Philosophy*. Cambridge: Cambridge University Press.

Henrich, J., R. Boyd & P. J. Richerson (forthcoming). Five misunderstandings about cultural evolution. In D. Sperber (ed.). *Epidemiology of Ideas*, Open Court Publishing Company.

Hesse, M. (1963). *Models and Analogies in Science*. London: Sheed and Ward.

Heyes, C. M. & H. C. Plotkin (1989). Replicators and interactors in cultural evolution. In M. Ruse (ed.). *What the Philosophy of Biology is*, (pp. 139-162). Dordrecht: Kluwer.

Heyes, C. M. (1994). Social learning in animals: Categories and mechanisms. *Biological Reviews 69*, 207-231.

Heyes, C. M. & B. G. Galef (ed.) (1996). *Social Learning in Animals*: *The Roots of Culture*. San Diego: Academic Press.

Holland, J. H. (1975). *Adaptation in Natural and Artificial Systems*. Ann Arbor, MI: University of Michigan Press.

Hull, D. L. (1980). Individuality and selection. *Annual Review of Ecology and Systematics 11*, 311-332.

Hull, D. L. (1982). The naked meme. In H. C. Plotkin (ed.). *Learning, Development, and Culture*, (pp. 273-327). Chichester: Wiley.

Hull, D. L. (1988a). *Science as a Process*: *An Evolutionary Account of the Social and Conceptual Development of Science*. Chicago: Chicago University Press.

Hull, D. L. (1988b). Interactors versus vehicles. In H. Plotkin (ed.). *The Role of Behavior in Evolution*, (pp. 19-50). Cambridge, MA: MIT Press.

Hull, D. L. (1999). The use and abuse of Sir Karl Popper. *Biology and Philosophy 14*, 481-504.

Hull, D. L. (2000). Taking memetics seriously: Memetics will be what we make it. In R. Aunger (ed.). *Darwinizing Culture: The Status of Memetics as a Science*, (pp. 43-67). Oxford: Oxford University Press.

Hull, D. L., S. S. Glenn & R. E. Langman (2001). A general account of selection: biology, immunology and behavior. *Behavioral and Brain Sciences 24*, 511-528.

Hull, D. L. (2001). *Science and Selection*: *Essays on Biological Evolution and the Philosophy of Science*. Cambridge: Cambridge University Press.

Hull, D. L. & J. Wilkins (2005). Replication. *Stanford Encyclopedia of Philosophy (Spring 2005 Edition)* (http://plato.stanford.edu/archives/spr2005/entries/replication/).

Ingold, T. (1994). Introduction to Culture. In T. Ingold (ed.). *Companion Encyclopedia of Anthropology: Humanity, Culture and Social Life*, (pp. 329-349). London: Routledge.

Jablonka, E. & M. J. Lamb (1995). *Epigenetic Inheritance and Evolution*: *The Lamarckian Dimension*. Oxford: Oxford University Press.

Jablonka, E. (2004). From replicators to heritably varying phenotypic traits: The extended phenotype revisited. *Biology and Philosophy 19*, 353-375.

Jacob, F. (1982). *The Possible and the Actual*. Seattle: University of Washington Press.

James, W. (1979 [1880]). Great Men and Their Environment. In F. H. Burkhardt, Bowers, F. & Skrupskelis, I. (ed.). *The Works of William James, Bd. 6: The Will to Believe and Other Essays in Popular Philosophy*, (pp. 163-189). Cambridge, MA: Harvard University Press.

Jeffreys, M. (2000). The meme metaphor. *Perspectives in Biology and Medicine 43*, 227-42.

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*, 849-957.

Jerne, N. K. (1967). Antibodies and learning: Selection versus instruction. In G. Quarton, Melnechuk, T. & Schmitt, F. (ed.). *The Neurosciences: A Study Program*, (pp. 200-205). New York: Rockefeller University Press.

Jerne, N. K. (1985). The generative grammar of the immune system. *Science 229*, 1057-09.

Kahneman, D., P. Slovic & A. Tversky (ed.) (1982). *Judgment Under Uncertainty*: *Heuristics and Biases*. Cambridge: Cambridge University Press.

Katz, M. L. & H. S. Rosen (1998). *Microeconomics*. Boston: Irwin/ McGraw-Hill.

Keesing, R. M. (1974). Theories of culture. *Annual Review of Anthropology 3*, 73-97.

Keller, L. (ed.) (1999). *Levels of Selection in Evolution*. Princeton, NJ: Princeton University Press.

Kettlewell, B. (1973). *The Evolution of Melanism*: *The Study of Recurring Necessity with special Reference to Industrial Melanism in the Lepidoptera*. Oxford: Oxford University Press.

Kim, J. (1978). Supervenience and nomological incommensurables. *American Philosophical Quarterly 15*, 149-156.

Kimura, M. (1983). *The Neutral Theory of Molecular Evolution*. Cambridge: Cambridge University Press.

Kimura, M. (1992). Neutralism. In E. F. Keller & Lloyd, E. A. (ed.). *Keywords in Evolutionary Biology*, (pp. 225-230). Cambridge, MA: Harvard University Press.

Kitcher, P. (1992). Gene: Current usages. In E. F. Keller & Lloyd, E. A. (ed.). *Keywords in Evolutionary Biology*, (pp. 128-131). Cambridge, MA: Harvard University Press.

Kitcher, P. (1998). Explanation. In E. Craig (ed.). *Routledge Encyclopedia of Philosophy, Version 1.0.*, (pp. online). London: Routledge.

Koestler, A. (1964). *The Act of Creation*. London: Hutchinson.

Koza, J. R. (1992). *Genetic Programming*: *On the Programming of Computers by Means of Natural Selection*. Cambridge, MA: MIT Press.

Koza, J. R., F. H. Bennett, D. Andre & M. A. Keene (1999). *Genetic Programming III*: *Darwinian Invention and Problem Solving*. San Francisco: Kaufman.

Krebs, J. R. & N. B. Davies (ed.) (1997 [1978]). *Behavioural Ecology*: *An Evolutionary Approach*. Oxford: Blackwell.

Kroeber, A. L. & T. Parsons (1958). The concept of culture and of social system. *American Sociological Review 23*, 582-583.

Kroeber, A. L. & C. K. Kluckhohn (1963 [1952]). *Culture: A Critical Review of Concepts and Definitions*. New York: Vintage Books. (org. in 1952. Papers of the Peabody Museum 47, 1-223).

Kronfeldner, M. E. (2001, manuscript) *Zur Kreativität des Denkens*: *Einem Begriff auf der Spur*. Universität Regensburg (Magisterarbeit in Philosophie, Philosophische Fakultät 1).

Kronfeldner, M. E. (2005). Zum Begriff der psychologischen Kreativität. In G. Abel (ed.). *Kreativität: XX. Deutscher Kongress für Philosophie*, (pp. 19-30). Berlin: Universitätsverlag TU Berlin.

Kronfeldner, M. E. (forthcoming). Cultural evolution and Lamarck: Inheritance, patterns and processes. forthcoming in *Biology and Philosophy*.

Kuper, A. (1999). *Culture*: *The Anthropologists Account*. Cambridge, MA: Harvard University Press.

Kuper, A. (2000). If memes are the answer, what is the question? In R. Aunger (ed.). *Darwinizing Culture: The Status of Memetics as a Science*, (pp. 175-188). Oxford: Oxford University Press.

Kuper, A. (2002). Culture. *Proceedings ot the British Academy 112*, 87-112.

Kutschera, F. v. (1999). Rationality and Reason. In G. Brüntrup & Tacelli, R. (ed.). *The Rationality of Theism*, (pp. 255-269). Dordrecht: Kluwer.

Laland, K. N., J. Odling-Smee & M. W. Feldman (2000). Niche construction: biological evolution and cultural change. *Behavioral and Brain Sciences 23*, 131-175.

Laland, K. & J. Odling-Smee (2000). The evolution of the meme. In R. Aunger (ed.). *Darwinizing Culture*, (pp. 121-141). Oxford: Oxford University Press.

Laland, K. N. & G. R. Brown (2002). *Sense and Nonsense*: *Evolutionary Perspectives on Human Social Behavior*. Oxford: Oxford University Press.

Laland, K. N. (2004). Extending the extended phenotype. *Biology and Philosophy 19*, 313-325.

Lamarck, J. d. (1984 [1809]). *Zoological Philosophy*: *An Exposition with Regard to the Natural History of Animals*. Chicago: Chicago University Press. (trans. by Hugh Elliot, org. 1809. Philosophie zoologique. Bruxelles: Culture et Civilisation).

LeMahieu, D. (1976). *The Mind of William Paley*: *A Philosopher and His Age*. Lincoln: University of Nebraska Press.

Lennox, J. G. (1992). Teleology. In E. F. Keller & Lloyd, E. A. (ed.). *Keywords in Evolutionary Biology*, (pp. 324-333). Cambridge, MA: Harvard University Press.

Levi-Strauss, C. (1962). *La pensée sauvage*. Paris: Plon.

Lewens, T. (2004). *Organisms and Artifacts*: *Design in Nature and Elsewhere*. Cambridge, MA: MIT Press.

Lewis, W. S. (ed.) (1960). *Horace Walpole's correspondence (The Yale Edition, Vol. 20)*. New Haven: Yale University Press.

Lewontin, R. C. (1968). The concept of evolution. In D. Sills (ed.). *International Encyclopedia of the Social Sciences*, (Vol. 5, pp. 202-210). New York: Macmillan.

Lewontin, R. C. (1970). The units of selection. *Annual Review of Ecology and Systematics 1*, 1-18.

Lewontin, R. C. (1985). Evolution as theory and ideology. In R. Levins & Lewontin, R. C. (ed.). *The Dialectical Biologist*, (pp. 9-64). Cambridge, MA: Harvard University Press.

Lewontin, R. C. (1985 [1983]). The Organism as the Subject and Object of Evolution. In R. Levins & Lewontin, R. C. (ed.). *The Dialectical Biologist*, (pp. 85-106). Cambridge, MA: Harvard University Press. (org. in 1983. Scientia 118, 63-82).

Lewontin, R. C. (1991). *Biology as Ideology*. New York: Harper.

Lloyd, E. A. (1992). Units of Selection. In E. F. Keller & Lloyd, E. A. (ed.). *Keywords in Evolutionary Biology*, (pp. 334-340). Cambridge, MA: Harvard University Press.

Lloyd, E. A. (2001). Units and levels of selection: An anatomy of the units of selection debates. In R. S. Sing, Krimbas, C.B., Paul, D.E. & Beatty, J. (ed.). *Thinking About Evolution: Historical, Philosophical, and Political Perspectives*, (pp. 267-291). Cambridge: Cambridge University Press.

Lloyd, E. A. (forthcoming 2005). Why the gene will not return. *Philosophy of Science 72*.

Lorenz, K. (1941). Kants Lehre vom Apriorischen im Lichte gegenwärtiger Biologie. *Blätter für deutsche Philosophie 15*, 94-125.

Lovejoy, A. O. (1936). *The Great Chain of Being*: *A Study of the History of an Idea*. Cambridge, MA: Harvard University Press.

Luhmann, N. (1997). *Die Gesellschaft der Gesellschaft, 2. Bde.*. Frankfurt: Suhrkamp.

Lynch, A. (1996). *Thought Contagion*: *How Belief Spreads Through Society*. New York: Basic Books.

Mach, E. (1905). *Erkenntnis und Irrtum*: *Skizzen zur Psychologie der Forschung*. Barth: Leipzig. (engl. transl. (1976). Knowledge and Error: Sketches on the Psychology of Enquiry, transl. by Thomas J. McCormack and Paul Foulkes, ed. by Brian McGuiness, Dordrecht: Reidel).

Manser, A. (1965). The Concept of Evolution. *Philosophy 40*, 18-34.

Maynard-Smith, J. (1982). *Evolution and the Theory of Games*. Cambridge: Cambridge University Press.

Maynard-Smith, J. & E. Szathmary (1995a). *The Major Transitions in Evolution*. New York: Freeman.

Maynard-Smith, J. & E. Szathmary (1995b). The major evolutionary transitions. *Nature 374*, 227-32.

Maynard-Smith, J. (1995c). Genes, memes and minds: Review of Daniel C. Dennett's 'Darwin's Dangerous Idea'. *New York Review of Books 42*, 46-48.

Maynard-Smith, J. & E. Szathmary (1999). *The Origins of Life*: *From the Birth of Life to the Origin of Language*. Oxford: Oxford University Press.

Maynard-Smith, J. (2000). The concept of information in biology. *Philosophy of Science 67*, 177-194.

Mayr, E. (1959). Darwin and the evolutionary theory in biology. In B. J. Meggers (ed.). *Evolution and Anthropology: A Centennial Appraisal*, (pp. 1-10). Washington, DC: Anthropological Society of Washington.

Mayr, E. (1963). *Animal Species and Evolution*. Cambridge, MA: Harvard University Press.

Mayr, E. (1975). The Unity of the Genotype. *Biologisches Zentralblatt 94*, 377-388.

Mayr, E. (1976 [1962]). Accident or design: The paradox of evolution. In E. Mayr (ed.). *Evolution and the Diversity of Life: Selected Essays*, (pp. 30-43). Cambridge, MA: Harvard University Press. (org. in 1962. Proceedings of the Darwin Centenary Symposium of the Royal Society of Victoria: The Evolution of Living Organisms, 1-14.).

Mayr, E. (1976 [1972]). Lamarck revisited. In E. Mayr (ed.). *Evolution and the Diversity of Life: Selected Essays*, (pp. 222-250). Cambridge, MA: Harvard University Press. (org. in 1972. Journal of the History of Biology 5, 55-94).

Mayr, E. (1976 [1974]). Teleological and teleonomic: A new analysis. In E. Mayr (ed.). *Evolution and the Diversity of Life: Selected Essays*, (pp. 383-404). Cambridge, MA: Harvard University Press. (org. in 1974. Boston Studies in the Philosophy of Science 14, 91-117).

Mayr, E. (1982). *The Growth of Biological Thought*: *Diversity, Evolution, and Inheritance*. Cambridge, MA: Harvard University Press.

McLaughlin, P. (2001). *What Functions Explain*: *Functional Explanations and Self-Reproducing Systems*. Cambridge: Cambridge University Press.

Medawar, P. M. (1981 [1953]). A Commentary on Lamarckism. In P. Medawar (ed.). *The Uniqueness of the Individual*, (pp. 63-87). New York: Dover.

Medawar, P. (1982 [1977]). Unnatural Science. In P. Medawar (ed.). *Pluto's Republic*, (pp. 167-183). Oxford: Oxford University Press. (org. in 1977. New York Review of Books, 9 June.).

Mednick, S. A. (1962). The associative basis of the creative process. *Psychological Review 69*, 220-232.

Merton, R. K. (1967 [1949]). *Social Theory and Social Structure*. rev. and enl. ed. New York: Free Press.

Merton, R. K. & E. Barber (2004). *The Travels and Adventures of Serendipity*: *A Study in Historical Semantics and the Sociology of Science*. Princeton, NJ: Princeton University Press.

Mesoudi, A., A. Whiten & K. N. Laland (2004). Is human cultural evolution Darwinian? Evidence reviewed from the perspective of 'The Origin of Species'. *Evolution 58*, 1-11.

Midgley, M. (2002). Choosing the Selectors. *Proceedings of the British Academy 112*, 119-133.

Millikan, R. G. (2000). Reading Mother Nature's Mind. In D. Ross, Brook, A. & Thompson, D. (ed.). *Dennett's Philosophy: A Comprehensive Assessment*, (pp. 55-75). Cambridge, MA: MIT Press.

Mills, S. K. & J. H. Beatty (1994 [1979]). The propensity interpretation of fitness. In E. Sober (ed.). *Conceptual Issues in Evolutionary Biology*, (pp. 3-23). Cambridge, MA: MIT Press. (org. in 1979. Philosophy of Science 46, 263-286).

Mithen, S. (1996). *The Prehistory of the Mind*: *The Cognitive Origins of Art, Religion and Science*. London: Thames and Hudson.

Mithen, S. (ed.) (1998). *Creativity in Human Evolution and Prehistory*. London: Routledge.

Mumford, M. D. (1999). Blind variation or selective variation? Evaluative elements in creative thought. *Psychological Inquiry 10*, 344-348.

Nagel, E. (19777). *Teleology Revisited and Other Essays in the Philosophy and History of Science*. New York: Columbia University Press.

Nahm, M. C. (1956). *The Artist as a Creator*: *An Essay of Human Freedom*. Baltimore: John Hopkins Press.

Newell, A., J. C. Shaw & H. A. Simon (1957). Empirical explorations of the logic theory machine: A case study in heuristics. *Proceedings of the Western Joint Computer Conference . Institute of Radio Engineers*.

Nickles, T. (2003). Evolutionary models of innovation and the Meno problem. In L. V. Shavinina (ed.). *International Handbook on Innovation*, (pp. 54-78). Amsterdam: Elsevier.

Orr, A. H. (2004). A Passion for evolution: Review of Dawkins' 'A Devils's Chaplain'. *New York Review of Books 51*, 27-29.

Paley, W. (1825 [1802] ). *Natural Theology*: *Or Evidences of the Existence and Attributes of the Deity, Collected from the Appearances of Nature*, (The Works of William Paley, Vol. 5). London: Rivington.

Parsons, T. (1951). *The Social System*. New York: Free Press.

Paul, D. (1992). Fitness: Historical perspectives. In E. F. Keller & Lloyd, E. A. (ed.). *Keywords in Evolutionary Biology*, (pp. 112-114). Cambridge, MA: Harvard University Press.

Peel, J. (1971). *Herbert Spencer*: *The evolution of a sociologist*. London: Heinemann.

Perkins, D. N. (1998). In the country of the blind: An appreciation of Donald Campbell's vision of creative thought. *The Journal of Creative Behaviour 32*, 177-191.

Perkins, D. N. (1999). How Cheetah writes Hamlet. *Psychological Inquiry 10*, 348-350.

Pitt-Rivers, A. H. L. F. (1875). On the evolution of culture. *Journal of the Anthropological Institute 7*, 357-89.

Plotkin, H. (1994). *Darwin Machines and the Nature of Knowledge*. Cambridge, MA: Harvard Univesity Press.

Plotkin, H. (2000a). Culture and psychological mechanisms. In R. Aunger (ed.). *Darwinizing Culture: The Status of Memetics as a Science*, (pp. 69-82). Oxford: Oxford University Press.

Plotkin, H. (2000b). People do more than imitate. *Scientific American 283*, 72.

Plotkin, H. (2002). *The Imagined World Made Real*: *Towards a Natural Science of Culture*. London: Penguin.

Poincaré, H. (1982 [1908]). Mathematical creation. In H. Poincaré (ed.). *The Foundations of Science*, (pp. 383-394). Washington, DC: University Press of America. (transl. of Poincaré 1908. L' invention mathematique. In H. Poincaré. Science et Methode, (pp. 43-63). Paris: Flammurion. ).

Popper, K. R. (1972). *Objective Knowledge*: *An Evolutionary Approach*. Oxford: Clarendon Press.

Popper, K. R. (1974a). Intellectual autobiography. In P. A. Schilpp (ed.). *The Philosophy of Karl Popper*, (Vol. 1, pp. 3-184). LaSalle, IL: Open Court.

Popper, K. R. (1974b). Campbell on the evolutionary theory of knowledge. In P. A. Schilpp (ed.). *The Philosophy of Karl Popper*, (Vol. 2, pp. 1059-65). LaSalle: Open Court.

Popper, K. R. (1978). Natural selection and the emergence of mind. *Dialectica 32*, 339-355.

Popper, K. R. (1984). Evolutionary epistemology. In J. W. Pollard (ed.). *Evolutionary Theory: Paths into the Future*, (pp. 239-255). New York: Wiley.

Reader, S. M. & K. N. Laland (2003). *Animal Innovation*. Oxford: Oxford University Press.

Richards, R. J. (1977). The natural selection model of conceptual evolution. *Philosophy of Science 44*, 494-507.

Richards, R. J. (1987). *Darwin and the Emergence of Evolutionary Theories of Mind and Behavior*. Chicago: University of Chicago Press.

Richerson, P. J. & R. Boyd (2005). *Not by Genes Alone*: *How Culture Transformed Human Evolution*. Chicago: University of Chicago Press.

Rindos, D. (1985). Darwinian selection, symbolic variation, and the evolution of culture. *Current Anthropology 26*, 65-77.

Roberts, R. M. (1989). *Serendipity*: *Accidental Discoveries in Science*. New York: Wiley.

Rogers, E. M. (1995). *Diffusion of Innovations*. 4. ed. New York: Free Press.

Rosenberg, A. (1978). The supervenience of biological concepts. *Philosophy of Science 45*, 368-386.

Rosenberg, A. (1998). Methodology of social science. In E. Craig (ed.). *Routledge Encyclopedia of Philosophy, Version 1.0*, (pp. online). London and New York: Routledge.

Rothenberg, A. & C. R. Hausman (1976). *The Creativity Question*. Durham, NC: Duke University Press.

Rothenberg, A. (1986). Artistic creation as stimulated by superimposed versus combined-composite visual images. *Journal of Personality and Social Psychology 50*, 370-81.

Rubinstein, A. (1998). *Modeling Bounded Rationality*. Cambridge, MA: MIT Press.

Runciman, W. G. (1998). *The Social Animal*. London: Harper Collins.

Ruse, M. (1974). Cultural evolution. *Theory and Decision 5*, 413-40.

Ruse, M. (1977). Karl Popper's philosophy of biology. *Philosophy of Science 44*, 638-61. (repr. in Ruse, M. (ed.) (1988). But Is It Science: The Philosophical Question in the Creation/Evolution Controversy, (pp. 156-176). Buffalo, New York: Prometheus Books).

Ruse, M. (1996). The Darwin Industry: A Guide. *Victorian Studies 39*, 217-235.

Ruse, M. (ed.) (1998a). *Philosophy of Biology*. Amherst, New York: Prometheus Books.

Ruse, M. (1998b). Meme. In P. Bouissac (ed.). *Encyclopedia of Semiotics*, (pp. 404-405). Oxford: Oxford University Press.

Ruse, M. (2003). *Darwin and Design*: *Does Evolution Have a Purpose*. Cambridge, MA: Harvard University Press.

Sanderson, S. K. (1990). *Social Evolutionism*: *A Critical History*. Cambridge, MA: Blackwell.

Schooler, J. W. & Melcher J. (1995). The ineffibility of insight. In S. M. Smith, Ward, T. & Finke, R. A. (ed.). *The Creative Cognition Approach: Theory, Research, and Applications*, (pp. 97-133). Cambridge, MA: MIT Press.

Schooler, J. W. & S. Dougal (1999). Why creativity is not like the proverbial typing monkey. *Psychological Inquiry 10*, 351-356.

Scriven, M. (1959). Explanation and prediction in evolutionary theory. *Science 130*, 477-482.

Searle, J. R. (2001). *Rationality in Action*. Cambridge, MA: MIT Press.

Seifert, C. M., D. E. Meyer, N. Davidson, A. L. Patalano & I. Yaniv (1995). Demystification of cognitive insight: Opportunistic assimilation and the prepared-mind perspective. In R. J. Sternberg (ed.). *The Nature of Insight*, (pp. 65-124).

Shannon, C. & W. Weaver (1949). *The Mathematical Theory of Communication*. Urbana: University of Illinois Press.

Sherry, D. & B. Galef (1984). Cultural transmission without imitation: Milk bottle opening by birds. *Animal Behaviour 32*, 937-938.

Silverstein, A. M. (1989). *A History of Immunology*. San Diego: Academic Press.

Simmel, G. (1895). Über eine Beziehung der Selektionslehre zur Erkenntnistheorie. *Archiv für Systematische Philosophie 1*, 34-45. (repr. in Simmel, G. (1922) Zur Philosophie der Kunst - Philosophische und kunstphilosophische Aufsätze. (pp. 111-125). Potsdam: Kiepenheuer ).

Simon, H. A. (1959). Decision-making in Economics and Behavioral Science. *The American Economic Review*, 253-83.

Simon, H. A. (1981). *The Sciences of the Artificial*. 2. ed. Cambridge, MA: MIT Press.

Simonton, D. K. (1988). *Scientific Genius*: *A Psychology of Science*. Cambridge: Cambridge University Press.

Simonton, D. K. (1995). Foresight in insight? A Darwinian answer. In R. J. Sternberg (ed.). *The Nature of Insight*, (pp. 465-494). Cambridge, MA: MIT Press.

Simonton, D. K. (1997). Creative productivity: A predictive and explanatory model of career trajectories and landmarks. *Psychological Review 104*, 66-89.

Simonton, D. K. (1999a). *Origins of Genius*: *Darwinian Perspectives on Creativity*. Oxford: Oxford University Press.

Simonton, D. K. (1999b). Creativity as blind variation and selective retention: Is the creative process Darwinian. *Psychological Inquiry 10*, 309-328.

Simonton, D. K. (2001a). Creativity as cognitive selection: The blind-variation and selective-retention model. Commentary on Hull et al. *Behavioral and Brain Sciences 24*, 554-556.

Simonton, D. K. (2001b). Creativity as a secondary Darwinian process. *Bulletin of Psychology and the Arts 2*, 33-39.

Simonton, D. K. (2003). Human creativity: Two Darwinian analyses. In S. M. Reader & Laland, K. N. (ed.). *Animal Innovation*, (pp. 309-325). Oxford: Oxford University Press.

Skagestad, P. (1978). Taking evolution seriously: Critical comments on D.T. Campbell's evolutionary epistemology. *The Monist 61*, 611-621.

Skinner, B. F. (1953). *Science and Human Behavior*. New York: Free Press.

Smart, J. (1963). *Philosophy of Scientific Realism*. London: Routledge and Kegan Paul.

Smith, E. A. & B. Winterhalder (ed.) (1992). *Evolutionary Ecology and Human Behavior*. New York: deGruyter.

Smith, S. M., T. B. Ward, R. A. Finke & (ed.) (1992). *The Creative Cognition Approach*. Cambridge, MA: MIT Press.

Smith, E. A. (2000). Three styles in the evolutionary analysis of human behavior. In L. Cronk, Chagnon, N.A. & Irons, W. (ed.). *Adaptation and Human Behavior: An Anthropological Perspective*, (pp. 27-46). New York: DeGruyter.

Snow, C. P. (1969). *The Two Cultures*: *And a Second Look*. Cambridge: Cambridge University Press.

Sober, E. & R. C. Lewontin (1982). Artifact, cause and genic selection. *Philosophy of Science 49*, 157-180. (repr. in Brandon & Burian 1984: 109-132.).

Sober, E. (1992). Models of cultural evolution. In P. Griffiths (ed.). *Trees of Life: Essays in Philosophy of Biology*, (pp. 17-39). Dordrecht : Kluwer.

Sober, E. (1993). *Philosophy of Biology*. 1. ed. Oxford: Oxford University Press.

Sober, E. & D. S. Wilson (1994). A critical review of philosophical work on the units of selection problem. *Philosophy of Science 61*, 534-555.

Sober, E. (2000). *Philosophy of Biology*. 2. ed. Boulder: Westview.

Sober, E. (2001). The two faces of fitness. In R. S. Singh, Krimbas, C.B., Paul, D.B. & Beatty, J. (ed.). *Thinking about Evolution: Historical, Philosophical, and Political Perspectives*, (Vol. 2, pp. 309-321). Cambridge: Cambridge University Press.

Söderquist, T. (1994). Darwinian overtones: Niels K. Jerne and the origin of the selection theory of antibody formation. *Journal of the History of Biology 27*, 481-529.

Spencer, H. (1864). *Principles of Biology, 2. vols.*. London: Williams and Norgate.

Sperber, D. (1996). *Explaining Culture*: *A Naturalist Approach*. Oxford: Blackwell.

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. 163-174). Oxford: Oxford University Press.

Spohn, W. (2002) Die vielen Facetten der Rationalitätstheorie (*Forschungsberichte der DFG-Forschergruppe "Logik in der Philosophie"*, Universität Konstanz, Nr. 38).

Stein, E. & P. Lipton (1989). Where guesses come from: Evolutionary epistemology and the anomaly of guided variation. *Biology and Philosophy 4*, 33-56.

Sterelny, K. & P. Kitcher (1988). The return of the gene. *Journal of Philosophy 85*, 339-361.

Sterelny, K., K. Smith & M. Dickinson (1996). The extended replicator. *Biology and Philosophy 11*, 377-403.

Sterelny, K. & P. E. Griffiths (1999). *Sex and Death*: *An Introduction to Philosophy of Biology*. Chicago, London: University of Chicago Press.

Sterelny (forthcoming a). Memes revisited *(draft of Nov 2004)*.

Sterelny, K. (forthcoming b). Evolution and evolvability of culture *(draft of Oct 2003)*.

Sternberg, R. J. & J. E. Davidson (ed.) (1995). *The Nature of Insight*. Cambridge, MA: MIT Press.

Sternberg, R. J. (1998). Cognitive mechanisms in human creativity: Is variation blind or sighted? *Journal of Creative Behavior 32*, 159-176.

Sternberg, R. J. & T. I. Lubart (1999). The concept of creativity: Prospects and paradigms. In R. J. Sternberg (ed.). *Handbook of Creativity*, (pp. 3-15). Cambridge: Cambridge University Press.

Sternberg, R. J. (1999b). Darwinian creativity as a conventional religious faith. *Psychological Inquiry 10*, 357-359.

Stich, S. P. (1990). *The Fragmentation of Reason*: *Preface to a Pragmatic Theory of Cognitive Evaluation*. Cambridge, MA: MIT Press.

Szathmary, E. (1999). Chemes, genes, memes : A revised classification of replicators. *Lectures on Mathematics in the Life Sciences 26*, 1-10.

Thagard, P. (1988). Against evolutionary epistemology. In P. Thagard (ed.). *Computational Philosophy of Science*, (pp. 101-111). Cambridge, MA: MIT Press.

Thorndike, E. L. (1898). Animal intelligence: An experimental study of the associative process in animals. *Psychological Review Monographs 2*, 1-109.

Tomasello, M. (1999). *The Cultural Origins of Human Cognition*. Cambridge, MA: Harvard University Press.

Tonegawa, S. (1983). Somatic generation of antibody diversity. *Nature 302*, 575-581.

Toulmin, S. (1972). *Human Understanding*: *The Collective Use and Evolution of Concepts*. Princeton, NJ: Princeton University Press.

Tylor, E. B. (1958 [1871]). *Primitive Culture*: *The Origins of Culture*. New York: Harper & Row.

Vayda, A. P. (1994). Actions, variations, and change: The emerging anti-essentialist view in anthropology. In R. Borofsky (ed.). *Assessing Cultural Anthropology*, (pp. 320-329). New York: McGraw Hill.

Wallas, G. (1926). *The Art of Thought*. New York: Harcourt-Brace.

Ward, T. B., S. M. Smith & R. A. Finke (1999). Creative cognition. In R. J. Sternberg (ed.). *Handbook of Creativity*, (pp. 189-212). Cambridge: Cambridge University Press.

Waters, K. (1991). Tempered realism about the force of selection. *Philosophy of Science 58*, 553-573.

Weingart, P., R. Boyd, W. H. Durham & P. J. Richerson (1997). Units of culture, types of transmission. In P. Weingart, Peter J. Richerson; S.D. Mitchell, & S. Maasen (ed.). *Human by Nature: Between Biology and the Social Sciences*, (pp. 299-325). Mahway, NJ/ London: Lawrence Erlbaum Ass.

Weisberg, R. W. (1993). *Creativity*: *Beyond the Myth of Genius*. New York: Freeman and Company.

Weismann, A. (1892). *Das Keimplasma*: *Eine Theorie der Vererbung*. Jena: Gustav Fischer. (transl. 1893. The Germ Plasm: A Theory of Heredity. Transl. by. N. Parker and H. Ronfeldt, London. ).

Wiener, P. P. (1949). *Evolution and the Founders of Pragmatism*. New York: Harper.

Wilder, R. L. (1968). *Evolution of Mathematical Concepts*. Milton Keynes: Open University Press.

Wilkins, J. S. (1998). What's in a meme?: Reflections from the perspective of the history and philosophy of evolutionary biology. *Journal of Memetics 2*, http://jom-emit.cfpm.org/1998/vol2/wilkins_js.html.

Wilkins, J. (1999). Memes ain't (just) in the head. *Journal of Memetics 3*, http://jom-emit.cfpm.org/1999/vol3/wilkins_j.html.

Williams, G. C. (1966). *Adaptations and Natural Selection*: *A Critique of Some Current Evolutionary Thought*. Princeton, NJ: Princeton University Press.

Wilson, E. O. (1975). *Sociobiology*: *The New Synthesis*. Cambridge, MA: Harvard University Press.

Wilson, D. S. (1999). Flying over uncharted territory: Review of 'The Meme Machine' by Susan Blackmore. *Science 285*, 206.

Wimsatt, W. C. (1980). Reductionistic research strategies and their biases in the units of selection controversy. In T. Nickles (ed.). *Scientific Discovery: Case Studies*, (Vol. 2, pp. 213-259). Dordrecht: Reidel.

Wimsatt, W. C. (1999). Genes, Memes and Cultural Heredity. *Biology and Philosophy 14*, 279-310.

Wright, L. (1973). Functions. *Philosophical Review 82*, 139-68.

Zirkle, C. (1946). The early history of the idea of the inheritance of acquired characters and of pangenesis. *Transactions of the American Philosophical Society 35*, 91-151.