Thiemo Breyer *Editor*

Epistemological Dimensions of Evolutionary Psychology

Springer

Epistemological Dimensions of Evolutionary Psychology

Thiemo Breyer
Editor

Epistemological Dimensions of Evolutionary Psychology

Springer

Editor
Thiemo Breyer
a.r.t.e.s. Graduate School for the Humanities
University of Cologne
Cologne
Germany

ISBN 978-1-4939-1386-2 ISBN 978-1-4939-1387-9 (eBook)
DOI 10.1007/978-1-4939-1387-9
Springer New York Heidelberg Dordrecht London

Library of Congress Control Number: 2014955339

© Springer Science+Business Media New York 2015
This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

To Olli, companion in rites of passage

Preface

Evolutionary Psychology (EP) is currently one of the most fascinating fields of interdisciplinary research on the human mind, brain, and behaviour. According to EP, the brain is an organ whose computational capacities have been selected for because they provided advantages of survival in the course of evolution. Cognitive abilities are seen as the results of specific reactions to evolutionary pressures, which have shaped our ancestors: 'Natural selection shapes domain-specific mechanisms so that their structure meshes with the evolutionarily stable features of their particular problem domains. Understanding the evolutionarily stable feature of problem domains—and what selection favoured as a solution under ancestral conditions—illuminates the design of cognitive specialisations' (Cosmides and Tooby 1994). In this volume, some of the epistemological premises and problems of this scientific paradigm shall be addressed to critically assess the status of EP in contemporary discussions in the philosophy of mind and cognitive science.

Some of the fundamental ideas of EP had been discussed already at the beginning of the twentieth century. For a few decades, though, they were suppressed by the upsurge of three more prominent paradigms: (1) behaviourism and its focus on ontogeny as opposed to phylogeny; (2) philosophical and common sense convictions that the individual is essentially free and can overcome biological determinations; (3) cultural relativism and its assumption that culture is the main factor that shapes social groups and the minds of their members. Since the 1970/1980s, however, the ideas of EP, coupled with the success of sociobiology (Wilson 1975), have become increasingly more important as a *theoretical* bridge between the natural, behavioural, social, and cognitive sciences investigating human nature as well as a rich domain of *empirical* research (cf. the journal *Evolutionary Psychology*, established in 2003).

In light of this development, at least three bundles of theoretical issues concerning the rationality and validity of EP arguments can be addressed: (1) the relationship between the biological foundations of cognitive processes and mental states; (2) the relationship between organism, brain, and brain modules; (3) the relationship between the things EP tries to explain and the way it does explain them.

It is sometimes supposed that EP contains a type of *naturalistic fallacy* (Moore 1903), that is the misconception that higher-order principles (such as ethical principles) could be derived from what happens in the natural world. More generally, it is supposed that EP is characterised by a strong reductionism, namely the attempt to reduce mental phenomena to neuronal phenomena and from there ultimately to genetic phenomena. In the philosophy of mind, the problems of reductionist programmes are widely discussed. Apart from the suggested possibility of *multiple realisations* (Fodor 1974) and the pluralism of descriptive levels (Dupré 1993), there might be the 'phenomenological' problem that EP needs to identify cognitive processes by applying folk-psychological or introspective notions of mental phenomena, which are then subjected to an evolutionary explanation. In other words, the formation of EP's explananda is not accomplished in terms of 'hard science' alone—it must be known before the scientific analysis what we mean by concepts such as 'will', 'action', 'deliberation', 'desire', 'emotion', etc. How can philosophy help here with detailed descriptions of the structures of consciousness and experience and the conceptual analysis of folk-psychological vocabularies?

Another line of criticism stems from the arguments developed by Bennett and Hacker (2003), who outline a critical framework for the evaluation of recent approaches in neuroscience. They remind us that for some brain scientists and neurophilosophers, the brain does all kinds of things: it believes (Crick), knows (Blakemore), interprets (Edelman), questions itself (Young), contains symbols (Gregory), represents information (Marr), and makes decisions (Damasio). Behind these assertions lies, in Bennett and Hacker's view, a *mereological fallacy*, because the conception of the 'brain' is unduly inflated by ascribing activities and powers to it that are usually applied to organisms as wholes. Even if one admits that there is empirical evidence for the correlation of complex subjective *wholes* such as decision-making and specific physical *parts* of those processes such as neuronal firings, the identification of a part with a whole would be an invalid inference.

The concerns Bennett and Hacker raise with regard to neuroscience seem relevant with regard to EP as well, where it is often claimed that particular brain modules fulfil complex functions that are usually attributed to subjects. In other words, if it does not make sense to say that the brain (part of the organism) does all these complex things, then it makes even less sense to say that a module (part of the brain) does them. The mereological fallacy would be transposed to a more fine-grained level and, as some would argue, such a shift does not solve the problem, but makes it even more serious. What would thus be needed is a *mereological analysis* that helps to ground an *organismic ontology*.

Another problem of EP could be seen in the temporal structure of the development of psychological faculties and features and its translation into a scientifically adequate mediation. The *narrative fallacy* (Taleb 2007), which can be attributed to various scientific models of explanation, amounts to the retrospective construction of a story to make a particular occurrence appear plausible. In the case of EP theorising, some think that this fallacy consists in using current cognitive and behavioural phenomena (P_c) to explain past phenomena (P_p), which are *then* believed to be the necessary conditions for the existence of P_c. The idea that seemingly legitimises this line of reasoning is that every P_c, for example a behavioural disposition of an individual of a certain species, has only survived until the present because its correlating P_p was beneficial for the survival of the species in question.

By assuming that the behaviour of modern humans can only be explained in terms of adaptations that occurred in a Stone Age environment, EP presupposes that human evolution has not produced drastic changes ever since, at least when it comes to such factors as brain size and structure. This heuristic presupposition is called, in paleodemographics, the *principle of uniformity* (Grupe et al. 2005). From a philosophical point of view, this assumption raises the question of how we can achieve epistemic access to the 'lifeworld' of our ancestors. Two methodical possibilities of EP to deal with this are (1) the analysis of *relics* and (2) the *back projection* from present hunter-gatherer societies to prehistoric times. But do these suffice to reconstruct the lifeworld of individuals at the time when crucial adaptations are believed to have taken place? First, the inference from a fragmented physical reality to a complex psychological reality seems to leave many parameters undefined; and second, modern day hunter-gatherers do not exist in isolation from processes of globalisation and social change, so that they cannot be regarded as models of an archaic human condition. These theoretical and practical questions of how the explananda as well as the explanations of EP are constructed, and many more issues, are discussed in this volume to elucidate the ways we think about human evolution on the one hand, and the argumentative structure of EP explanations and narratives on the other.

The recent literature is characterised by fierce attacks of philosophers against the programme of EP (e.g. Buller 2006; Lewontin 2007; Richardson 2007), and equally emphatic defences or counter-attacks by proponents of EP (e.g. Pinker 2009; Buss 2011; Hoch-Olesen et al. 2011). Some fear that EP 'might constitute a genuine threat to the contemporary moral order' (Hagan 2005), while others celebrate its Darwinistic strategy as the key to our understanding of ourselves: 'Natural selection has a special place in science because it *alone* explains what makes life special' (Pinker 1997).

This volume brings together the expertise of philosophers and psychologists to explore the interdisciplinary ground for fruitful discussions in the middle sphere between such extreme positions by investigating the epistemological dimensions of EP. The point of departure for the collaborative work on this volume was the conference 'Epistemological Foundations of Evolutionary Psychology', which took place at the University of Heidelberg's *Internationales Wissenschaftsforum* (IWH) from March 16–17, 2012. From there, the project has developed for quite a while, as new authors joined in and as the outlook was specified. I am grateful for all the discussions we have had at the conference and beyond and thank all authors for their fascinating contributions.

My special thanks go to Christopher Gutland for his invaluable work as a research fellow in the project 'Anthropology of Intersubjectivity', in the context of which the conference and this volume are embedded. The project was funded from 2011–2013 by the *Baden-Württemberg Foundation* (Az. 1.16101.08), to which I owe my gratitude. Finally, I would like to thank Birgit Schubert for her editorial assistance.

Cologne T. Breyer
March 2014

Contents

Contributors

Alan Costall Department of Psychology, University of Portsmouth, Portsmouth, UK

Malte H. Dahlgrün Department of Philosophy, Humboldt University of Berlin, Berlin, Germany

Ulrich J. Frey University of Halle, Halle (Saale), Germany

Michael Hampe Swiss Federal Institute of Technology, Zurich, Switzerland

Daniel Haun University of Jena, Jena, Germany

Christine Hennighausen Department of Media Psychology, University of Würzburg, Würzburg, Germany

Stefanie Hoehl Institute of Psychology, University of Heidelberg, Heidelberg, Germany

Andreas Mayer Institute of Psychology, University of Osnabrück, Osnabrück, Germany

Alexander Neumann Maastricht University, Maastricht, Netherlands

Harriet Over University of York, York, UK

Sabina Pauen Institute of Psychology, University of Heidelberg, Heidelberg, Germany

Robert C. Richardson Department of Philosophy, University of Cincinnati, Cincinnati, (OH), USA

Felix D. Schönbrodt Department of Psychology, Ludwig-Maximilians-University, Munich, Germany

Frank Schwab Department of Media Psychology, University of Würzburg, Würzburg, Germany

Sven Walter University of Osnabrück, Osnabrück, Germany

Michael Zehetleitner Department of Psychology, Ludwig-Maximilians-University, Munich, Germany

Chapter 1
Philosophy and Evolution

Michael Hampe

1.1 The Dispute Concerning the Naturalisability of Values

Nowadays, the discussion of biological evolutionary theory within philosophy occurs primarily within the context of *naturalisation projects,* which prioritise scientific descriptions and explanations of man over those from religion, the social or the cultural sciences. The dispute over the question whether human concepts of values might not ultimately be the result of natural circumstances is as old as philosophy itself and refers to both practical values, expressed in terms of "good" and "bad", as well as theoretical values, which may manifest themselves in a belief being characterised as "true" or "false". Is the "good" or "just" action the one of the (physically) fitter? Is the true belief the one that has promoted survival? Do "right" actions and thoughts therefore only result from the power developed by individual people or groups of human beings faced with threats to their survival? Philosophy has been dealing with such problems since Plato's *Protagoras* dialogue. In 1740, David Hume unmasked the omnipresent *naturalistic fallacy* within this context by separating the realm of facts from that of norms, declaring that norms ("ought") are falsely derived from the description of facts ("is"), and thereby appeared to have finally put a stop to naturalisation efforts of this kind (Hume 1978, p. 460; cf. Flew in Caplan 1978). These efforts however received new impetus from Darwin's evolutionary theory. Building on the theories of sociobiology (Wilson 1975; Caplan 1978; Clutton-B. and Harvey 1978; King's College Sociobiology Group 1982; Randall 1983), the 1970s and 1980s saw a resurgence of *evolutionary ethics*, a concept originally developed by Herbert Spencer (Spencer 1893; Richards 1993; Farber 1998). Furthermore, the application of insights from perceptual psychology and neuroscience resulted in the emergence of ideas leaning towards *evolutionary epistemologies* (Campbell 1974; Fenk 1990; Lorenz 1973; Popper 1973, 1994;

M. Hampe (✉)
Swiss Federal Institute of Technology, Zurich, Switzerland
e-mail: hampe@phil.gess.ethz.ch

T. Breyer (ed.), *Epistemological Dimensions of Evolutionary Psychology,*
DOI 10.1007/978-1-4939-1387-9_1

Riedl and Delpos 1996; Vollmer 1975). These disciplines were deemed viable not only by biologists but also by academic representatives of philosophy. However, they were unable to refute the diagnosis that inferences from facts to norms are *fallacies*. Since the second half of the nineteenth century and throughout the twentieth century, the public fascination surrounding new insights in genetics, behavioural and neurological sciences regarding the development of the biological conditions of human action and knowledge (e.g. Dawkins 1978, 1986; Pinker 1997, 2002), which were usually presented in a popularised manner and with a fittingly revolutionary emphasis, continually disregarding the difference between *conditions* (causal, structural or perhaps anatomical in nature) and *reasons* (as decisions for actions or claims to knowledge), has brought about the false impression that evolutionary biology may one day answer normative questions after all.

There is no need to recapitulate the stages of this circular journey, with biologists demanding that philosophers should finally acknowledge some new finding concerning the human genome or the neuronal apparatus, and philosophers trying to make biologists take into account the difference between conditions and causes on the one hand and reasons and justifications on the other (see Lütterfelds 1987, 1993; Gräfarth 1997; Meyer 2000). This confrontation has been evident ever since the time of Plato and will probably resurge in every scientific generation.

This circular motion could now also branch off into two directions. On the one hand, it is possible that instead of continuing to produce naturalistic fallacies, the use of the terminology of justification and reason will simply decrease in both casual and academic language within the fields of ethics, law and science, whereas the use of causal modes of expression will increase. This, however, could then no longer be reconstructed as a process triggered by well-justified insights, as the act of characterising insights as being "well justified" would no longer be possible without a performative contradiction. Rather, a kind of cultural revolution would have to result in the replacement of norm-based criminal jurisdiction by organic therapies, and of education based on arguments regarding norm-regulated behaviour by physical punishment and drill. Nowadays, this is surely nothing but a negative utopia of the kind portrayed in certain science fiction fantasies (cf. Burgess 1962). Biologists also view cultural development as a largely autonomous process. Thomas Henry Huxley, for example, regarded the norms of society as being in opposition to nature (Huxley 1989, p. 31) and Richard Dawkins saw "memes" as the units of cultural development, as opposed to and clearly distinguishable from genes, which he considered to be the decisive objects of biological evolution. Whilst recognition of the autonomy of cultural developments is still widespread, even though philosophers consider the naturalistic fallacy argument to be compelling, it is nevertheless conceivable that in a fully scientific society, another mode of thought could gain acceptance at some far away point in the future, (as happened in the Age of Enlightenment, when philosophy prevailed over theology). As previously explained, however, this would be nothing less than a cultural revolution.

On the other hand, it is conceivable that this circular motion within the discussion of naturalism will branch off in the direction of a new religious dogmatism, perhaps derived from Islam or Christian fundamentalism, as all fundamentalist interpretations of Mosaic religions are adverse to those naturalisms which are motivated by evolutionary ideas. Thus, for instance, Christian fundamentalism is once again making itself heard in the dispute pertaining to evolutionary theory between the life sciences on the one hand and philosophy, jurisprudence as well as the critical social and cultural sciences on the other hand. This development is however limited to (but not only) the USA, where the State of Kansas (with the support of the former President, George W. Bush) ruled that the biblical story of creation and the assumption that nature is a product of an “intelligent design” is to be taught as an “equal hypothesis” alongside Darwinian evolutionary theory (Dennett 1995; Schrader 2007). Within this context, creationists have abused a diagnosis which applies to *any* accepted theory and actually stems from the philosophy of science, i.e. that evolutionary theory is merely a *hypothesis* and has thus not *definitively* been proven to be true. On many occasions, the hypothetical nature of evolutionary theory is portrayed as something that negatively sets it apart from other scientific theories. The scientific philosophy of evolutionary theory does not, however, allow for parallels to be drawn between accounts of creation and Darwin’s assumptions, because according to enlightened philosophy, the texts on the history of creation do not represent any systems of hypothesis (i.e. theories) whatsoever. From a philosophical perspective, creationism conflates the truism that scientific theories are in principle hypothetical, with a distorted or outdated pre-enlightenment view of the biblical story of creation, that is, the view that a theory on the development of facts is reflected therein.

A brief look at the history of cosmologies and theories of creation since antiquity will provide more insight into what is at stake here. Since antiquity, there have been varying alternative views in cosmology. In Plato’s *Timaeus,* for example, the world was *technically created* by the Demiurge, whereas in *De Caelo,* Aristotle posited that it was neither made nor developed, but that the world is *eternal*. Both of these ideas contrast with the concept of a world as found in Democritus and Lucretius, namely that it came about *by chance,* has subsequently gone through a process of development and will again cease to exist. Along with Christianity came the dominance of Platonic technomorphic cosmogonical views, which changed again from the seventeenth century onwards with the development of historically critical biblical studies (since Spinoza’s (1994) *Theologico-Political Treatise*) and the demythologisation of Christianity (as well as the demythologisation of the New Testament in the twentieth century, with authors like Bultmann 1941). Since that time, the biblical texts have been understood *metaphorically,* with the “accounts” of creation no longer being read as representations of facts, but instead viewed to be *manifestations of general attitudes and world views* (cf. Wittgenstein 1966, p. 54f.). They provide guidance in a life in which nature and the living beings existing in nature are treated *as if* they had been created by a supremely wise and good being. This reflected, critical view spread throughout modernity, at least within the cultural spheres dominated by Christianity. However, it remains to be seen whether or not the forms of religious fundamentalism which resort to a pre-modern interpretation of religious texts (as can be seen throughout the world) might be able to gain cultural influence in the longer term. Aided by a distorted use of insights from the philosophy of science, these forms of fundamentalism may thus be able to assert themselves against the general spread of Darwinian evolutionary theory, and in doing so, to replace the enlightened religious and the corresponding hypothetical scientific understanding of the world by religious propaganda.

Now, the history of changing interpretations can itself be considered as a possible argument against the idea of a "natural" determinacy of human thought and action. This is because cultural processes of the kind described above, which bring about a fundamental change to the landscape of discourse and the established self-definitions and finally result in a revision of people's *self-conception,* can be interpreted as proof of the *semiotic autonomy* of people, even if one does not welcome them in a philosophical sense. They show that the way in which people relate to themselves, aided by systems of signs, is not naturally *determined,* nor is it informed by truth alone; it rather "drifts" in a cultural sense, and changes faster than genetic material. However, the idea of a semiotic autonomy of this kind being the reason for a possible drift from a religious self-conception to one which is defined by culture and history, continuing to a self-conception based on naturalist evolutionary biology and perhaps back again to a fundamentalist religious image of mankind, is in contradiction with the idea of biological determinism in the sense of evolutionary epistemology, according to which the development of the human self-image must also be subject to a process of adaptation to reality. Thus, individual and collective autonomy, including semiotic autonomy, could only be regarded as an illusion (Hampe 2006, pp. 72–75) and the historical processes based on this autonomy would remain inexplicable. It must therefore be asked whether a self-conception of human knowledge and action that is purely based on evolutionary biology (which, in such a pure form, is probably not actually defended by anyone within biology) and renounces any terminology of justification as supposedly unfounded, does not necessarily lead to a performative contradiction. Such a self-conception could not be shown to be theoretically justifiable, nor could it provide normative legitimisation for the practical cultural consequences in the drift of human self-conception, which would form the very basis for its own acceptance (cf. Hampe 2007, pp. 11 f., 171–177).

1.2 The Boston Metaphysical Club and the Issue of the Generalisability of Evolutionary Theory

It would be a daunting task to try to trace the effects of Darwinian evolutionary theory in the areas of philosophy not connected to the project of naturalising the theory of human cognition and action, given that the influence of evolutionary theory can be found in completely different schools of philosophy. It has had an impact on authors whose works have entered into the philosophical canon, for example Nietzsche and Foucault, as well as others who have now been largely forgotten, such as Teilhard de Chardin and James Baldwin. Given that it is not possible to provide even a rough overview of these wide-reaching effects of evolutionary theory in the different schools and levels of philosophy within the scope of this chapter, only one *metaphysical* problem will be investigated here: the problem of the *nature and origin of the mind.* This is a central problem for philosophy as a whole, but evolutionary theory played a decisive role in this discussion. Two types of theories about the nature and the origin of the mind are considered here: *pan-psychism* and the *emergent theory* of the mind. Despite being incompatible, they both received great impetus from evolutionary theory. Both conceive the mind as a product of natural development. Emergent theory claims that the mind "emerges" as a new, unpredictable property as the organic matter becomes increasingly complex during the development of life. Conversely, the term "pan-psychism" stands for the belief that (1) all living beings form an evolutionary continuity and (2) people possess minds. It therefore follows that all living beings must in some form or another possess a mind. In order to understand these two theories in their historical context, we must take into account the assessments of the general philosophical "scope" of evolutionary theory at the time of its origin.

The fact that evolutionary theory equally supported/nurtured two incompatible philosophical views of the mind shows in a very general sense an important aspect of its effect on philosophy. Just as it is the case with other results from other sciences, this successful biological theory has not led to a *decision* between competing philosophical views, but has merely changed the discursive landscape in which these philosophical theories compete with one another.

It was in Boston in the second half of the nineteenth century where the relevance of evolutionary theory to philosophy was discussed by an association of individuals who would later become famous American intellectuals, amongst them Chauncey Wright, Charles Sanders Peirce, Oliver Wendell Holmes and William James. As of 1872, this group referred to itself as the "metaphysical club", reading and debating Darwin's *On the Origin of Species* (1859) immediately after its publication (Menand 2001, pp. 120–140). In his discussion of Darwin's theory, Chauncey Wright was the first to warn against transferring this theory to other sciences without having an empirical basis to do so, as well as against philosophical generalisations per se. According to Wright, Darwin's theory could only be applied to the development of living beings; it was compelling within this context, but not applicable to inorganic matter or to the cultural and intellectual spheres of reality. Wright also thought that no general cosmological principles could be inferred from the theory. Instead, he viewed the development of the inorganic cosmos as a kind of "cosmic weather" which we still cannot understand completely, even in light of Darwin's theory of evolution (Wright in Madden 1958, pp. 106–117). For him, cosmology was not a science at all, neither could it be made into one using evolutionary theory; rather, the development of the world as a whole tumbles into an "infinite variety of manifestations of causes and laws" without there being any apparent natural tendency that would allow for the theoretical standardisation of the relevant factors (Wright in Madden 1958, p. 113). Instead of drawing an analogy between the development of the world as a whole and that of living beings, Wright's inclination was to compare the development of the world with that of the weather and the climate: a complex, contingent process which cannot be grasped by one discipline alone.

With these reflections, he stood in opposition to the transference of evolutionary theory to other sciences, which began to occur soon after the publication of Darwin's *On the Origin of Species*. He was also challenging generalisations such as those made by one of Darwin's most well-known contemporaries, Herbert Spencer, who first coined the phrase "survival of the fittest" in his *Principles of Biology* before it was adopted by Darwin in later editions of *On the Origin of Species*. Spencer accordingly became a target of criticism from Wright, who viewed Spencer's "speculations" as nothing other than abstract representations of cosmological terms "which the human mind spontaneously supplies in the absence of facts". The assumption of a teleology of the evolution of the cosmos towards the complex or even towards chaos reflects the workings of a "mythical instinct" in the idea of as aiming at an ultimate "order" or "disorder" (Wright in Madden 1971, p. 18 f.). According to Wright, this effort is not justified by evolutionary theory. It must be noted that this critique would also affect evolutionary generalisations of thermodynamics to cosmology, which became widespread as the theory of the "heat death of the universe" in the wake of physicist Ludwig Boltzmann's theory of gases (cf. Boltzmann 1896). Wright also voiced this critique against reflections from Spencer, such as the following:

> Everywhere and to the last, the change at any moment going on forms a part of one or other of the two processes. While the general history of every aggregate [*of matter,* M.H.] is definable as a change from a diffused imperceptible state [*of matter*, M.H.] to a concentrated perceptible state, and again to a diffuse, imperceptible state; every detail of the history is definable as a part of either the one change or the other. This, then, must be that universal law of the redistribution of matter and motion […]. The [*natural*, M.H.] processes thus everywhere in antagonism and everywhere gaining now a temporary and now a more or less permanent triumph of the one over the other, we call *Evolution* and *Dissolution*. Evolution under its simplest and most general aspect is the integration of matter and concomitant dissipation of motion; while Dissolution is the absorption of motion and concomitant disintegration of matter. (Spencer 1860/1971, p. 59)

Within the framework of his "system of philosophy", Herbert Spencer's books provided an evolutionary psychology, pedagogy, and sociology and were widely disseminated despite (or perhaps because of) the prevalence of empty phrases such as those cited above. The situation is completely different with Charles Sanders Peirce, another universalising evolutionist of the "metaphysical club". Although the conclusions Peirce drew from evolutionary theory were less trivial than those drawn by Spencer, they were just as general. His evolutionary cosmology therefore provides a good basis from which the contrast between emergentism and pan-psychism can be characterised.

1.3 Evolution of Natural Laws and Pan-Psychism

Wright's view was criticised by Peirce, who insisted that philosophy should become much more evolutionary than with Spencer. By this he meant that Spencer's assumption of quasi-mechanical laws of evolution controlling the aggregation or dissipation of matter was still trapped in a pre-Darwinian conception of nature. Peirce wrote,

> Spencer is, in short, no evolutionist; but only a semi-evolutionist […] *Thorough-going* evolutionism is called for in philosophy […]. Now the only way to do that is to show in some way that *law* may have been a product of growth, of evolution. Then we must make some principle of growth more fundamental than any mechanical law. The suggestion to which this leads is obvious. It is *that matter is mind* which has come under the almost complete domination of habit. That at first only mind existed, a vast unpersonalized manifold of mind. (Peirce 1991, pp. 129, 138, emphasis by M.H.).

If, following Aristotle, one interprets the constancy of natural species as laws of reproduction, which ensure that what comes from a frog is another frog, and that what comes from a human being is another human being and not the other way around, then with his theory of the evolution of the species, Darwin *historicised* the laws of reproduction (Gilson 1971). For both Plato and Aristotle, the constancy of the species was *the* paradigm of nature's regularity (the regularities of the heavenly bodies' motions are not included here as they were not viewed as being natural by the authors preceding Galileo, because no laws of motion existed for the irregular accelerations and decelerations of bodies under the moon). Peirce takes Darwin's historicisation of the species as a kind of historicisation of natural laws in general, generalising it into a conception of nature that states that *all* laws have a history, even the physical laws such as the law of gravity. For him, "the only possible way of explaining the laws of nature and uniformity in general [is in] accepting it as a result of evolution, i.e. however, accepting that they are not absolute and that they cannot be traced precisely" (Peirce 1991b, p. 271).

But what does this historicism have to do with the call for matter to be conceived as mind? Peirce believes that the *development* of regularities can be observed in each mind from an internal perspective, whereby one, as a sentient being, is able to ask what it is like to instantiate a given regularity by oneself, when the mind *forms habits* of perception and action. Furthermore, when speaking of the formation of habits, it can be said that the probability of a habitual perception or action being realised increases if it has occurred more frequently in the past. Hence, the solidity of a habit can be interpreted as an indicator of the length of its history, or, in other words: The more fixed a habit, the older the history of its formation; the more exceptions still permitted by a habit, the younger it must be.

This train of thought leads Peirce to develop a speculative evolutionary theory of regularities, where the inorganic laws of nature—for example, the laws of matter—represent very old habits of the world. These are habits that were formed at a very early stage of the cosmos and are therefore realised today with a probability of 1. Their history has come to an end. The regularities of organic nature are also a product of nature's long development process; however, they are not yet realised with absolute necessity. It can therefore be concluded that these regularities are "younger" than the inorganic laws. Finally, the regularities of the cultural world and personal life are the most variable, contain the highest number of exceptions and are therefore the youngest. Yet the entire natural process is one of mental development and the taking of habits. With its unshakeable necessities, inorganic matter is thus nothing other than a mind congealed in old habits.

Through the principle of the taking of habits, this Peircean concept of nature is arguably as closely connected to Lamarck as to Darwin, as Peirce himself also acknowledges. (Darwin himself had believed at some stage during the development of his theory that habitually acquired structures could become hereditary.) However, the discussions of *The Origin of Species* in the “metaphysical club” are likely to have provided Peirce with the impetus for the development of his universalised evolutionary theory of the cosmos. Peirce’s account of the history of the universe begins with an unstructured sequence of qualities of sensation and random combinations of these. He sees an intellectual law at work above the surface of this random sea of combinations of sensations, namely the aforementioned law of habit-taking which reinforces itself and brings forth all of the regularities, from the physical laws to the cultural patterns. He writes,

> if the laws of nature are results of evolution, this evolution must proceed according to some principle; and this principle will itself be of the nature of a law. But it must be such a law that it can evolve or develop itself. Not that if absolutely absent it would create itself perhaps, but such that it would strengthen itself, and looking back into the past we should be looking back [to] times in which its strength was less than any given strength, and so that at the limit of the infinitely distant past it should vanish altogether. Then the problem was to imagine any kind of law or tendency which would thus have a tendency *to strengthen itself*. Evidently it must be a tendency toward generalization […]. Now the generalizing tendency is the great law of mind, the law of association, the *law of habit taking* […]. Hence I was led to the hypothesis that the laws of the universe have been formed under a universal tendency of all things toward generalization and habit-taking. (Peirce 1992, p. 241)

It was not biology alone, but also Peirce’s reception of the philosophy of German idealism, particularly Hegel and Schelling, that fed into his development of an evolutionary philosophy of nature. Schelling had already used the term “evolution” in his *System of Transcendental Idealism* in 1800 in relation to his theory of living beings:

> Organisation [of living beings, M.H.] is succession hampered and, as it were, coagulated in its course […]. But now it appears from this very deduction of life […] that there can be no distinction between living and non-living organizations […]. Since the intelligence is to intuit itself as active in the successions throughout the whole of organic nature, every organisation must also possess life in the wider sense of the word, that is, must have an inner principle of motion within itself. The life in question may well be more or less restricted; the question, therefore: whence this distinction? reduces itself to the previous one: whence the graduated sequence in organic nature? But this scale of organisations merely refers to different stages in the evolution of the universe. (Schelling 1800/1978, p. 124f.)

Schelling presupposes a hierarchy of living beings (a widespread assumption since Neo-Platonism), which he however historicises in an evolution of the universe. The hierarchy is based on intelligence and forms a temporal sequence. Despite the knowledge of this text, Peirce however joins the development of nature and of mind in a pan-psychist evolutionary cosmology, whereas according to Schelling’s thinking, the development of the mind is separated from that of nature, as he sees transcendental and natural philosophy as complementary systems that run parallel to one another. The structural analogies of evolutionary thought in Schelling and Peirce have been researched in detail (Esposito 1980; Hampe 1999). According to

Peirce, the universe hypothetically ends in a perfect system of laws in which all regularities are coordinated and nothing continues to develop.

The French philosopher Henri Bergson, too, posited that mental principles are effective in evolution and that there exists a teleology of the development of what is living. In 1907, he wrote,

> A hereditary change […] which continues to accumulate […] to build up a more and more complex machine, must certainly be related to some sort of *effort*, but to an effort of far greater depth than the individual effort, far more independent of circumstances, an effort common to most representatives of the same species […]. So we come back […] to the idea we started from, that of an original impetus of life [*elan vital,* M.H.], passing from one generation of germs to the following generation of germs through the developed organisms which bridge the interval between the generations. This impetus […] is the fundamental cause of variations, at least of those that are regularly passed on, that accumulate and create new species. (Bergson 1907/1921, p. 88)

Here, unlike Peirce, the mental principle is not that of habit formation; rather, it is that of the *effort* involved as organisms develop into formations of an increasing level of adaptation and complexity as higher stages of development are reached.

Finally, Alfred North Whitehead's cosmology, which can be interpreted as panpsychist, must be mentioned within this context. With the help of evolutionary theory, Whitehead attempts to provide an answer for the question about the cause of the emergence of life in the first place. For this reason, he generalises the idea of the "survival of the fittest" to one that focuses on the being persisting the longest. According to Whitehead, if one compares the individuals and types of inorganic and organic existence, one must notice that inorganic forms of existence often persist longer than organic forms. He then asks why living beings had emerged in the first place, if development is only concerned with persistence. In 1927, he writes,

> I must […] join issue with the evolutionist fallacy suggested by the phrase "the survival of the fittest." The fallacy does not consist in believing that in the struggle for existence the fittest to survive eliminate the less fit […]. The fallacy is the belief that fitness for survival is identical with the best exemplification of the Art of Life. In fact life itself is comparatively deficient in survival value. The art of persistence is to be dead. Only inorganic things persist for great lengths of time. A rock survives for eight hundred million years; whereas the limit for a tree is about a thousand years, for a man or an elephant about fifty or one hundred years, for a dog about twelve years, for an insect about one year. The problem set by the doctrine of evolution is to explain how complex organisms with such deficient survival power ever evolved. (Whitehead 1927/1958, p. 2)

Similarly to Peirce and to the other founder of pragmatism, William James, Whitehead's own explanation of this fact consists in the postulation that the whole of reality consists of nothing other than experience. Yet experience entails an inherent effort to increase the *intensity* of experience (Whitehead 1929, III.II). This effort is realised in living beings at the expense of the duration of survival. Whilst a living being does not "survive" for as long as a rock, its experience is more intense than that of a rock. According to Whitehead, however, the latter is to be interpreted as a system of "dormant" units of experience, not unlike Leibniz' Monadology. The experience of living beings, including the conscious experience of man, is therefore nothing other than an experience that has grown to great intensity and can be found throughout nature.

Nevertheless, it must be emphasised that neither Peirce's cosmology nor the cosmologies of Henri Bergson or Alfred North Whitehead are justified by Darwinian evolutionary theory or what has become of it. Even though Darwin himself occasionally referred to a "higher development", such a form of development is not necessarily connected with his idea. Organisms do not always have to become increasingly complex. According to the interpretation that has been accepted within biology, evolution has no goal and aims, neither at increasing regularity or complexity, nor at intensifying experience. All of these teleologies have been *added* to the Darwinian concept of a history of living nature by the authors mentioned in order to structure their cosmologies. It may be possible to confirm these teleologies one day, but as for now, they are purely hypothetical.

In Germany, Ernst Haeckel and Gustav Theodor Fechner developed similar pan-psychic cosmologies to Peirce, Bergson and Whitehead, with Haeckel (1863) being the most influential advocate of Darwin's evolutionary theory on the continent. In his talk at the 38th Assembly of German Natural Scientists and Doctors in Stettin on 19 September 1863, he helped to draw a great amount of attention to Darwinism in Germany. In his "Monism", Haeckel (1899) generalised Darwinism by positing what he referred to in his *Riddle of the Universe* as the "Psychoplasm", which he viewed as the "material substrate" of all mental activities. These activities, he said, also occur in plants and lower animals. By contrast, it is the "neuroplasm" that is important for the mental life of man as well as for higher animals in general (Haeckel 1899). By adopting this position, Haeckel wanted to challenge psychophysical dualism, which assumed a mental realm of reality that was independent of the body. Even before Darwin's time, a similar theory was brought forward in Germany by Gustav Theodor Fechner a pupil of Lorenz Oken. Independently of Darwin's evolutionary theory, but influenced by Schelling's philosophy of nature, the physicist and philosopher remembered today primarily for his insights into psychophysics published a book about the souls of plants in (1848), as well as a pan-psychic cosmology in 1861. The continuity of the living forms as demonstrated by evolutionary theory provided empirical plausibility for the speculation found in all of these pan-psychic ideas: that the mental as an experience in an inner world characterises all forms of life. This speculation was deemed to "fill" the supposedly insurmountable gap between the material and the mental.

1.4 Dialectical Materialism and Pragmatist Process Philosophy

Similar to Peirce, Karl Marx's and Friedrich Engels's theory was influenced by the natural philosophy of German idealism and by Darwin. However, unlike Peirce and the aforementioned theorists following him, Marx and Engels did not develop a pan-psychist evolutionism. Instead, they believed that they had found evolutionary regularities which were responsible for making the development of the human realm as necessary and predictable as that of the other organisms: "Darwin has interested us in the history of Nature's Technology, i.e., in the formation of the organs of plants and animals, which organs serve as instruments of production for sustaining life. Does not the history of the productive organs of man, of organs that are the material basis of all social organisation, deserve equal attention?" (Marx and Engels 1996, p. 375).

Marx and Engels did however not yet speak of emergence within this context. They much rather believed that the development of life is a process that occurs in stages, taking qualitative leaps whenever a new stage is reached. From this moment on, the development is governed by new, but equally necessary laws. They affirmatively quote the following condensed exposition of this thought by *Capital* reviewer Kaufmann in a later edition of their main work:

> As soon as society has outlived a given period of development, and is passing over from one given stage to another, it begins to be subject also to other laws. In a word, economic life offers us a phenomenon analogous to the history of evolution in other branches of biology. The old economists misunderstood the nature of economic laws when they likened them to the laws of physics and chemistry. A more thorough analysis of phenomena shows that social organisms differ among themselves as fundamentally as plants or animals. […] The scientific value of such an inquiry lies in the disclosing of the special laws that regulate the origin, existence, development, death of a given social organism and its replacement by another and higher one. (Review of *Capital* by I.I. Kaufmann, as cited in Marx and Engels 1996, p. 18)

In the spirit of Marx and Engels, Kaufmann seems to understand the evolution of animals and plants both as a *nomological* matter as well as the emergence of something *new*. Before Darwin, laws of nature were viewed in physics and chemistry as regularities that described recurring sequences of motion in terms of their necessity (in the impact law of Cartesian kinematics, for instance) or the conservation of mass or energy (such as in the chemical law of the conservation of matter from original substances and the products of their reactions), whereas the Darwinian laws of coincidental variation of the progenies of a population and of natural selection represent laws of the *emergence of novelties*. It is this thought that Marx and Engels want to apply to the development of economies and societies in the human realm. Capitalism as a form of economy and the society accompanying it also constitutes a novelty in comparison to mercantilism and feudalism. But this novelty of economics and societal life has developed by necessity and will either continue to evolve by necessity or be "overcome" in a communist society, as Marx and Engels surmised. This generalisation of Darwinism and Hegelianism was then interpreted as evolutionary *dialectical materialism* by Stalin in 1938 and turned into the ideological basis of Soviet communism (Stalin 1938/1961). Here, mind emerges as something new during the "shift" from quantity (complexity) to a new quality of matter.

If the theory of evolution is regarded substantially as an explanation of the emergence of order, the question arises whether the choice between mind and matter is of any particular interest at all. In 1761, 100 years before Darwin, David Hume had already pointed out the following:

> For aught we can know a priori, matter may contain the source or spring of order originally within itself, as well as mind does; and there is no more difficulty in conceiving, that the several elements, from an internal unknown cause, may fall into the most exquisite arrangement, than to conceive that their ideas, in the great universal mind, from a like internal unknown cause, fall into that arrangement. […] Suppose […] that matter were thrown into

> any position, by a blind, unguided force; it is evident that this first position must […] be the most confused […] and so on through many successions of changes and revolutions […]. Thus the universe goes on for many ages in a continued succession of chaos and disorder. But is it not possible that it may settle […] so as not to lose its motion and active force […] so as to preserve an uniformity of appearance, amidst the continual motion and fluctuation of its parts? This we find to be the case with the universe at present. (Hume 1948, pp. 20, 54 ff.)

In saying that a material structure has been arranged by a mind, we hardly have an answer to the question of how the mind arrived at this arrangement. For Hume, the principle of self-organisation is the alternative to this predicament: Through some "blind [mentally] unguided force", a material constellation simply comes into being and is relatively stable even if it is unlikely to last forever. Do we need to comment any further on how order comes about? By Hume's lights, we do not. Whether we call what has thus organised itself (without reflectivity) "matter" or "mind" is not really of great importance. Most philosophical authors have therefore not only abandoned the speculation about an ultimate goal for the development of living beings or nature as a whole but also given up speculating about a "first principle" and an "origin" of this development. It is presumed that stable orders come about by coincidence at some point in material or mental processes and will one day dissolve again in an equally contingent manner.

Perhaps the most "mature" effect of the theory of evolution can therefore be found in the *awareness of contingency* that has developed more recently in Pragmatism. According to John Dewey, Darwin's theory of evolution produced an "intellectual face-about" from what is fixed and permanent in the logic of knowledge, in human attitudes towards morality and religion, and in politics (Dewey 1965, p. 2). A process view of the world and knowledge, which abandons the idea of ultimate certainty, has developed (cf. Dewey 1929). John Dewey's explanation of this process view of the world, in which everything comes and goes without being governed by a single principle or goal, is as follows:

> The Christian idea of this world and this life as a probation is a kind of distorted recognition of the situation; […] in truth anything which can exist at any place and at any time occurs subject to tests imposed upon it by surroundings, which are only in part compatible and reinforcing. These surroundings test its strength and measure its endurance. As we can discourse of change only in terms of velocity and acceleration which involve relations to other things, so assertion of the permanent and enduring is comparative. The stablest thing we can speak of is not free from conditions set to it by other things. That even the solid earth mountains, the emblems of constancy, appear and disappear like the clouds is an old theme of moralists and poets. […] It is no cause or source of events or processes [underlying all change, M.H.] no absolute monarch; no principle of explanation; no substance behind or underlying changes (Dewey 1925/1958, p. 70ff.).

1.5 Modern Emergentism

During the twentieth century, as one of the variations of non-reductive physicalism, emergentism develops into the most influential theory concerning the relation between matter and mind. This non-reductive materialism takes the place of nineteenth century evolutionary (or dialectical) materialism, as physics rises to become the ontologically most important discipline of all sciences for many philosophers of this period. The victory of atomism and the search for the ultimate particles of matter in particle physics have given support to elementarist conceptions in ontology. Together with the theory of evolution, this reinforced the conviction that reality "ultimately" consists of objects which are identified by physics, and that the evolution of nature leads to these basic physical objects permanently increasing in complexity. The consequence of this complexification of matter, which occurs as part of evolution, is that systems, which ultimately consist of elementary physical particles, possess properties that these particles themselves do not possess. While it is true that the system's properties are ultimately dependent on the existence and the properties of its parts (they "supervene" on them), they cannot be reduced to their parts, for example by definitions, and are "new" compared to these parts in the development of the respective system (they "emerge").

The debate on different forms of supervenience and emergence is rather technical and has meanwhile become rather extended (cf. Weber and Stephan as cited in Krohs and Toepfer 2005). The subject of interest lies in what is referred to as *diachronic structure emergentism* (following Weber 2005, p. 103). According to Weber, the novelty or unpredictability of living structures can arise for two reasons: because the material world is understood as fundamentally indeterministic in a quantum mechanics sense or, more interestingly here given that quantum phenomena are hardly likely to play a role in the macro area of living beings and their evolution, because developments of living beings follow the laws of *deterministic chaos* (cf. Smith 1998). In the latter case, the functional properties of organisms would not be "deducible" from their predecessors, as minor variations in the initial conditions of the stages of development observed can lead to extensive changes in later phases of evolution. Weber describes this form of unpredictability as follows:

> The structure of a newly arising system can be unpredictable for various reasons. The assumption of an indeterministic universe, for instance, immediately implies the unpredictability of newly forming structures. From the perspective of emergence theory, however, it would not be very interesting if the formation of a new structure only counted as unpredictable because its coming about is not determined. Besides, most emergentists assume that the formation of new structures also follows deterministic laws, anyway. But even formations of structure can be unpredictable in principle for us, namely when they follow laws which must be considered as appertaining to deterministic chaos. (Weber 2005, p. 100)

Especially in molecular evolution theory, many believe that developments are to be reconstructed as ones of a deterministic chaos (cf. Eigen 2000). According to the laws of deterministic chaos, minimally, maybe technically not even measurably, different initial values of analogous processes, which can be described by the same mathematical function, can lead to completely different and unpredictable material realisations (Metzler 1998). In organic development, we are undoubtedly constantly dealing with such minimal differences, which can have far-reaching and unforeseeable consequences. This occurs not only in genetics but also in the evolution of ecological systems within which organisms adapt. The large differences in anatomy and function of the organic, which can be traced back to such minimal differences in the initial conditions of developments, then lead to the phenomenon of emergence in the sense of unpredictables. At the same time, however, the emergent is seen as dependent on the material structures whose evolution is to be traced. Consequently, the radically new phenomena are not regarded as the result of radically new causal factors, but only as the result of a new relation between well-known causal factors that have come about in the course of evolutionary history. The dependence on these familiar causal factors or materials is also called “supervenience” in this context.

The most obvious example of the relation between emergence and supervenience is art. A painting cannot exemplify the aesthetic properties it possesses, for instance that of being “beautiful”, unless it possesses certain physical properties; even its aesthetic properties depend on its physical properties. Had it been painted in different, say, lighter or darker colours, the reflective properties of the material to which the paint was applied would be different, and so on. This, in turn, would alter the aesthetic properties of the painting. But an art critic is hardly going to investigate a painting with a luxmeter in order to determine the painting’s physical pattern of light reflexion. The critic’s judgment will rather be made relative to the tradition of painting, possibly taking into account the artistic development of the painter, and it will refer to properties such as beauty, ugliness, originality, etc., which do not appear in any physical theory. While there cannot be any objects of art without there being physical objects, the artistic properties are of course new, and aesthetics can thus not be reduced to physics.

A painting is a product that the artist (more or less) intended to produce and one might think that its emergent aesthetic properties derive from the artist’s creativity or the artistic criticism of the observer (“in the eyes of the beholder”). Applying the concept of emergence to the observation of living beings however refutes the supposition that it is mental activity or a reception idealism that endow physical systems with properties that cannot be described in physical theories.

Important points on the path towards the prevalence of this modern concept of emergence regarding the description of the development of new properties of biological systems, including the mental abilities of animals, were books published in the 1920s by Samuel Alexander (1920), Roy Wood Sellars (1922), C. Lloyd Morgan (1923) and C.D. Broad (1925). After first being criticised (e.g. by Hempel and Oppenheim 1948), the philosophical theories of the evolution of life and mind presented here gained a great deal of influence in the recent past, albeit in a new theoretical guise. The most impressive of these is Arno Ros’s synthetic materialism, which operates using mereological explications (Ros 2005, 2008) and successfully continues the tradition of non-reductive materialism and emergentism by means of a separate theory of justification as a “transition between conceptual fields” (Ros 1990, III, p. 298 f.). Unlike causal explications, which concern the change of an object due to the law-like impact of another object or event, mereological explications deal with the appearance of properties in a complex, internally structured object due to the relations among its parts (cf. Ros 2008). However, this is not the place to address all the details of this development from the 1920s until today, particularly in terms of the logical reconstruction of explications and theoretical reductions (in addition to Ros, cf. Beckermann et al. 1992).

Today, the evolution of living beings from their origins, i.e. starting from when life evolved out of inorganic matter, is regarded as an evolution of physicochemical systems (cf. Küppers 1983). From the formation of coacervates to that of the human being, these systems were constantly developing new properties. The first thing that comes to mind here is the property of self-preservation in an encapsulated membrane and the creation of an internal space through which energy flows (metabolism) as well as, in the case of animals, the abilities of perception and locomotion. The properties of living beings certainly rely on a physical basis and the properties of the mental on a biological basis: Without matter there can be no living beings, and without a nervous system, there can be no mind. Nevertheless, physical elementary particles do not possess the functional properties of metabolism and procreation possessed by living beings, and neurons do not possess the properties of intentionality that characterise mental states. Most biologists and philosophers believe that a change in the physical or biological basis of an individual entails a change in its biological or mental properties. At the same time, it is not possible to predict which combination of physical and biological systems in the history of evolution brings about certain kinds of functional or mental properties or abilities. In this sense, evolution can be regarded as new properties arising in an emphatic sense. The transitions from inanimate matter—which is not covered by a membrane and does not form internal spaces—to the first metabolising unicellular organisms, and from living beings—which do not yet intentionally turn towards an external world—to animals that are endowed with a mind and are capable of representing things, constitute the most important emergences of life and mind in the evolution of a nature that consists of purely material structures. They are each formations of new part-whole relations between structures of matter in which former wholes (e.g. cells) become elements of larger, more comprehensive wholes (e.g. multi-cell organisms). This formation of new part-whole relations is captured in conceptual transitions that were traditionally mistakenly interpreted as descriptions of the appearance of new causal factors (“life force” or the “mind”) (cf. Driesch’s concept of entelechy and the ensuing so-called problem of vitalism, Driesch 1909, II, p. 137).

It is in this context that the special position of natural kinds in the area of the development of life and in biology must be understood. Elementary physical particles and chemical elements have unchangeable properties. Even if they are imagined today as having come about in the development of the universe, they still only are what they are when they have a very specific mass, a very specific charge, or a very specific atomic weight. The charge of an electron is therefore also called a *natural constant,* and the atomic mass of elements such as hydrogen or gold cannot be changed. Biological kinds, however, change enormously over the course of the development of living beings; in fact, their ability to change is the very condition for the appearance of something new in the history of organisms.

This ability to change is the reason why the role of mathematics is much smaller in biology than it is in physics and chemistry, as there are no constant relations between magnitudes in which the gradable properties of organisms falling under a particular type could be stated and pinned down as essential characteristics. At best, very large numerical variations can be stated here, which is something that could never be realised in physics: Living beings ranging all the way from the microscopic size of a unicellular organism to a dinosaur or whale can possess comparable metabolic (citric acid cycle) and reproductive (mitosis and meiosis) competences;

a human being instantiates the genetic characteristics of what is human in a range of mass that lies between roughly 3 kg (birth weight) and 560 kg (a person by the name of Manuel Uribe, the heaviest man in the world according to the 2007 Guinness Book of Records). Within these variations, the relevant individuals still exemplify the *homo sapiens* species. Nevertheless, the fact that biology regards kinds as having evolved historically, pertaining to a kind cannot be used as a mark of an unchangeable essence. This is why, ever since the recognition of the historical variability of its objects, biology has ceased to be an essentialist science such as physics and chemistry continue to be; hence there will be no exact natural laws in biology (cf. Mayr 1982, p. 45; Weber 2005, p. 84). Biology is the only natural science so far that has replaced essentialist-nomological explanations with historical–genealogical observations. As the founder of the theory of evolution, Darwin also set the paradigms for the *development of methods* for an anti-essentialist philosophy and history (such as Nietzsche's and Foucault's genealogy) that are still applied today (cf. Sarasin 2009). Philosophies that move within these paradigms will consider historical evidence to be more important for knowledge about a permanently changing reality of concrete processes than evidence supplied by mathematics. Thus, despite the impossibility of an evolutionary epistemology, Darwin would appear to be the decisive alternative to the Platonist ideal of knowledge, which was guided by the idea of mathematical evidence of eternal abstract objects being the very ideal of knowledge.

References

Alexander, S. (1920). *Space, time and deity. The gifford lectures at Glasgow 1916–1918* (Vols. 1–2). London: Macmillan.

Beckermann, A., Flohr, H., & Kim, J. (Eds.). (1992). *Emergence or reduction? Essays on the prospects of nonreductive physicalism*. Berlin: Walter de Gruyter.

Bergson, H. (1907). *L'Evolution créatrice*. Paris: F. Alcan.

Bergson, H. (1907/1921). *Creative evolution* New York: Barnes & Noble.

Boltzmann, L. (1896). *Vorlesungen über Gastheorie*. Leipzig: Barth.

Bultmann, R. (1941). *Neues Testament und Mythologie*. Hamburg: Gütersloher.

Burgess, A. (1962). *A clockwork orange*. London: W. W. Norton & Company.

Broad, C. D. (1925). *The mind and its place in nature*. London: Routledge.

Campbell, D. T. (1974). Evolutionary epistemology. In P. A. Schlipp (Ed.), *The philosophy of Karl Popper* (Vol. 1, pp. 413–459). Illinois: The Open Court.

Caplan, A. (Ed.). (1978). *The sociobiology debate. Readings on the ethical and scientific issues concerning sociobiology*. New York: Harper Collins.

Clutton-B., T. H., & Harvey, P. H. (Eds.). (1978). *Readings in sociobiology*. San Fransisco: Freeman & Company.

Dawkins, R. (1978). *The selfish gene*. London: Oxford University Press.

Dawkins, R. (1986). *The blind watchmaker*. London: W. W. Norton & Company.

Dennett, D. (1995). *Darwin's dangerous idea: Evolution and the meaning of life*. New York: Simon & Schuster.

Dewey, J. (1929). *The quest for certainty. A study of the relation of knowledge and action*. New York: Minton, Balch.

Dewey, J. (1925/1958). *Experience and nature*. Chicago: Open Court.

Dewey, J. (1965). *The influence of darwinism on philosophy and other essays in contemporary thought*. Bloomington: Indiana University Press.

Driesch, H. (1909). *Philosophie des Organischen. Gifford-Vorlesungen, gehalten an der Universität Aberdeen in den Jahren 1907–1909*. Leipzig: W. Engelmann.

Eigen, M. (2000). *Stufen zum Leben. Die frühe Evolution im Visier der Molekularbiologie*. Munich: Piper.

Esposito, J. L. (1980). *Evolutionary metaphysics: The development of Peirce's theory of categories*. Ohio: Ohio University Press.

Farber, P. L. (1998). *The temptations of evolutionary ethics*. Los Angeles: University of California Press.

Fechner, G. T. (1848). *Nanna oder über das Seelenleben der Pflanzen*. Leipzig: Voss.

Fechner, G. T. (1861). *Über die Seelenfrage. Ein Gang durch die sichtbare Welt um die unsichtbare zu finden*. Leipzig: Voss.

Fenk, A. (Ed.). (1990). *Evolution und Selbstbezug des Erkennens*. Wien: Boehlau.

Flew, A. (1978). From is to ought. In Caplan, A. (Ed.), *The sociobiology debate. Readings on the ethical and scientific issues concerning sociobiology* (pp. 142–162). New York: Harper Collins.

Gilson, E. (1971). *From Aristotle to Darwin and back again. A journey in final causality, species and evolution*. London: Sheed & Ward.

Gräfarth, B. (1997). *Evolutionäre Ethik? Philosophische Programme, Probleme und Perspektiven der Soziobiologie*. Berlin: Gruyter.

Haeckel, E. (1863). Über die Entwicklungstheorie Darwins. Öffentlicher Vortrag in der Allgemeinen Versammlung deutscher Naturforscher und Ärzte zu Stettin, am 19. September 1862 (Amtlicher Bericht über die 37. Versammlung, 17).

Haeckel, E. (1899). *Die Welträthsel. Gemeinverständliche Studien über Monistische Philosophie*. Bonn: Strauss.

Hampe, M. (1999). Komplementarität und Konkordanz von Natur und Erkenntnis. Anmerkungen zu Schelling und Peirce. In H. Eidam, F. Hermenau, & E. Stedderoth (Eds.), *Kritik und Praxis. Zur Problematik menschlicher Emanzipation* (pp. 96–106). Lüneburg: zu Klampen.

Hampe, M. (2006). *Erkenntnis und Praxis. Zur Philosophie des Pragmatismus*. Frankfurt a. M.: Suhrkamp.

Hampe, M. (2007). *Eine kleine Geschichte des Naturgesetzbegriffs*. Frankfurt a. M.: Suhrkamp.

Hempel, C. G., & Oppenheim, P. (1948). Studies in the logic of explanataion. *Philosophy of Science*, *15*, 135–175.

Hume, D. (1948). *Dialogues concerning natural religion*. London: Hafner.

Hume, D. (1978). *A treatise of human nature*. Oxford: Oxford University Press.

Huxley, T. H. (1989). *Evolution and ethics*. Princeton: Princeton University Press.

King's College Sociobiology Group. (Ed.). (1982). *Current problems in sociobiology*. Cambridge: Cambridge University Press.

Krohs, U., & Toepfer, G. (Eds.) (2005). *Philosophie der Biologie: Eine Einführung*. Frankfurt a. M.: Suhrkamp.

Küppers, B.-O. (1983). *Molecular theory of evolution. Outline of a physico-chemical theory of the origin of life*. Berlin: Springer.

Lloyd Morgan, C. (1923). *Emergent Evolution*. London: Methuen.

Lorenz, K. (1973). *Die Rückseite des Spiegels. Der Abbau des Menschlichen*. Munich: Piper.

Lütterfelds, W. (1987). *Transzendentale oder evolutionäre Erkenntnistheorie*? Darmstadt: Wissenschaftliche Buchgesellschaft.

Lütterfelds, W. (1993). *Evolutionäre Ethik zwischen Naturalismus Idealismus*. Darmstadt: Wissenschaftliche Buchgesellschaft.

Madden, E. H. (1958). *The philosophical writing of chauncey wright*. New York: Liberal Arts.

Marx, K., & Engels, F (1996). *Collected works* (Vol. 35). London: International.

Mayr, E. (1982). *The growth of biological thought*. Cambridge: Harvard University Press.

Menand, L. (2001). *The Metaphysical Club,* New York: Farrar, Straus and Giroux.

Metzler, W. (1998). *Nichtlineare Dynamik und Chaos*. Stuttgart: Teubner.

Meyer, H. (2000). *Traditionelle und Evolutionäre Erkenntnistheorie*. Hildesheim: Georg Olms.

Peirce, C. S. (1991a). *Naturordnung und Zeichenprozess: Schriften über Semiotik und Naturphilosophie*. Frankfurt a. M.: Suhrkamp.

Peirce, C. S. (1991b). *Schriften zum Pragmatismus und Pragmatizismus*. Frankfurt a. M.: Suhrkamp.

Peirce, C. S. (1992). *Reasoning and the Logic of Things,* Harvard University Press.

Pinker, S. (1997). *How the mind works*. London: W. W. Norton & Company.

Pinker, S. (2002). *The blank slate. The modern denial of human nature*. New York: Penguin.

Popper, K. (1973). *Objektive Erkenntnis. Ein evolutionärer Entwurf.* Hamburg: Hoffmann und Campe.

Popper, K. (1994). *Alles Leben ist Problemlösen: Über Erkenntnis, Geschichte und Politik*. Munich: Piper.

Randall, D. S. (1983). *Evolution. From molecules to men*. Cambridge: Cambridge University Press.

Richards, R. J. (1993). Birth, death and resurrection of evolutionary ethics. In M. H. Nitecki & D. V. Nitecki (Eds.), *Evolutionary ethics* (pp. 113–139). New York: SUNY.

Riedl, R., & Delpos, M. (Eds.). (1996). *Die evolutionäre Erkenntnistheorie im Spiegel der Wissenschaften*. Vienna: WUV.

Ros, A. (1990). *Begründung und Begriff. Wandlungen des Verständnisses begrifflicher Argumentation* (Vols. 1–3). Hamburg: Meiner.

Ros, A. (2005). *Materie und Geist. Eine philosophische Untersuchung*. Paderborn: Mentis.

Ros, A. (2008). Mentale Verursachung und mereologische Erklärungen. Eine einfache Lösung für ein komplexes Problem. *Deutsche Zeitschrift für Philosophie, 56,* 167–203.

Sarasin, P. (2009). *Darwin und Foucault. Genealogie und Geschichte im Zeitalter der Biologie*. Frankfurt a. M.: Suhrkamp.

Schelling, F. W. J. (1800/1978). *System of transcendental idealism*. Virginia: University of Virginia Press.

Schrader, C. (2007). *Darwins Werk und Gottes Beitrag: Evolutionstheorie und Intelligent Design*. Stuttgart: Kreuz.

Sellars, R. W. (1922). *Evolutionary naturalism*. Chicago: Open Court.

Smith, P. (1998). *Explaining chaos*. Cambridge: Cambridge University Press.

Spencer, H. (1893). *Principles of ethics*. New York: Harper & Brothers.

Spencer, H. (1860/1971). *First principles*. London: Williams and Norgate.

Spinoza, B. de (1994). *Theologisch-politischer Traktat*. Hamburg: Meiner.

Stalin, J. (1938/1961). *Über dialektischen und historischen Materialismus*. Berlin: Moritz Diesterweg.

Vollmer, G. (1975). *Evolutionäre Erkenntnistheorie*. Stuttgart: Hirzel.

Weber, M. (2005). Supervenienz und Physikalismus. In U. Krohs & G. Toepfer (Eds.), *Philosophie der Biologie* (pp. 71–87). Frankfurt a. M.: Suhrkamp.

Whitehead, A. N. (1927/1958). *The function of reason*. Princeton: Nabu Press.

Whitehead, A. N. (1929/1973). *Process and reality*. Cambridge: Simon & Schuster.

Wilson, E. O. (1975). Sociobiology. *A New Synthesis,* Harvard University Press.

Wittgenstein, L. (1966). *Lectures and conversations on aesthetics, psychology and religious belief.* Oxford: Oxford University Press.

Chapter 2
The Broad Foundations of Adaptationist-Computational Evolutionary Psychology

Malte H. Dahlgrün

2.1 What Evolutionary Psychology Is

Sociological Introduction What is known as *evolutionary psychology* is a broad scientific paradigm and disciplinary synthesis that is little more than a quarter of a century old as this chapter is written. Its beginnings are probably best identified with the publication of the first of a string of programmatic papers by Cosmides and Tooby (1987), as well as a paper by Symons (1987).[1] The scientist couple Leda Cosmides and John Tooby, David Buss, and Steven Pinker, have been the leading first-generation theorists of evolutionary psychology and continue to play central roles. Buss, Cosmides and Tooby, as well as Martin Daly and Margo Wilson, produced some of the best-known empirical results of the first years of evolutionary psychology.[2] In the course of the 1990s and 2000s, PhD students of these scientists developed various specializations and dispersed widely, geographically

---

This chapter is dedicated to the memory of my mother.

---

[1] If, as seems sensible enough, we choose to identify a first wave of major works in evolutionary psychology that were published by 1990, these would also include Daly and Wilson (1988), Buss (1989), Cosmides (1989), Pinker and Bloom (1990), Tooby and Cosmides (1989, 1990a, b), Buss and Barnes (1986), and Crawford et al. (1987). Evolutionary psychology was foreshadowed in the areas of human sexuality by the books of Symons (1979) and Daly and Wilson (1983/1978, Chaps. 11–12). The famous edited collection *The Adapted Mind* (Barkow et al. 1992), and the book-length treatise by Tooby and Cosmides (1992) it contains, are sometimes viewed by external observers as founding documents of the field. But evolutionary psychology was already well underway before this volume was published and considerable portions of the more significant material in it had already been published elsewhere.

[2] Valuable autobiographical remarks on early meetings of these scientists are provided by Buss (2008/1999, p. xvi f.) or Tooby and Cosmides (2005, p. 15, n. 3).

M. H. Dahlgrün (✉)
Department of Philosophy, Humboldt University of Berlin, Berlin, Germany
e-mail: dahlgrun@gmail.com

T. Breyer (ed.), *Epistemological Dimensions of Evolutionary Psychology,*
DOI 10.1007/978-1-4939-1387-9_2

speaking, with a number of them becoming influential contributors to the field in their own right. Many leading researchers in evolutionary psychology—from long-established senior figures to the youngest up-and-coming—are neither identical to any of the above nor former students of them. Many have their own theoretical emphases within the broad shared framework I wish to articulate in this chapter, yet they self-identify as 'evolutionary psychologists' and are regarded within the field as contributing to the same scientific enterprise. Evolutionary psychology is a quite rapidly developing field, it is bigger than ever before, and it continually attracts new and open-minded students who come to evolutionary psychology with the mindset that pursuing experimental work in this field is a perfectly natural thing to do.[3]

The reason I make these sociological observations is that they run contrary to an old, negative stereotype about evolutionary psychology. According to this influential stereotype, evolutionary psychology consists in the activities of a small coterie of researchers, or perhaps in activities controlled by such a coterie. Reflections of this stereotype are found in labels such as 'Santa Barbara school', based on the fact that Tooby, Cosmides, and Symons conducted their research at the University of California at Santa Barbara for many years. In fact, evolutionary psychology is not only a large field, but also a theoretically and methodically diverse one.

Aims of this Chapter Mapping the intra-field diversity of evolutionary psychology would be a formidable task worth executing systematically in a different work. Relatedly, it would be useful to compile some widespread misconceptions—short of out-and-out straw men—about what defines evolutionary psychology. These misconceptions would include misattributions of the following putatively defining characteristics: ideas that are merely typical of the field, derivative tenets, and inessentially strong emphases only endorsed by specific evolutionary psychologists. But this is beyond the scope of the present work.

In this chapter, I will try to state carefully what evolutionary psychology is positively *committed to,* as a field. I will not say very much about what it is not committed to, and I will say even less about specific criticisms of the field. Next to articulating evolutionary psychology's basic commitments, the central theme here will be that these commitments are considerably *broader than usually assumed outside the field*, although the commitments are certainly not trivial. Misconceptions about evolutionary psychology, for the main part, enter into the present discussion only via allegations of narrowness. Though I will not begin to map the field's theoretical diversity in this chapter, the fact that there is room for substantive differences of theoretical emphasis within evolutionary psychology will hopefully become clear by implication of which commitments are stated here, and which are *not* stated.

---

[3] It is interesting to note in this context that, while philosophers of special sciences generally try to be closely in touch with the activities of the sciences they specialize in, this norm does not seem to be widely observed when it comes to the human evolutionary behavioural sciences. There is a relative dearth of interaction between philosophers and the research communities in this broad field.

When it comes to evolutionary psychology, foundational clarifications and replies to fundamental criticisms are peculiarly closely connected. Erroneously narrow and ill-informed conceptions of what evolutionary psychology is committed to are so pervasive, even among the less than hostile, that mere clarification by itself may constitute evolutionary psychology's most important avenue of defence.

The Essence of Evolutionary Psychology Evolutionary psychology can be viewed as combining *two ways of assimilating the study of the mind to the natural sciences*. It does so by combining two broad and standard research paradigms from psychology and evolutionary biology: (1) a broadly *computational* (information-processing) approach to the mind, and (2) an *adaptationist* (or, roughly equivalently: selectionist) approach to the accordingly conceived mind's architecture, where this adaptationism is based on the modern evolutionary theory of natural selection rather than pretheoretical, intuitive notions of function. Against this background, the research program generally known as evolutionary psychology might more accurately be called *adaptationist-computational evolutionary psychology*. This is how I understand the label 'evolutionary psychology' in what follows. Still more fully stated, the nature of evolutionary psychology is this:

> **(EP)** Evolutionary psychology is experimental psychology in a broadly computational framework, organized, unified, and heuristically guided by adaptationist nativism based on modern evolutionary theory.

That is all there is to it. No stronger commitments than those in (EP) unite the scientists who unhesitatingly identify themselves as evolutionary psychologists. Nothing more specific underlies their use of the terms 'evolutionary psychologist' or 'evolutionary psychology'. But nothing less will do either. The present chapter aims to throw into sharper relief these basic commitments of adaptationist-computational evolutionary psychology, explaining them in some detail. For now, let me anticipate the following.

The broadly *computationalist* aspect, here as elsewhere, consists in positing mechanically interacting internal representations in the explanation of behaviour that are located at a functional level of description beyond that of common-sense psychology—or at least in requiring one's work to naturally lend itself to such a kind of explanation. Evolutionary psychology's *adaptationism* is based on the modern theory of natural selection and can be factored into a basic, ontological component, and an additional, disciplinary component. The disciplinary adaptationism of evolutionary psychology makes claims about how psychology is to be properly pursued, given that the ontological hypothesis is true. Ontological adaptationism hypothesizes that the human mind develops on the basis of many more naturally selected, content-specific acquisition mechanisms (and other information-processing adaptations) than have been traditionally assumed in mainstream psychology. This is the central idea of evolutionary psychology. It will be explained later what this means.

Ontological adaptationism has not been spelled out in (EP) for the sake of readability. But it is entailed by the inclusion of disciplinary adaptationism. This in turn is represented in (EP) through the requirements that psychology be organized, unified, and heuristically guided by adaptationist theory.

Psychological nativism is essential to evolutionary psychology if anything is, but it is an inextricable part of the adaptationist commitment. Evolutionary psychology makes no commitment to nativism independently of its adaptationist stance. This is why it is not treated as a separate pillar of evolutionary psychology here. Nonetheless, I will explain nativism in the context of evolutionary psychology's adaptationist commitments.

The Tooby-Cosmides Emphases[4] If, in the following, repeated reference is made to works written in the 1990s by first-generation promoters of evolutionary psychology, this is not due to a lack of awareness of subsequent publications and developments. It is because my present interest lies in identifying and clarifying the fundamental commitments of evolutionary psychology and because, since its beginnings, no *new* potential tenets that might play this role have emerged. This is not to say that, conversely, *all* of the potentially defining ideas which have been associated with evolutionary psychology since its beginnings have really turned out to be defining ones. In fact, that is not the case.

In particular, since the early days of evolutionary psychology, Tooby and Cosmides have tended to advocate adaptationist-computational evolutionary psychology in a package together with a bundle of special emphases which are inessential to evolutionary psychology as such. Many card-carrying evolutionary psychologists do not share these particular emphases to the same degree, and many endorse other emphases instead. (There is no clearer testament to this than Miller (2000), a leading, almost paradigmatic evolutionary psychologist who I believe actually goes too far in his purported rejection of elements of the Tooby-Cosmides 'metatheory', as he calls it.)

What is unfortunate in this context is the following. Firstly, Cosmides and Tooby have not bothered to separate their special emphases from the commitments that constitute evolutionary psychology as such—a field they have always supported and been part of. Secondly, even though methodologically conscious evolutionary psychologists must have been long aware of the difference between their field's essential commitments and the Tooby-Cosmides emphases, even influential figures have only rarely commented on it explicitly in writing. Thirdly, by common consent, Cosmides and Tooby are co-founders of evolutionary psychology. This combination of facts has been deeply misleading. It makes it hard to blame outsiders one-sidedly for identifying evolutionary psychology tout court with the more specific package of theorems promoted by Cosmides and Tooby.[5]

Why have evolutionary psychologists rarely stated in print that the Tooby-Cosmides emphases are far from universally shared in the community of self-identified evolutionary psychologists which Cosmides and Tooby themselves feel part of?

---

[4] The order of names in this descriptor is arbitrary with regard to the authors' contributions; it is chosen purely for rhythm and rhyme.

[5] This is not to lay the entire blame for the present type of misconception on Cosmides, Tooby, and other evolutionary psychologists. Philosophers of science who publish on evolutionary psychology, for example, can fairly be expected to dig deep enough to know better—but as far as I can discern, they frequently have not.

Mainly for strategic reasons, surely. One publicly deemphasizes differing theoretical hunches and emphases when the shared foundation is more substantive and one's research enterprise is still continuously challenged (Kurzban 2010a) to justify its very existence. Detractors of evolutionary psychology may balk at the comparison, but the situation seems at least sociologically reminiscent of how disagreeing evolutionary scientists publicly close ranks in the face of creationism (cf. Dawkins 2003).

2.2 On a Broad/Narrow Distinction Regarding 'Evolutionary Psychology'

It is popular nowadays, especially among authors critical of standardly so-called evolutionary psychology, to draw a broad/narrow distinction between two ways of understanding the compound noun 'evolutionary psychology' (Scher and Rauscher 2003a; Rauscher and Scher 2003a; Heyes 2000; Wilson 2003; Buller 2005; Dunbar and L. Barrett 2007; Heyes and Huber 2000: passim). The distinction and its rationale can be roughly summarized as follows. On a broad and natural understanding, 'evolutionary psychology' might mean something like:[6] psychology that is informed by evolutionary theory and the mind's evolutionary past. But Tooby, Cosmides, Buss, and colleagues established the term 'evolutionary psychology' with a far more specific and programmatically tendentious meaning than the generic meaning naturally derivable from the component meanings of this term. Consequently, it is desirable to restore an innocuous, generic meaning to the expression 'evolutionary psychology'. Sometimes, a separate label is suggested (or a special notational convention such as capitalization, see Buller 2005) to distinguish evolutionary psychology as a specific school of thought from the broad field.

An example of this way of framing things is provided by a 2003 volume entitled *Evolutionary Psychology: Alternative Approaches*, which contains various noteworthy contributions by well-known authors. Its editors Scher and Rauscher insert the following footnote towards the beginning of each chapter:

> EDITOR'S NOTE: In this book, the term 'narrow evolutionary psychology' signifies the approach to evolutionary psychology developed by Cosmides, Tooby, Buss, et al. This term was chosen not to imply that this approach has an inappropriately narrow point of view, but merely to suggest that the approach adopts a narrower range of assumptions than 'broad evolutionary psychology' (or, just 'evolutionary psychology'). This latter term signifies evolutionary psychology generally, practiced with any of a very broad range of assumptions possible within the general framework of evolutionary approaches to psychology.

---

[6] I am trying to be charitable here in suggesting how 'evolutionary psychology' might plausibly by understood by those without prior knowledge or expectations.

Incidentally, it requires no small amount of mental effort to understand the label 'narrow evolutionary psychology'[7] to be as void of suggestions of parochiality as the editors assure their readers. After all, the value-free logico-philosophical term that would be customarily used in such a context is 'strong', not 'narrow'—and evolutionary psychologists who would happily introduce themselves to others as 'narrow evolutionary psychologists' are presumably yet to be found. Be this as it may, I will not adhere to a twofold terminological distinction of the kind just described. My previously introduced usage implies this already. Before I explain this, a couple of terminological clarifications are in order.

It is true, and certainly evident from characterizations like (EP), that the term 'evolutionary psychology' has a significantly more specific meaning than one that might be determined by a competent speaker of English simply by understanding its component expressions. Where misunderstanding threatens, we can and should use the more accurate term 'adaptationist-computational evolutionary psychology'—or simply 'adaptationist-computational psychology', given that the modifier 'evolutionary' is strictly speaking rendered redundant by 'adaptationist'. Or, as I chose to do earlier, we can expressly stipulate to use 'evolutionary psychology' with the meaning of 'adaptationist-computational evolutionary psychology' and then freely avail ourselves of the shorter, less cumbersome term, so entrenched in the scientific community already.

But here is why I think we should reject a terminological revisionism along the 'broad/narrow' lines above. Consider the assumptions it is based upon. Firstly, such revisionism takes for granted that the term 'evolutionary psychology' as such has a clear, broad, innocuous meaning, on which it refers to a doctrinally uncommitted field of research. This is the positive assumption about 'broad' evolutionary psychology. Secondly, this terminological revisionism assumes that the broad and natural meaning needs to be freed from the stranglehold of a rigid and narrowly doctrinally committed school of thought that Cosmides, Tooby, Buss, et al. established under the name 'evolutionary psychology'. This is the negative assumption about 'narrow' evolutionary psychology. Let me comment on these in reverse order.

Rejecting the negative ('narrow') part of the conception underlying the broad/narrow distinction is part of the point of this chapter. Judging from the characterizations I have come across, whenever a broad/narrow distinction along the lines above is made, the conception of 'narrow' evolutionary psychology assumed as a foil is *practically always too specific to reflect actual understandings of 'evolutionary psychology' by self-identified evolutionary psychologists*, including paradigmatic representatives. It is likewise too specific to do any justice to the boundaries of the scientific community to which paradigmatic, self-identified evolutionary psychologists belong—that is, whom they associate with, whom they collaborate with, and whom they constructively discuss their ideas with. What would be problematic is if avowed evolutionary psychologists really did understand the term 'evolutionary

---

[7] See also the editors' note in D.S. Wilson (2003, n. 1), who is described as having independently arrived at the same terminology—apparently also with no intention of insinuating narrow-mindedness.

psychology' as implying the endorsement of each tenet regularly attributed to them; or even, arguably, if they understood it as implying endorsement of every single theoretical emphasis pushed by Cosmides and Tooby. But this is not the case. Rather, a large scientific community of self-identified evolutionary psychologists exists who, whatever else they may believe, share a commitment to pursuing psychology roughly along the lines of (EP).

As for the positive ('broad') part of the conception underlying the broad/narrow distinction, a full criticism of it would involve commenting on the exact justifications offered by different authors for assuming the existence of a broader, untendentious field of evolutionary psychology. Suffice it to sketch the following general response here.

Once one avoids attributing inappropriately strong ideas to standardly so-called, self-described evolutionary psychologists, the question becomes pressing as to how an allegedly 'broad' evolutionary psychology is supposed to be more inclusive than standard evolutionary psychology at all. For example, it cannot be the potential inclusion of animal behaviour research and cross-species comparisons in one's considerations (*pace* Heyes 2000, p. 3) that sets 'broad' evolutionary psychology apart from standardly so-called evolutionary psychologists (see, e.g., co-founders of the field, Daly and Wilson 1995, 1999, 2005). Or, to mention another central aspect, it is unlikely that a refusal to commit to a computational, information-processing approach to psychology, understood along the broad lines explained below, motivates insistence on a 'broad' evolutionary psychology.

Fundamentally, I submit, there really is only one issue relative to which an interesting alternative evolutionary psychology might be distinguished from the standardly so-called evolutionary psychology that is the subject of this chapter. It is the issue of *adaptationist nativism*. One might distinguish from standardly so-called evolutionary psychology a *non-adaptationist* evolutionary psychology whose adherents are much more sceptical about the number of human-specific psychological adaptations which natural selection may have shaped, and whose adherents are much more optimistic about how many human psychological phenomena can be explained through learning from experience—including social learning and cultural transmission—rather than through specific innate learning adaptations. A distinction of schools of thought in recent decades along these lines might make sense.

However, the crucial point for present purposes is this. Such an alternative, non-adaptationist evolutionary psychology *cannot impartially be claimed to form a broader, more innocuous scientific undertaking*. To assume this without argument would be to beg the very question that separates such a non-adaptationist evolutionary psychology from its adaptationist counterpart. A non-adaptationist, alternative evolutionary psychology is not a broader, more modest, more open-minded research program—at least not ipso facto. It is rather a research program which places its bets differently: much nearer to the empiricist end on a continuum spanning between adaptationist nativism and empiricism in human psychology.

Having clarified this, let me now return to standardly so-called evolutionary psychology, the adaptationist-computational psychology characterized in (EP).

2.3 On Fundamental Opposition and Methodological Misgivings

Fundamental opposition and disdainful comments on evolutionary psychology are still remarkably widespread in the philosophy of science and large parts of the cognitive or biological sciences, to say nothing of the social sciences. A frank disclosure therefore seems in order. The present chapter is written by ‘a sympathetic observer of the scene’, to borrow a self-description used in the same context by Richard Dawkins (2005, p. 975). After years of observing discourse in and on evolutionary psychology and studying relevant works from a thoroughly uncommitted standpoint, I have found myself agreeing with the assessment that evolutionary psychology is ‘subject to a level of implacable hostility which seems far out of proportion to anything sober reason […] might sanction’ (ibid.).[8] Evolutionary psychology’s central rationale is sound, I take it: The human mind/brain can be expected to show numerous effects of natural selection, no differently from the extracerebral body in general.

Attacks on the basic legitimacy of evolutionary psychology tend to be disconcertingly ill-informed about their target. This has often been pointed out in defences of evolutionary psychology, and it continues to be the case. Fundamental criticisms continue to rely more on prejudices, superficial misattributions, and popular vulgarizations than on attentive readings of the works that thoughtful evolutionary psychologists have actually written. Readers wishing to form an opinion on critiques of evolutionary psychology are referred to the works of Kenrick (1995), Alcock (2001), Pinker (2002 and 1997: passim), Kurzban (2002), Hagen (2005), Dawkins (2005), Tybur et al. (2007), Tooby and Cosmides (1992, pp. 33–49 and passim);[9] more indirectly but foundationally: Ketelaar and Ellis (2000), Barrett (2006), Barrett and Kurzban (2006); and perhaps most instructively, in response to specific high-profile attacks: Tooby et al. (2003), Buss and Reeve (2003), Daly and Wilson (2008), Holcomb (2005) or Machery and Barrett (2006).

My aim in this chapter is not to respond directly and in detail to specific criticisms of evolutionary psychology, as many of the aforementioned works commend-

---

[8] I have omitted from the Dawkins quote the added remark, ‘or even common politeness’, because I wish to focus on content here. Moreover, matters of politeness are hardly the problem in this context independently of content. Those openly hostile towards evolutionary psychology believe that their sometimes extraordinary reactions are justified by crass and pernicious ideas promoted by evolutionary psychologists. If I were to address any ethical issues here, it would be the lack of epistemic humility that is evidenced by critics commenting snidely or offensively on an entire field they often know remarkably little about.

[9] In several respects, the genre to which the works cited here respond could be extended back at least into the 1980s, that is, into the so-called sociobiology controversy (see Segerstråle 2000; Dawkins 1985). Indeed, some of the above-cited works address attacks on human sociobiology and evolutionary psychology alike, and there is very considerable overlap between the charges these two have had to face (just-so storytelling, genetic determinism, political misgivings, etc.). However, I am assuming that, for the familiar reasons first advanced in early evolutionary psychology, much of human sociobiology was fundamentally flawed in a way that evolutionary psychology was not (Cosmides and Tooby 1987, pp. 277–283; Symons 1987, 1989; cf. Laland and Brown 2011/2002, Chap. 3). Accordingly, the responses referenced above were chosen specifically with evolutionary psychology in mind.

ably do. My main aim is rather to provide a positive characterization of evolutionary psychology, and to separate it from inessential or spurious ideas—and to do this clearly enough to make the identification of straw man attacks a relatively straightforward affair, rendering superfluous long lists of further disclaimers as to all the things that evolutionary psychology is not.

What I want to do in the remainder of the present section is to address three *methodological types of misgivings* about evolutionary psychology. These concern: (1) the quality of specific works in the field, (2) the scope of topics addressed in evolutionary psychology, and (3) the potentially myopic scope of methods employed therein.

Specific Studies Apart from misconceptions about the aims and claims of evolutionary psychology, criticisms often reduce to dismissive remarks about the quality or interpretation of *specific studies* which a critic has come across. Presumably, no good evolutionary psychologist would deny that a great amount of flawed studies, inferences, and claims have been put forth by self-identified evolutionary psychologists in the past. But evolutionary psychology hardly fares worse in this respect than other behavioural fields of inquiry, and evolutionary psychology is still in its childhood. The two or three most important replies that I would make to critics of specific studies are these.

Firstly, surely one ought to judge a scientific paradigm by its *better* work, by work that realizes the paradigm's potential, rather than by the worst of it. Anyone criticizing evolutionary psychology on the basis of specific studies ought to consider whether they are observing this point, or whether they are measuring evolutionary psychology by unfair standards by which they would not measure other research paradigms. Relatedly, what would be unsettling is if shoddy studies were nowadays regularly found among the celebrated and most-cited writings within the evolutionary psychology literature. But that can hardly be claimed to be the case. The scientific community's attention, its awards, the tenured professorships at top-tier research universities, the establishment of citation classics, the publications in high-impact journals: these do very much tend to go to authors of strong rather than poor works in evolutionary psychology. And if the rationale of evolutionary psychology is correct, the good work it brings forth offers more epistemic progress per study than comparable work in corresponding areas of non-evolutionary psychology.

Finally, banal as it may seem to point this out, evolutionary psychologists do not claim single studies or experiments to establish hypotheses about the existence of a psychological adaptation. It is many pieces of evidence together, plus the lack of a satisfactory alternative empiricist explanation of relevant phenomena, that jointly and in the long run make the postulation of specific psychological adaptations plausible.

Topical Emphases People new to the field frequently get the impression that certain topical emphases—above all, sex and mating, but also parenting, social exchange, etc.—and a concomitant marginalization of other topics are somehow central to the identity of evolutionary psychology. This impression is incorrect. The adaptationist approach defined in (EP) does not mandate a topical scope restriction.

On the contrary, it mandates its absence. Such a scope restriction is incompatible with evolutionary psychology's aspiration of unifying psychological research.

It is, however, quite understandable how one might get the incorrect impression that specific topical foci are central to evolutionary psychology's identity. The efforts of the evolutionary psychology research community are still unevenly distributed across the range of phenomena constituted by the human mind and its behavioural products. Similarly, the limited scope of topics covered in the field's leading textbook (Buss 2008/1999) and the relative allotment of space to the topics it does cover are bound to confuse in light of the book title's unqualified reference to 'evolutionary psychology' and the subtitle's promise to describe 'the new science of the mind' tout court.[10] For an early tome in the field that is more evenly dedicated to a broad scope of themes and mental phenomena, compare Pinker (1997).

Methods and Disciplines Sometimes, a heavy reliance on certain methods or disciplines (the difference between these is not clear-cut) is portrayed as constitutive of the identity of evolutionary psychology. It is probably true that certain methods, typically convenient and inexpensive, still predominate in the field. But it is questionable whether the situation here is any worse than in realms of experimental psychology unguided by evolutionary thinking (cf. Baumeister et al. 2007). More importantly, evolutionary psychologists' choice of methods is not subject to any *programmatically driven* restrictions. On the contrary, good evolutionary psychologists welcome any method, if it helps discover the evolved architecture of the mechanisms of the human mind. Methodological overviews by evolutionary psychologists include a broad range of methods and data sources. They encourage their use. They sometimes even admonish them (Schmitt and Pilcher 2004; Schmitt 2008; Simpson and Campbell 2005; Buss 2008/1999, pp. 59–69, 42–49).

It is, in particular, provably wrong that evolutionary psychologists shun fields such as game theory, or behavioural genetics, or developmental psychology, or even neuroscience. Evolutionary psychologists are at work in all these fields. To claim otherwise is only to prove oneself ignorant of the literature. Not every *individual* evolutionary psychologist draws on the knowledge and methods of many different disciplines. But *between* them, evolutionary psychologists do aim to do so. That at least is the regulative ideal. Evolutionary psychology is programmatically keyed to *heavy disciplinary eclecticism*. Pinker, for one, has always been enthusiastic about this point, as in the following passage on what is required by good adaptationist explanations in psychology:[11]

> Unfortunately for those who think that the departments in a university reflect meaningful divisions of knowledge, it means that psychologists have to look outside psychology if they

---

[10] See D.S. Wilson's (1999) review of the textbook's first edition—a highly differentiated and competent critique published in the leading journal of the human evolutionary behavioural sciences. Wilson's remarks on topical partiality reflect the same impression which I had upon first encountering the textbook.

[11] See also Pinker (2005b, p. xiv f.) and the entertaining remarks in Pinker (2005a, p. 19). The presently considered claim goes beyond what Cosmides et al. (1992, p. 3 and 14, n. 1) call 'conceptual integration', which imposes the weaker requirement of cross-disciplinary consistency.

> want to explain what the parts of the mind are for. To understand sight, we have to look to optics and computer vision systems. To understand movement, we have to look to robotics. To understand sexual and familial feelings, we have to look to Mendelian genetics. To understand cooperation and conflict, we have to look to the mathematics of games and to economic modelling. (1997, p. 38)

This list could be continued. It gives an idea of why definition (EP) is not stated so as to single out specific disciplines as the privileged providers of the knowledge and methods that evolutionary psychologists tap into.

Specifically, palaeoanthropology and hunter-gatherer studies are eminently important in constraining hypotheses about psychological adaptations, for they provide psychologists with knowledge about the ecological conditions in which such adaptations would have evolved. Hunter-gatherer studies, like other studies on non-Western subjects, are of course also indispensable in testing hypotheses about putatively universal human psychological adaptations, once these have been advanced (Henrich et al. 2010). But a host of other disciplines need to feature in evolutionary psychology, too. As with regard to topics, if some disciplines and methods still rarely show up in the field, this may simply reflect the short history of evolutionary psychology so far—and its limited human resources.

I now turn to the two broad paradigms that evolutionary psychology combines: first, computationalism in psychology and cognitive science (Sects. 2.4 and 2.5); then, modern evolutionary adaptationism (Sect. 2.6) as applied to the human mind (Sects. 2.7 and 2.8), which forms the more distinctive aspect of evolutionary psychology as a way of doing human psychology.

2.4 Computationalism

The *computational* or *information-processing* approach to psychology originated at the birth of cognitive science in the 1950s and 1960s, sometimes called the cognitive revolution (Bechtel et al. 1998; Mandler 2007, Chap. 10; Neisser 1967). This approach is widely labelled ‘cognitive psychology’ or ‘cognitivism’ by evolutionary psychologists, but I will avoid these terms in what follows (*pace* Barrett and Kurzban 2006, n. 2). The basic reason for avoiding them receives an added degree of urgency when considering the characteristic scope of vision of evolutionary psychology.

On a standard understanding that conforms with traditional usage and etymology, the adjective ‘cognitive’ refers only to processes of *knowledge acquisition* (perception, reasoning, memory, attention, etc.), while excluding motivation and emotion. The same is true of the term ‘cognitive psychology’. One need only consult the tables of contents of volumes with this title to see this, from Neisser’s (1967) classic to contemporary textbooks. The traditional meaning of ‘cognitive’ is congruent with classical cognitive science’s peculiarly restricted scope of vision. For a few decades after its inception, cognitive science was almost all about knowledge acquisition, with motivation and emotion pushed to the margins. Moreover, the phenomena of knowledge acquisition themselves tended to be investigated as if they had never evolved in the service of behaviour regulation.

This tradition of classical cognitive science is the contrary of what evolutionary psychologists stand for. For while evolutionary psychology is in the typical case openly computationalist, at the same time it emphasizes the many facets of *behaviour regulation*. Indeed, it expects to find influences of behaviour regulation all over the realm of knowledge acquisition itself. The general reason for this is that only behavioural effects can be acted upon by natural selection. One can of course stipulate to use terms as one sees fit. But in light of the double fact that even psychological scientists still frequently used 'cognitive' specifically for knowledge acquisition, and that evolutionary psychology precisely counteracts traditional emphases on knowledge acquisition, it is not a good idea to refer to evolutionary psychology as promoting a 'cognitivist' approach.

Psychology is computational to the extent that it takes mental processes to consist in *mechanical transitions between semantically evaluable states* and other states. These other states can be other semantically evaluable states (representations) or physiological outputs. What makes it appropriate to call these state transitions mechanical is that they causally respond in uniform ways to specific non-semantic (physical, formal) properties of the representations entering into them. Stated only in the abstract, this kind of statement is not too helpful. Let me explain matters with the help of a couple of contrasting examples.

Consider someone who is trying to declutter his or her private household. For one item after another, she has to decide whether to stow it in a box and keep it, or whether she should discard it. In situations like these, a practically boundless variety of thoughts or emotions can come to bear on one's decision. For any item, one might consider which practical uses it might have, whether one expects to use it in the future, whether it has significant cash value, whether one has the same kind of item already, whether its storage is costly, and so on. One might feel emotional attachment to the object; on the other hand, perhaps the object reminds one of somebody one would rather not be reminded of. Somehow, these kinds of considerations and emotions automatically pop up in considering the item that needs to be sorted, and somehow a decision results. There must be certain mechanisms that bring the behavioural results about, and it is highly nontrivial how they work. But these nonobvious mechanisms must involve the causal interaction of stored *internal representations* with each other, and with other states.

Consider by contrast, a robot pursuing the sole task of putting an item in a box if (and only if) it weighs less than 1 kg. Or perhaps, if the item has a red colour. Or if it emits alpha rays. Ascertaining the presence of conditions like these does not require an understanding of what the conditions ascertained 'mean', in any sense of this word (for example, that 1 kg is a unit of mass, that many items with a lot of mass in one's household will make one's next move arduous, and so on). The robot is simply responding in uniform ways to specific physical, non-semantic properties. This lies at the heart of our ascription of a mechanical mode of operating to the system.

The robot's behaviour in the simple example just given is obviously, paradigmatically mechanical. The human's behaviour and mental goings-on in the parallel example a paragraph earlier is paradigmatically non-mechanical. What computational cognitive science tries to do is to find explanations of a feat like this and countless other human feats and tendencies, through the *mechanistic decomposition of mediating mental states and processes* into simpler elements and processes, such that these interactions ultimately bottom out in interactions that *are* paradigmatically mechanical. The mechanisms invoked are posited to operate on *representations* implemented in the brain. To the extent that they do, they can be said to perform *computations*. The basic idea of a computational, information-processing psychology is no more complicated than that.

A closely connected aspect of computationalism tends to receive less attention than it deserves: Even though the explanations of computational psychology crucially involve positing semantically evaluable states (‘intentional states’, in philosophers’ lingo), they are, at least in large part, *not situated at the level of common-sense explanations of behaviour*. The latter explanations are couched in terms of states like beliefs, desires, and intentions. Moving beyond this level constitutes progress because, even though the belief-/desire-based psychology we utilize in everyday life is way better than no intentional psychology at all, it only goes so far. It is a lot worse than anything we would expect from a mature scientific psychology.

Why? For one thing, explanations in folk psychology are often awfully crude. For another thing, they only tend to posit states that are in principle accessible to consciousness. Last but not least, I think, they often invite the rationalization of actions, motivations, and feelings in terms of this already inadequate inventory, even when there is nothing to rationalize and only brute causation is involved.

While many of the internal representations posited in a computational psychology stand for entities and properties which we readily refer to in everyday discourse, a great many of them stand for entities, properties and relations for which we have no convenient vernacular labels. The same is true in spades of the mechanisms that are postulated by computational psychological theories. Relatedly, folk categories such as those of belief or the self, central to ordinary discourse about persons, may well turn out as heterogeneous and inadequate from the viewpoint of computational psychology.

The last aspect I wish to point out is that computational psychology as such abstracts away from neuroanatomy. It thus goes naturally with the ontological doctrine about mental states known as *functionalism* (Block 1995). The functionalist idea is that what makes a state a certain kind of mental state is its having a certain causal potential—a functional role—rather than its being constituted by a certain kind of material. That is, the essence of mental states is relational and medium-independent, rather than residing in intrinsic properties of the underlying brain tissue. Hence, the slogan, learnt by generations of cognitive science students, that *the mind is what the brain does*. It is not simply what the brain *is*.

While it nowadays often meshes with so-called cognitive neuroscience and the search for neural correlates of mental functions, the characteristic activity of computational psychology is the pursuit of abstractly specified representations and procedures. To say that these representations and procedures are abstractly specifiable is not to deny that inspecting the brain and localizing mental functions can be of considerable help in identifying them (cf. Bechtel and Richardson 1993/2010; Mundale 2002) and sharpening our understanding. Indeed, consistently with what I remarked about evolutionary psychology's permeability to other disciplines, a number of evolutionary psychologists already are doing cognitive neuroscience (see, e.g., the contributions in Platek and Shackelford 2009).

Nor does a computational functionalist as such deny that low-level neuroscientific facts can *constrain* the causal properties of information-processing mechanisms which she is interested in. The leading evolutionary psychologists were forthright about this already decades ago. Cosmides and Tooby (1994, p. 46) addressed the 'currently fashionable' idea that findings about neurons, neurotransmitters, and cellular development 'will eventually place strong constraints on theory formation at the cognitive level', and they agreed: 'Undoubtedly they will'. But, they argued, these constraints would be similarly pervasive as the basic neural tissue encountered in all animals, and facts about this neural tissue do not begin to answer the question of which computational programs are executed by different animals in performing species-typical activities (such as web spinning in spiders, echolocation in bats, or birdsong). Answers to that kind of question must lie in organizational properties more abstract than specific biochemical properties of neural tissue.

In the same functionalist vein, Pinker comments on the relationship between understanding the brain and understanding the mind. Referring to the simplest information-processing units, or 'microcircuits', from which computational programs are assembled, he is happy to admit the thoroughly substrate-dependent nature of their causal powers—but he contends that the process of discovering and mapping our psychological programs will need to go on for a long time before knowledge about their neural substrates can become relevant to them:

> What those microcircuits can do depends only on what they are made of. Circuits made from neurons cannot do exactly the same things as circuits made from silicon, and vice versa. For example, a silicon circuit is faster than a neural circuit, but a neural circuit can match a larger pattern than a silicon one. These differences ripple up through the programs built from the circuits and affect how quickly and easily the programs do various things, even if they do not determine exactly which things they do. My point is not that prodding brain tissue is irrelevant to understanding the mind, only that it is not enough. Psychology, the analysis of mental software, will have to burrow a considerable way into the mountain before meeting the neurobiologists tunnelling through from the other side. (1997, p. 26)

Typical evolutionary psychologists can be taken to agree with this methodological bet. They can be understood as placing bets on the assumption that many years of relatively autonomous theorizing at a relatively abstract level of representations and procedures still lie ahead, before neurobiology at the cellular level can start to constrain those computationalist theories. In doing so, they simply share the stance of the rest of the large research community in cognitive experimental psychology.

In conclusion, the research program of evolutionary psychology is committed to a computational approach insofar as its practitioners assume the following:

1. that explaining behaviour requires positing a rich variety of internal *representations,*
2. that these representations cause the occurrence of other representations or physiological states through *mechanical*, meaning-blind procedures,

3. that many of these representations' contents, as well as the machinery processing them, are *not* already posited in *folk-psychological* discourse and are in principle inaccessible to consciousness,
4. that psychological theories can and should be couched at a predominantly *functional* level, prescinding from neurobiology.

2.5 Understanding Computationalism Broadly Enough

Sometimes the computationalist approach to the mind is regarded as inessential to evolutionary psychology. This negative assessment is usually expressed implicitly, by omission of explicit reference to computationalism in defining evolutionary psychology.[12] I should like to suggest that, when researchers knowledgeable of evolutionary psychology make this assessment, they are moved to do so mainly because of inordinately narrow understandings of what the computationalist idea amounts to.

Such narrow understandings are especially encouraged by the tenaciously and widely held notion that computationalism is tantamount to endorsing 'the computer metaphor' (or 'the computer analogy') of the human mind/brain.[13] This notion is inappropriate in two ways. For one thing, it suggests that generations of cognitive scientists have been committed to computationalism only insofar as they have been vaguely inspired by nonliteral talk. This is not true, because computationalism must be understood as a claim about what the human mind/brain literally is, or does. More relevantly for present purposes, to the extent that literal conclusions are drawn from the attribution of a 'computer metaphor', these are far too narrow to do justice to computationalism about the mind, as I now wish to explain.

Conventional Computers? Disclaimer 1 Insofar as an identification of computationalism with 'the computer metaphor' of the mind is understood as implying literal assumptions about the human mind/brain, the associated assumptions tend to cluster around actual properties of commercially available, *conventional digital computers*. Presumably, a large family of intuitive, hyper-specific misconstruals of computationalism is based on this association and continues to exert a grip on researchers outside traditional core disciplines of cognitive science.

It is worth noting that cognitive scientists already struggled against misconceptions of this kind decades ago. Take, for example, Fodor and Pylyshyn's remarks

---

[12] Buss sometimes does this in statements of what evolutionary psychology is. Cosmides and Tooby, in contrast, have always been significantly more outspoken about the centrality of computationalism to evolutionary psychology. It is no coincidence that contributions by former students of theirs address computational-representational questions more directly than many other researchers in the field do.

[13] Remarkably, some leading contributors to evolutionary psychology who are committed computationalists and experts in computational modelling have themselves described the computationalist approach as though it were tantamount to a mere 'computer metaphor' (Miller 2000, p. 42; Gigerenzer and Goldstein 1996).

towards the end of a classic 1988 paper, in which they assumed such misconceptions to be so pervasive that encounters with them could be presumed as a collectively shared experience among researchers teaching computational modelling at universities:

> Almost every student who enters a course on computational or information-processing models of cognition must be disabused of a very general misunderstanding concerning the role of the physical computer in such models. Students are almost always sceptical about 'the computer as a model of cognition' on such grounds as 'computers don't forget or make mistakes', 'computers function by exhaustive search', 'computers are too logical and unmotivated', 'computers can't learn by themselves; they can only do what they're told', or 'computers are too fast (or too slow)', or 'computers never get tired or bored' and so on. (1988, p. 146)

The general response to this is that such misgivings mistake specific ways of implementing computational processes for the hypothesis of computationalism itself.

Only Classicism? Disclaimer 2 The paper from which the above quote is drawn defended the classical variety of computationalism about cognition against a rival which began to attract much attention and a broad following in the mid-1980s: the connectionist paradigm (Rumelhart et al. 1986; McClelland et al. 1986; Macdonald and Macdonald 1995; Clark 1989; Bechtel and Abrahamsen 1991). Someone's conception of computationalism might be abstract enough to avoid the erroneous restriction to specific properties of conventional computers, while still inappropriately excluding connectionist computationalism.

It needs to be briefly indicated here what the difference between classicism and connectionism consists in. The difference between the two is often stated with reference to the machinery of the so-called neural networks posited in connectionist models.[14] However, positing neural networks is not by itself what distinguishes connectionist modelling of cognitive processes. It is quite uncontentious that neural networks can *implement* basic representations or rules of the sort that classical computational models rely upon. Connectionism starts where expansionist claims are made regarding neural networks' ability to account for cognition, above the level at which basic representations and rules are implemented. The single most important difference between connectionist and classical computationalism consists in a defining resource of the latter, which connectionism claims can be supplanted by neural networks: Classical computationalism assumes that cognitive processes operate on *stored, discrete symbols* that stably retain their identities across recombinations into different structured representations. Classical computationalism is also often described as the 'symbolic' approach among cognitive scientists (Newell and Simon 1976; Newell 1980).[15] Given the discretely recombinable nature of the

---

[14] Neural networks are so called despite being only loosely inspired and quite unconstrained by properties of real neurons. This machinery involves nodes (visible or hidden), activation levels, weighted connections (excitatory or inhibitory), learning rules, and other ingredients.

[15] Among philosophers, the paradigm often goes by the name of 'the language of thought hypothesis' (Fodor 1975, 1981), due to the language-like nature of any representational medium involving discrete symbols combinable into complex constituent structures.

representations it posits, one can describe classical computationalism as committed to digital computing.

There exists a terminological practice on which expressions like 'computationalism' or 'computational theory of mind' are reserved for the classical computationalist paradigm,[16] thereby excluding connectionism. This corresponds to the second unduly narrow conception of computationalism presently distinguished. Steven Pinker, who has contributed as much to foundational issues in the area of computation and cognition as anybody, states the matter crisply when he says that 'we need to set aside some red herrings. Connectionism is not an alternative to the computational theory of mind, but a variety of it, which claims that the main kind of information processing done by the mind is multivariate statistics' (1997, p. 114). Barring a form of eliminative connectionism that altogether tries to dispense with the notion of representation, connectionist modelling is within the bounds of the resources that evolutionary psychologists can avail themselves of.

Only Explicit Rules? Disclaimer 3 The next disclaimer specifically concerns the essential aspects of classical computationalism. Traditional characterizations of classical computationalism often emphasized that its symbols are manipulated in accordance with *rules*. Indeed, the phrase 'rules and representations' has sometimes been used synonymously for this paradigm. But this is probably not a good idea. On the one hand, if talk of symbol manipulation in accordance with rules were merely to mean that symbols enter into mechanical transformations *describable* by rules, then this would not be distinctive of classicism. On the other hand, if it is taken to mean (as it more typically is) that symbols are transformed according to stored, *explicitly represented rules*, then the idea is too strong. Rules of computation can be explicitly represented, as they are in the data structures known as programs. But they need not be, and they almost certainly very often are not, in real creatures.

Rigid Routes of Information Flow? Disclaimer 4 The final disclaimer I will make here concerns classical computationalism again. Information flow in classical computational architectures is often conceived as though it proceeded through rigidly prearranged pipelines (Barrett 2005). The information flows in one overall direction and, once processed, it ceases to be available to other specialized systems in a given layer of the architecture. The information processed is only locally available, in the sense that it is available to only one functional component at a given time—somewhat like a mailed letter will be in one processing stage and place at any given time (post box, mail van, distribution centre, conveyor belt, sorting machine, pigeonhole, mailman's bag, etc.).

In such architectures, for representations to become available to mechanisms capable of processing them, they need to be routed to them first. This requires a meta-mechanism to sort incoming information first—much like someone in a mail distribution centre who sorts incoming letters according to destination. And such a meta-mechanism will necessarily have an input domain that subsumes at least

---

[16] Philosophers are particularly prone to collapsing these two ideas. See, for some prominent examples, Cummins (1989), Carruthers (2006), and any of the books by Fodor cited in this chapter.

all of the input domains of the mechanisms to which it sends information. While this does not pose a full-blown regress problem,[17] it does stand in the way of the computationalist aim of mechanistic decomposition. Computational explanations are supposed to explain complex and intelligent capacities in terms of the regular interaction of simpler states. Once powerful meta-mechanisms are posited that can take a vast range of different inputs and make context-sensitive decisions on which mechanisms are suitable for processing them, one begins to posit capacities of similar sophistication as the cognitive capacities one set out to explain.

However, as the evolutionary psychologist Barrett (2005) reminds us, a rigid *pipe architecture* of the sort just sketched is not the only type of computational architecture one can avail oneself of. With nods to models by Selfridge and Neisser (1960)[18] and Holland et al. (1986), Barrett points out the option of a bulletin board architecture. In such an architecture, making information accessible to appropriately specialized components is not a task in its own right that needs to be accomplished anew for each piece of incoming information. Rather, an inherent feature of the systemis is *open access:* incoming information is made available in a common pool of information. This common pool—the public bulletin board, or perhaps better, representation soup—is monitored by many different mechanisms specialized for processing different inputs. Instead of passively waiting for appropriate representations to be routed to them along a rigid grid of pipes, the specialized mechanisms hover by the bulletin board like enzymes in a living cell, springing into action to perform their specific jobs once appropriate information shows up. The general access to information in the bulletin board system does not lead to obstructions or chaos because the specialized processors detect their substrates by lock-and-key style template matching. This matching can be highly specific. Once an enzyme-like computational mechanism has detected an appropriate substrate and performed its transformations on it, the processed substrate is *returned* to the common pool. It is thereby available to other processors within the same system again, rather than disappearing from sight.

As Barrett (2005, p. 271 f.) observes, one can easily build into this kind of architecture the capacity for enzymes to selectively switch each other on or off, depending on the context. It is fundamentally the access-generality of this architecture (metaphorically: many mini-machines floating in the same representational soup) that allows such innovations.

Enzymatic computation in a bulletin-board setting provides a good foundation for explaining the famed flexibility and global information-sensitivity of higher cognitive processes. These are properties that some have thought to pose a mystery intractable by computationalist means (Fodor 1983, 2000). In my own experience at least, the idea of relevantly enzyme-like computational mechanisms is a liberating

---

[17] This has been claimed by Fodor (2000, pp. 71–73) and admitted by others with regard to the pipeline-style architecture (Barrett and Kurzban 2006, p. 634).

[18] The locus classicus for Oliver Selfridge's pandemonium model is his (1959), a remarkably early contribution to cognitive science of lasting influence.

and central ingredient in thinking clearly about potential computational realizations of human psychological capacities.

Classicism, Connectionism, Evolutionary Psychology I have tried to convey the breadth of the idea that evolutionary psychology is committed to in virtue of endorsing computationalism. This has involved pointing out some things that, more specifically, classical/symbolic computationalism is not committed to, though we would have reasons to be apprehensive if it were: a rigid pipe architecture, explicit rules en masse, and various other features of conventional digital computers. Recognizing the broadness of classical computationalism (and not just computationalism tout court) is important. For whatever theoretical resources it may take to explain all facets of the mind, classical computations will plausibly have to be posited as underlying *many* mundane aspects of the mind. A few remarks still need to be made about the plausibility of classicism and connectionism, and about the role they play relative to evolutionary psychology.

Further above, I quoted some objections that newcomers were reported to standardly raise against computational models of human cognition, with conventional digital computers in mind. These objections reminded Fodor and Pylyshyn of the kind of disaffection connectionists regularly expressed about classical computationalism.[19] Fodor and Pylyshyn admitted that the connectionists advanced ‘relatively more sophisticated complaints’ (ibid.), but viewed these as being similarly easily dismissible as the standard misunderstandings by introductory-level students, addressing inessential implementational aspects. Let me just remark about this that not all the misgivings about symbolic computationalism which gave the rivalling paradigm of connectionism its allure (cf. Bechtel and Abrahamsen 1991, pp. 56–65) may be dismissible so easily. Adjudicating their relevance or irrelevance may depend on nonobvious answers to deep questions. Moreover, connectionism does seem attractive for *some* more than merely implementational mental phenomena.

That being said, in many important areas of the human mind there hardly exists a viable alternative to symbolic computationalism. For several reasons, it has seemed clear for a while that connectionism is incapable of accommodating large and important portions of everyday human cognition (Pinker 1997, pp. 114–129; Marcus 2001)—and not just evidently symbolic activities like the conscious application of rules learnt in a classroom. One large group of examples is provided by the ability to represent and keep track of *individuals*, which is deeply important even independently of the requirements of the modern world. To merely give some examples from the realm of group living, we need to entertain thoughts and emotions about individual conspecifics as we meet challenges of mate choice, parenting, reciprocating favours, or coalition formation. This requires us to be capable of entertaining concepts of individuals *as* individuals, as opposed to merely representing their properties. The mundane cognitive ability of representing

---

[19] ‘If we add to this list such relatively more sophisticated complaints as that “computers don’t exhibit graceful degradation” or “computers are too sensitive to physical damage”, this list will begin to look much like the arguments put forward by connectionists’ (Fodor and Pylyshyn 1988, p. 146). See also all of Sects. 4 and 5 in the same paper.

individuals as such, and to do so across time, turns out to be surprisingly hard to achieve by connectionist means (Marcus 2001, Chap. 5; Pinker 1997, pp. 114–118). Indeed, it seems quite impossible to achieve without resorting to central elements of a symbolic architecture.

Where evolutionary psychologists explicitly address the internal structure of computational adaptations,[20] it seems fair to say that the symbolic approach is adopted by default. This may reflect an appreciation (on some level) of what connectionist models cannot naturally do.

The other reason why evolutionary psychologists take the symbolic approach by default is that evolutionary psychology and connectionism naturally fall on opposite sides of the nativism-empiricism divide. Connectionism is very typically endorsed by cognitive scientists with empiricist leanings. Empiricists believe in explaining the manifold capacities of minds on the basis of a small set of general-purpose learning mechanisms, operating according to universally applicable principles like association, analogy, or induction. Connectionism naturally caters to this explanatory aspiration. The general-purpose learning rules which it stipulates to govern its neural networks are naturally construed as providing the general-purpose learning mechanisms that empiricists require.

Evolutionary psychology, by contrast, is committed to adaptationist nativism (see Sect. 2.7). Symbolic computationalism can naturally provide the content-specific information-processing adaptations and 'innate contents' that the evolutionary psychologist requires. It is therefore unsurprising that evolutionary psychology tends to be associated with symbolic computationalism and not with connectionism. Still, this ought not to blind us to the scope of resources that may be from case to case available to the evolutionary psychologist in explaining psychological capacities.

Broad, Not Banal In spite of the generality of the defining features of computationalism spelled out in the previous section, and in spite of the broadness of the computationalist approach which I have emphasized in the present section, endorsing the computationalist approach is not trivial. It has been, and still is, far from universally adhered to. Also, notice that if it were banal, then the inception and dissemination of computationalist thinking in the 1950s and 1960s could hardly have been regarded as a revolution (i.e., 'the cognitive revolution').

The computationalist approach is, for example, not behaviourist. Nor is it an introspectionist approach to investigating the mind. Relatedly, it is not phenomenological. It is not psychoanalytic or anything else that might broadly be subsumed under the rubric of psychodynamics. These facts by themselves allow us to identify vast research traditions in the history of psychology as non-computationalist. Moreover, in more recent decades, computationalism has not gone unchallenged. It is incompatible with strong forms of 'embodied' or 'enactivist' approaches to the mind, whose opposition to computationalism forms a central part of their identity. It has been opposed by ambitious forms of dynamical systems theory. It precludes

---

[20] At the present stage, many evolutionary psychologists do not yet do so. I take this to be a sign of the relative youth of the discipline.

any account that explains cognitive processes directly through intrinsic properties of neural tissue. And obviously it is not substance dualist in its explanation of mind and behaviour.

Having clarified this, I now turn to the adaptationist nature and methodology of evolutionary psychology. This requires briefly clarifying its general background first, which consists in the modern theory of natural selection.

2.6 The Modern Theory of Natural Selection

Adaptations are traits of biological systems whose effects caused them to be favoured by natural selection in the evolutionary past, i.e., to be reproduced more successfully than alternative variants. For a trait that is an adaptation, the effect that caused it to be favoured by natural selection in its evolutionary past is its biological *function*. Evolutionary psychology takes an *adaptationist* approach[21] to the human mind. It does so, fundamentally, in virtue of the extent to which it takes natural selection to have shaped and differentiated the human mind (more on this in Sect. 2.7).

When I said earlier that evolutionary psychology combines two broad, standard research paradigms from psychology and biology, these paradigms might easily have been referred to as computationalism about the mind plus adaptationism according to the modern evolutionary synthesis, or, the Modern Synthesis. Apart from the clarifications in the present section, however, I will not speak much of adherence to 'the Modern Synthesis' in connection with the commitments of evolutionary psychology. Nor do evolutionary psychologists themselves. I see two reasons for this.

Reason one is that it is really only in a broad sense that adherence to the Modern Synthesis is essential to evolutionary psychology and that, in this broad sense, the label 'Modern Synthesis' is effectively superfluous outside of historical contexts. In the broad sense, the attribution of this label does not add anything to saying that evolutionary psychologists endorse adaptationism consistently with the standard lines of how evolutionary biologists since the mid-twentieth century conceive of natural selection. There is room for a broad and exciting variety of theoretical emphases in this standard framework. There are differences in the extent to which researchers believe natural selection to have influenced the evolutionary process. But to the extent that it has, it is textbook wisdom that natural selection operates along the general lines envisaged in the modern evolutionary synthesis.

---

[21] An effectively synonymous term is 'selectionism'. I will understand this to describe the same position as 'adaptationism', only connoting a different perspective on it—that of the population-genetic dynamics of natural selection, rather than its phenotypic results. Daly and Wilson (1995, 1999) or Tooby and Cosmides (2005) use both terms, apparently with something like this distinction in mind.

Reason two for avoiding the label 'Modern Synthesis' it that it is more narrowly understood by some evolutionary theorists than I understand it to be—in effect, they understand it as a myopic research program requiring revision or replacement. If reactions of mainstream evolutionary biologists to revisionist criticisms could be summarized in a sentence, I think it would be this: Just because the modern evolutionary synthesis does not mention some important aspect of the evolution of life does not mean it has overlooked it or disallows it. Regarding various emphases which have been promoted in recent decades as missing elements or even rivals of the Modern Synthesis (Pigliucci and Müller 2010; Pigliucci 2007), mainstream evolutionary biologists are inclined to point out that they do not contradict the synthesis and are naturally integrable into it. For some insights, mainstream biologists will also point out that they are less novel than they are made out to be by their promoters. Finally, in some cases mainstream evolutionary biologists may feel the emphases to be overstated by their more enthusiastic supporters.

Evolutionary psychologists are inclined to respond along much the same lines to revisionist or revolutionary rhetoric directed against the Modern Synthesis. What they are inclined to say about allegedly neglected insights is that they may (it depends on which) provide a fuller picture of ways in which natural selection operates, but that they do not undermine natural selection's relative importance in the evolution of the human mind/brain. Many different theoretical insights are consistent with the more general emphasis on the importance of natural selection. I cannot begin to make this general assessment plausible here, however.

Few non-specialist academics are nowadays aware of how little acceptance there was in the scientific community after Darwin (1859) and until the 1940s (Bowler 1983/2003, Chap. 7; Mayr 1982, Chap. 11; Mayr 1991, passim) for evolution by natural selection as an explanation of anything interesting in the living world. It is only through the modern evolutionary synthesis that the theory of natural selection became widely established. The Modern Synthesis was worked out by biologists of various disciplines[22] in the 1930s and 1940s, building on the work of pioneering mathematical population geneticists in the preceding years. It assigns central importance to natural selection in the evolutionary process because it is organized around this concept. For what the Modern Synthesis is a synthesis *of* is simply Darwin's theory of natural selection and Mendelian genetics.

Two genetic insights above all were crucial to the synthesis, corresponding to the respective rejections of two previously widespread and natural-seeming ideas. Firstly, the endorsement of *hard inheritance*, entailing the rejection of the inheritance of acquired characteristics (Lamarckian inheritance). Secondly, the endorsement of *particulate inheritance*, entailing the rejection of a conception on which inheritance in sexually reproducing organisms is based on the blending of traits of parent organisms. These insights have not been seriously disputed by biological science to this day.

---

[22] Dobzhansky (1937) heralded the Modern Synthesis. Huxley (1942) coined the name in his somewhat more popular classic. For more historical background, see Mayr and Provine (1980), Mayr (1991, Chaps. 9–11), and Smocovitis (1996).

Since the modern evolutionary synthesis, natural selection is understood to be a gradual process acting on inherited elements—most fundamentally, genes—over many generations. Variation in these elements, the precondition for the occurrence of natural selection, is introduced by occasional genetic mutations. In rare cases, mutant genes, embedded in their complex genetic-developmental setting, result in phenotypic traits which lead their bearers to enjoy a higher rate of survival and reproduction than other bearers with other gene variants. This leads to natural selection. The mutant genes replicate more successfully than the alternative variants, *gradually spreading in populations across numerous generations*. Complex specialized traits, for example, organs like the human eye or the heart, are in turn explained through the transgenerational accumulation of component traits which gradually spread according to this principle.

What can be viewed as an updated version of the Modern Synthesis framework, inclusive fitness theory, established itself in the 1960s and 1970s on the basis of works by W.D. Hamilton (1964a, b) and G.C. Williams (1966). This changed the science of animal behaviour, which in the course came to be called 'sociobiology' (Wilson 1975) instead of 'ethology'. The inclusive fitness revolution, however, merely followed the ramifications of the population genetics underlying the Modern Synthesis and is best viewed as a logical outgrowth of it (Dawkins 1989). Consistently with the Modern Synthesis, inclusive fitness theory takes natural selection to fundamentally involve the differential replication of gene variants in populations (though this does not in itself preclude natural selection at higher hierarchical levels). Gene variants whose effects in a given causal surrounding tend to increase their number in subsequent generations spread in a population at the cost of gene variants whose effects leave fewer copies.

These replicator dynamics are very often aligned with, but explanatorily more fundamental than, individual organisms' boundaries and survival interests. For example, they plausibly explain why we do not live forever. Relatedly, they are capable of explaining basic forms of altruism (that is, benefitting others at a cost to oneself) that are directed at close kin. They also preclude naive notions of selection acting 'for the good of the species', formerly widely accepted among ethologists, and still widespread to the present day among nonbiologists who sporadically dabble with quasi-adaptationist explanations of phenomena (see Sect. 2.8).

2.7 Evolutionary Psychology's Ontological Adaptationism

Evolutionary psychologists generally do not distinguish between different kinds of adaptationism. They simply identify themselves as adaptationists, full stop. But we should distinguish two aspects of their adaptationist position. Firstly and most fundamentally, evolutionary psychologists are adaptationist in an *ontological* sense (which is implicitly comparative): They make a claim about what there *is*. In the present section, I will say more about the adaptations posited than was anticipated already in connection with (EP). Secondly, building upon the ontological assumption, evolutionary psychologists are adaptationist in a *disciplinary* sense (which might also be called methodological, though I prefer not to use this term for reasons explained in Sect. 2.8): They commit themselves to a certain way of *doing* psychology in finding out what they believe there is. This will be the subject of the subsequent section.

1. More Mental Adaptations 'Adaptationist', in what I take to be the most common sense of the term, is a thoroughly vague predicate. The greater one takes natural selection's influence on biological evolution to be, the more strongly adaptationist one is. The same is true, mutatis mutandis, for adaptationism about specific realms of the biological world. Evolutionary psychologists are adaptationist, fundamentally, in virtue of the extent to which they take natural selection to have shaped and differentiated the human mind. What extent must that be, in order for someone to count as an evolutionary psychologist? This question seems pressing because there is a straightforward sense in which all evolutionary scientists are adaptationists. Practically everyone with the relevant scientific education who is not committed to creationist religious extremism acknowledges that natural selection has played some important kind of role in the evolutionary process.

The sense of 'adaptationist' I am presently aiming at is implicitly comparative. Ascribing adaptationism to somebody in this sense implicitly involves comparing their view to predominant views in a given research community. Evolutionary psychologists believe that the human mind/brain *develops on the basis of many more adaptations than has been traditionally assumed* in psychology or, more generally, in the social, cognitive, and brain sciences at large. I am not going to be precise about what 'traditionally' means here, apart from saying that it means the larger part of the twentieth century (Tooby and Cosmides 1992; Pinker 2002). The late nineteenth and early twentieth century, by contrast, was a period in which 'instincts' were posited readily (McDougall 1908; James 1890; Angell 1907). As we have seen, however, in those decades the modern theory of natural selection was not yet available, nor was the information-processing approach to the mind. In many of the instinct-theoretical excesses of those years (see Boakes 1984), the absence of these scientific paradigms is painfully evident.

Apart from the fact that they must have evolved through natural selection along the general population-genetic lines sketched earlier, how are we to conceive of the adaptations positing which is characteristic of evolutionary psychology?

2. Computational Adaptations Let us first set aside a pleonasm which is of no help at all. One often hears evolutionary psychology described as positing *functionally specialized* adaptations. Such descriptions are redundant, given that 'adaptation' and 'function' are interdefined in the conceptual framework of evolutionary theory. To the extent that a trait can be said to be an adaptation at all, it is an adaptation specialized to perform a biological function: some causal role in virtue of which it was favoured by natural selection.

How, then, are the adaptations of evolutionary psychology functionally specialized? The general answer is provided by computationalism. The adaptations posited by evolutionary psychology are *computational adaptations,* in the sense that they have been selected for fulfilling certain causal roles in information-processing, which had certain behavioural consequences (which were, by hypothesis, on average reproduction-promoting).

Evolutionary psychology's postulated adaptations are standardly characterized as *computational mechanisms* (in these or similar terms). A computational mechanism can be defined by (1) the formally delimited class of inputs it accepts and (2) the transformations into outputs which it performs on them. However, this is inadequate as a general means of specifying the nature of computational adaptations in evolutionary psychology, for two reasons. For one thing, the pair of aspects just given only defines an unchanging, presumably adult, computational mechanism. This is deeply inadequate as a way of characterizing learning mechanisms. For another thing, not each component adaptation in a computational architecture constitutes an entire mechanism in the first place. I return to both of these points presently.

For the two reasons just anticipated, in contexts where we wish to refer to the totality of mental adaptations posited in evolutionary psychology, we do well to refer to them simply as computational adaptations. Less abstract labels—even the term 'computational mechanism'—may be intuitively easier to grasp, but only at the cost of being misleading or false when applied to all adaptations posited in evolutionary psychology. In fact, insidious terms such as 'module' only seem easier to grasp precisely because of irrelevant connotations which they introduce. The computational adaptations posited by evolutionary psychologists have been subject to a dismal history of misunderstandings. Given this and the lack of charitable intentions among critics, evolutionary psychologists cannot afford to use loose and, strictly speaking, false talk, trusting that everybody is going to figure out highly nonobvious matters by themselves.

3. Computational Adaptations as Learning Mechanisms The most important thing to keep in mind about most of the complex adaptations of evolutionary psychology, alongside the computationalist commitment, is that they are basically *learning mechanisms,* or similarly, acquisition mechanisms (I am somewhat undecided about which term to prefer). The point can be stated in the form of a recommendation: If in doubt, always think about a computational adaptation as a learning mechanism, an *adaptation for acquiring* a psychological trait, rather than as a specialized adult end result of development. Far too often, specific end results are envisaged when evolutionary psychology's computational adaptations are considered.

To characterize the computational adaptations of evolutionary psychology simply as mechanisms defined by specific input criteria and specific transformations is misleading. It implies an unchanging mechanism with a persisting mapping of inputs onto outputs. Since, by default, unchanging mechanisms are considered in psychology qua adult ones, this effectively implies a computational mechanism that is an end result of development. If we are to think of the naturally selected mechanisms posited by evolutionary psychology primarily as acquisition or learning mechanisms, then we must think of a more complex kind of entity: one whose mapping from inputs to outputs itself changes over time, due to growth patterns and information fed into the learning mechanism by experience.

As a reminder of the importance of trying to conceive of human computational adaptations as acquisition mechanisms, and as a reminder of the perils of unreflectively adultocentric conceptions, I find it helpful to occasionally recall the case of human language. It is obvious and undisputed that no adaptation exists for speaking a specific language such as German, Cantonese, or Swahili—even though the grammar and phonology of each of these languages will be firmly ingrained, in dedicated computational mechanisms, in the brain of a native speaker. But it is very well possible that there is an underlying language faculty, with a more abstract nature, which *is* a complex human adaptation (Pinker and Bloom 1990; Pinker 1994, 2003; Pinker and Jackendoff 2005). Such a claim is far from undisputed, to be sure (compare, e.g., Tomasello 1995; Hauser et al. 2002; Tomasello 2003, and open peer commentary on Evans and Levinson 2009, or Pinker and Bloom 1990), and it has been argued to the contrary that human language is an evolutionary by-product, based on 'universal aspects of human cognition, social interaction, and information processing' (Tomasello 2009, p. 471). But while it is obviously false that understanding a specific human language is an adaptation, it is not obviously false that a human language faculty embodying abstract principles of a universal grammar is an adaptation—and no scientist fails to appreciate this difference.

So suppose the language faculty is an adaptation. If so, it is a straightforward case of a computational adaptation with far more abstract and nonobvious properties than those of the computational mechanisms in which it results in the adult mind. At the same time the language faculty, if it is an adaptation, is nonetheless far more specialized than an all-purpose learning mechanism. In quite the same fashion, we must be prepared to think about a lot of computational adaptations as *mechanisms for acquiring more narrowly specialized computational mechanisms* in the adult brain—sometimes developing into a single mature mechanism, sometimes spawning several, independently maturing mechanisms according to need.[23] Whereas the mature computational mechanisms will often be responsive to highly specific inputs from our contemporary environments, with highly specialized procedures operating on them, the learning mechanisms forming the underlying adaptations will often be responsive to inputs of a nonobviously intermediate degree of specificity: more general than that of the mature computational specialization on the one hand, but still far more specific than those of perfectly domain-general learning mechanisms on the other.

4. Computational Adaptations that are not Learning Mechanisms It would simplify matters if we could conceive of all human computational adaptations simply as learning mechanisms. To be sure, thinking and speaking of human computational adaptations in these terms is vastly preferable to conceiving of them as innate mechanisms qua developmental end results. However, to do so across the board would be taking things too far. I see two reasons against restricting our conception of the computational adaptations posited by evolutionary psychologists to learn-

---

[23] This is also the gist of various remarks in deep but difficult foundational works by Barrett (2006) and Boyer and Barrett (2005), which are, besides, at least consistent with the subsequent remarks I make in the present paragraph. See also Barrett and Kurzban (2006). I must admit, however, that carefully re-reading Barrett (2006) has made me feel less clear than I previously recalled about how some of the arguments in it exactly run.

ing mechanisms. The first reason might be regarded as merely terminological. The second cannot.

Firstly, the possibility of *rigidly developing* computational mechanisms must be allowed for. There can be computational mechanisms that grow in a highly canalized fashion, according to a rigid assembly schedule, towards a specific outcome. Such developmentally canalized mechanisms[24] will typically require exposure to inputs of the sort they are prepared to process within critical periods of sensitivity in development. But the effects the experience has are not appropriately termed 'learning'. The visual system, and stereo vision in particular, provides a prime example of such an adaptation (Pinker 1997: pp. 238–241; Simons 1993). The development of stereo vision requires that some neurons in primary visual cortex receive inputs from the left eye and some of them receive inputs from the right eye. This aspect of development, as Pinker points out, needs to be viewed as a self-assembly requiring a very specific bit of information (which cannot be predicted in advance). But to sensibly speak of 'learning', as many understand the term, would presuppose a considerably larger space of possible psychological changes contingent upon experience.

The point just made could be regarded as negligible. Clearly, there is a continuum between (a) open developmental pathways that rely on the uptake of large amounts of unpredictable ambient information and are obviously describable as learning and (b) narrowly canalized developmental pathways that only depend upon specific informational inputs at certain junctures. The latter type of case might be treated as a limiting case of learning. One could stipulate to use the term 'learning' accordingly, given that it has already been used by scientific psychologists in ways remote from ordinary usage.

However, secondly, we must allow for the existence of adaptations that contribute to a computational architecture and its development even though they do nothing that is even the sort of thing that might be describable as learning. For example, consider a simple reasoning rule that reliably pops up early in development and performs its service unchanged for the rest of the organism's lifespan. Or consider a switch that has been installed by natural selection within a computational architecture. In fact, computational adaptations like the latter one do not just fail to qualify as learning mechanisms, they are not even whole mechanisms to begin with. Simple switches and other kinds of traits, among them quite abstract properties of computational architectures, can be naturally selected features of a computational architecture—computational adaptations—without being mechanisms. A fortiori, they are not learning mechanisms.

We need to give such bona fide computational adaptations their due and not succumb to the common habit of thinking of any computational adaptation as a mechanism. Still, it seems appropriate to think of learning mechanisms as forming the bulk of the human mind's information-processing adaptations, around which other computational adaptations can be organized.

---

[24] The developmental canalization metaphor is due to Waddington (1957).

5. Content Domains, Formal Input Domains Perhaps the notion most famously associated with the adaptations characteristically posited by evolutionary psychologists is that they are *domain-specific*. Sometimes commentators on evolutionary psychology ask what domain specificity might be and suggest that there may be some interestingly unresolved issue lurking here. There is not. The way in which the concept of domain specificity is used by psychologists, evolutionary or otherwise, is similar to the way in which the concept of adaptationism explained earlier is used. 'Domain-specific' is an implicitly comparative predicate, admitting of great differences of degree.

Fundamentally, domain specificity is a property of psychological learning mechanisms. A learning mechanism's degree of domain specificity is determined by how broad the class of properties, situations, and entities—or simply, *contents*—is for which the mechanism evolved to reliably develop into specialized cognitive or motivational capacities. Domain specificity is *content-specificity*.

There may be a learning mechanism specialized for acquiring the capacity to recognize mental states and explain actions of conspecifics; perhaps it contains several component learning mechanisms. There may be another learning mechanism specialized for acquiring a language; perhaps similarly complex. There will be mechanisms guiding the acquisition of sexual desires around puberty, narrowing down the anatomically conceivable range of objects of sexual desire with respect to sex, preferred physical appearance, and age of mates, not to mention less obvious potential desiderata. There may be a learning mechanism, specialized for avoiding incest, which selectively blocks the development of sexual desires when it comes to siblings. There may be another mechanism specialized for acquiring a vast memory store of different faces and recognizing them effectively. There will be mechanisms specialized for biasing one's acquisition of food aversion. Distinct from these, there will be mechanisms channelling our acquisition of food preferences (even though these are about food too). Yet again, there will be distinct mechanisms biasing our learning of which things to fear (even though these too are about things to avoid). And so the list of examples could be continued.

The contrasting theoretical position, in its extreme version, tries to explain the accomplishment of all these different tasks and many more through the operation of a few learning mechanisms on the basis of sensory-perceptual inputs, including the shaping influences of the social environment. This is empiricism. Its postulated learning mechanisms deal with all classes of contents equally, unequipped with any biases, procedures, representations, growth patterns, parametrisations, etc., that would prepare them for any specific class of situations, entities, and properties—contents—in the world (Tooby and Cosmides 1992; Pinker 2002, 1994: Chap. 13). This is what makes them *content-unspecific,* or equivalently, *domain-general*.

I have just explained the notion of domain specificity in terms of content domains. One may also find it helpful to identify a formal sense of 'domain specificity'. Computational mechanisms necessarily operate only on inputs that satisfy certain formal criteria—such is their meaning-blind, mechanical nature. Any computational mechanism therefore exhibits domain specificity in the sense of *formal input specificity*. A fortiori, this is true of naturally selected computational mechanisms posited in evolutionary psychology. Barrett and Kurzban, two former students of Cosmides and Tooby's, have maintained in a celebrated article that formal input specificity is all there is to the domain specificity of naturally selected computational mechanisms (they do not speak of learning or acquisition mechanisms). I criticize their claim below, since the matter is centrally relevant. Since I am puzzled by their position, let me make sure to be fair and state it in representative quotes of their own (all emphases are theirs):[25]

> We do not intend a reading of domain as *content domain*, in the folk sense of domains individuated by the *meaning* of their constituents. Rather, we define domains as individuated by the formal properties of representations because, we believe, this is the only possible means by which brain systems could select inputs. (2006, p. 630)
>
> Here we have suggested that domains should be construed in terms of the formal properties of information that render it processable by some computational procedure. In this sense, even the rules of so-called content-independent logics—for example, modus ponens—are domain-specific, in that modus ponens operates only on propositional representations of a certain form. (2006, p. 634)
>
> There are strong intuitions about what counts as a 'domain', many of which are not consistent with formal definitions […]. [F]or example, the 'domain' of a particular device could be as broad as *all object representations,* yet the device could still be a specialized, […] computational device. Intuitions prevent some psychologists from accepting the idea that a [specialized device] that processed all object representations is reasonably called domain specific, possibly because domain specificity implies to many differential processing of stimuli based on their 'meaning', rather than their 'formal properties'. The notion of meaning or content, however, is a folk notion: Human computational systems always process information based on formal properties […]. (2006, p. 634)[26]

Barrett and Kurzban go wrong here. It makes no sense to pit formal domain specificity against semantic domain specificity and to define domain specificity tout court over the formal notion, suggesting that the semantic notion is merely prescientific.

---

[25] The central treatment of domain specificity in Barrett and Kurzban (2006, p. 630) alone contains several quite confusing infelicities or errors which I pass over above. For example, the authors (B&K) introduce domain specificity and justify its formal nature thus: 'As a direct and inseparable aspect of this evolutionary process of specialization, modules [defined in the same paragraph as "functionally specialized mechanisms with formally definable informational inputs"] will become *domain-specific*: Because they handle information in specialized ways, they will have specific *input criteria*. Only information of certain types or formats will be processable by a specialized system' (emph. B&K's). It is unclear what B&K mean by saying that modules, in the 'evolutionary process of specialization', 'become' something they are claimed to be by definition. Moreover, while B&K stress domain specificity qua formal input specificity to be a necessary consequence of increased specialization in evolution, it is simply a necessary consequence of being a computational mechanism, whatever its origins. Also, the mechanism's formal input specificity need not increase in tandem with the specialization of its operations. Another oddity is the fact that the first passage I quote above follows on the heels of the following assurance: 'We wish to stress that we intend the broadest construal of the term *domain* to include, in principle, any possible means of individuating inputs'—which is precisely what they do not intend. This statement is most charitably reinterpreted as meaning, in marked difference, 'any possible [formal] means'. But this in turn is just an entirely inadequate individuation scheme, as I explain in the text.

[26] I have omitted from this quote, firstly, literature references and, secondly, uses of the precarious term 'module' which I do not believe add relevant content.

It is true that computational mechanisms can only operate on formally delimited input domains. It is not true, however, that the notion of a content domain is an expendable folk notion—and it is emphatically not true with regard to the selectionally relevant content domains of evolutionary psychology. Both notions are needed. Content domains are what formal domains evolved in response to.

In evolutionary psychology, the notion of a content domain is not a dispensable folk notion, but an indispensable notion that captures the biological functions of the many acquisition mechanisms posited in it. As illustrated above, these acquisition mechanisms were shaped by natural selection to reliably result in computational mechanisms that *regulate behaviour in different ways appropriate to different content domains* (i.e., types of situations, properties, entities in the world).[27] This is the evolutionary-biological sense in which the computational acquisition mechanisms are functionally specialized. The acquisition mechanisms were not shaped by natural selection to result in computational mechanisms that are formally input-specific in *semantically arbitrary* ways. On the contrary, a computational acquisition mechanism that evolved by natural selection will have a formal input domain that approximately *co-varies with the content domain* the mechanism evolved to deal with. Conversely, a computational acquisition mechanism that evolved by natural selection can co-vary with certain contents only via the mediation of *formal cues* that were reliably correlated with those contents in the environment in which the mechanism evolved.

Thus, there is no competition between the ideas of a formal domain and a content domain with regard to naturally selected computational mechanisms. Qua computational mechanisms, the mechanisms must be sensitive solely to formal properties of representations (e.g., geometric aspects of visual representations that are schematically face-like). Qua mechanisms evolved by natural selection, the mechanisms must have evolved in response to some selection pressures in the world correlated with those formal properties—selection pressures which ipso facto count as involving contents of some kind or another (e.g., faces of conspecifics). Neither the formal domain nor the content domain is a theoretically dispensable notion. What counts is to not mix up the two.

Even though Barrett and Kurzban's official theoretical statements fail to respect the points just made, they are of course not news to them (nor to other evolutionary psychologists). Doubtlessly, Barrett and Kurzban are on some level aware of them—and indeed, in plain contradiction of their official statements, they still regularly use 'domain' in a content sense throughout their 2006 paper and in other writings, apparently in implicit appreciation of its indispensability. Most notably perhaps, Barrett often emphasizes the distinction (cf. Sperber 1994) between the 'proper domain' of a computational mechanism that evolved by natural selection (the set of things it is designed to respond to, e.g., faces of conspecifics) and its 'actual domain' (the set of things it can actually process, e.g., face-like surfaces). This

---

[27] Typically, though not necessarily, 'the world' will be the external environment.

central theoretical distinction he tends to emphasize is purely content-based.[28] The reason it is necessary is precisely that these different types of domains are indistinguishable from a formal input point of view, but different in content (hence potentially vastly different in terms of causal potential). Barrett and Kurzban's critique of a content notion of domain is incoherent and fails to do justice to the fundamental importance of the content-based notion of a domain in evolutionary psychology.

The reason Barrett and Kurzban give for preferring a formal notion over a content notion of a domain is the fact that computational mechanisms can only process representations based on formal properties. If two leading evolutionary psychologists take this definitional truth about computational mechanisms (the 'formality condition', Fodor 1980) to cast a bad light on the notion of a content domain, even though the latter is crucial to adaptationist thinking about computational mechanisms, this ought to make us suspicious that some confusion has occurred. In fact, a confusion appears to be in play here that is well-known from the science of animal behaviour (Tinbergen 1963; Bolhuis and Giraldeau 2005). It rests on conflating different levels of psychobehavioural explanation—in particular, the levels of proximate-causal and evolutionary-adaptationist explanation. Questions located at the first of these levels ask *how* a behavioural pattern or mental phenomenon occurs. These are requests for explanations in terms of proximate causal origins. Questions located at the latter level ask *why* a behavioural pattern or mental phenomenon occurs. These are requests for explanations in terms of evolutionary origins, including possible adaptive functions.

Where these levels and types of questions are conceptually run together, noncompeting types of explanation are easily treated as though they were in competition with each other. This, I submit, is what happens when Barrett and Kurzban treat a formal individuation of domains as preferable to a content-based individuation of domains. Claims about formal input domains and claims about functional content domains (or, 'proper domains', see above) are answers to different questions. Knowing the formal input domain tells you something about *proximate causation:* which formal properties of inputs switch the mechanism on.[29] Knowing the functional content domain tells you something about *biological function:* which situations, properties, or entities in the world the mechanism evolved as an adaptation for.

6. Environmental Content-Specificity Evolutionary psychology is not compatible with the postulation of simply any conceivable large collection of computational acquisition mechanisms. Rather, *many* of its postulated acquisition mechanisms

---

[28] It is easy to be misled about this fact by Barrett's definitions of actual versus proper domains in terms of different classes of 'inputs processed'. This sounds as if it referred to different classes of proximal stimuli. In fact, it refers to proximal stimuli of the same class that come from different classes of distal stimuli.

[29] Strictly speaking, insofar as we speak about computational *acquisition* mechanisms, rather than mature computational mechanisms, the input domain itself would have to be described as changing over developmental time, especially at critical junctures. I allow myself to simplify here, given that this aspect is not presently the focus of attention (quite apart from the fact that it tends not to be observed by other authors anyway).

(though certainly not all) will have evolved to process inputs from *specific environmental content domains*. By this, I mean specific classes of situations, entities, or properties instantiated in our spatiotemporal surroundings. Acquisition mechanisms with environmental content domains provide most of the typical examples of computational adaptations posited by evolutionary psychologists. Some of these examples were listed earlier.

Notice that not all content domains of acquisition mechanisms are environmental domains. In principle, anything can be the content of a representation processed by a computational mechanism. Accordingly, we need to have a sufficiently open conception of potential content domains of computational mechanisms. Apart from environmental constellations, possible contents might, for example, consist in bodily states of oneself that are monitored by specialized computational mechanisms. Or the contents might be other representational states of oneself, registered by metarepresentational states subsequently subjected to processing. Contents might also consist in abstract objects and relations between them, such as mathematical facts, or the logical rule of modus ponens mentioned in one of the quotes above. That having been said, it is environmentally domain-specific psychological adaptations that are most frequently associated with evolutionary psychology.

The commitment of evolutionary psychology to many of its naturally selected acquisition mechanisms being environmentally content-specific usually goes unmentioned when evolutionary psychologists state their view of the human mind. Typically, with sundry terminological variations, they simply express a commitment to the existence of something like ‘many functionally specialized computational devices’ in the human mind. This standardly issued type of definition is compatible with positing many computational devices that are functionally specialized in virtue of their formal input criteria and operations, without being functionally specialized in virtue of particular environmental content domains. But this will not do if we are to capture the distinctive commitments of evolutionary psychology. The idea of specialization for many environmental content domains is indeed profoundly characteristic of adaptationist-computational evolutionary psychology. It is part of what knowledgeable researchers mean by ‘evolutionary psychology’. Take the aspect of ecological content-specificity away and you are looking at a research program so different that one would need a different label for it.

Suppose someone assumes the following model of the human mind. The human mind rests on developmental mechanisms that treat all kinds of perceptually registered contents according to more or less the same principles: speech sounds, habitats, food tastes, various behavioural indicators of various mental states, confrontation with a predator, snakes, images of naked nubile women, observations of social conformity, reprimands, observations about local cultural practices, and many more. How does this work? ‘Well’, says the friend of the present model, ‘for one thing, we learn much through our culture and social environment. But for another thing, not entirely unlike a conventional general-purpose computer with its thousands of parts, we have a great many *content-unspecific* mechanisms for processing and storing information. This rich equipment of content-unspecific computational adaptations allows us to learn and infer many different things. It reliably yields motivational and cognitive outputs appropriate to vastly different ecological content domains. It makes us flexibly intelligent'.

While we must not assume evolutionary psychologists to be particularly impressed by this idea, it is not obviously false. Indeed, from the evolutionary psychologist's viewpoint, it is surely less implausible than a view which also assumes the mind to treat all kinds of ecological content identically, while *not even* positing many domain-general computational adaptations to deal with them. But the latter kind of view is just standard empiricism—which in turn has been endorsed to considerable or undiluted degrees by many psychologists. Thus, we should also expect the kind of view just imagined to be actually endorsed by researchers, in varying versions and degrees of strength. We must expect there to exist fairly moderate views which are to varying degrees inconsistent with evolutionary psychology as standardly understood, but which standard self-descriptions by evolutionary psychologists fail to exclude.

I conclude that we ought to include reference to some amount of ecological content-specificity in adequate characterizations of evolutionary psychology's ontological adaptationism, if it is not to include too much.

The present section can be concluded with a disclaimer. There is a persistent myth that evolutionary psychologists disallow highly *domain-general* learning mechanisms (or other computational adaptations). I do not think it is necessary to comment extensively on this. Of course they allow them. Evolutionary psychologists only deny these the privilege of providing the default explanations for motivational and cognitive phenomena. They deny that domain-general learning mechanisms are preferable a priori over explanations in terms of content-specific computational adaptations (e.g., Cosmides and Tooby 1997, p. 159; Buss 1995).

7. The Nativism Implied In positing a multitude of computational adaptations along the lines explained, evolutionary psychologists position themselves on the side of nativism in the long-standing struggle between nativism and empiricism about the mind. Indeed, one way of viewing evolutionary psychology is that it basically just *is* the nativist-computational approach, where this approach is, firstly, systematically heuristically guided and organized (see Sect. 2.8) by the modern theory of natural selection (see Sect. 2.6) and, secondly, where this approach is applied to the entire mind rather than only the parts of it which previously existing nativist-computational research communities have already focused upon.

A learning mechanism's receptivity to a specific class of inputs in its initial state, and its disposition to produce specific outputs when presented with them, can, where applicable,[30] be conceived of as primitive forms of innate representing. Together with the procedures that effect the initial transformations of the learning mechanism, these representations can be conceived of as a primitive form of innate knowledge. Depending on the learning mechanism, one might also appropriately view as innate various robustly developing representations, procedures, architectural features, and consequent capacities acquired by the mind in the course of

---

[30] The outputs need not necessarily be representations, they can also be physiological responses; hence the qualification.

subsequent maturation and interaction with the environment. The representational–computational equipment can sensibly count as *the more innate, the more underdetermined the knowledge it embodies is by environmental stimuli*. In saying this, I am implying how I think the concept of innateness needs to be understood—and implicitly is—in the context of cognitive science. Given the controversial status of the concept of innateness, a few more words on it are in order against the background of evolutionary psychology.

The concept of innateness has been argued to be problematic in virtue of promoting the conflation of different biological properties (e.g., Griffiths 2002; Mameli and Bateson 2006), and one cannot nowadays help oneself to the term ‘innate’ without clarifying what one means by it. In my judgement, the understanding of the term by many cognitive scientists is not fundamentally confused, and it is quite appropriate to use the terms ‘innate’ or ‘nativism’ in the context of evolutionary psychology. Indeed, we are missing something very important if we do not do so. Viewed against the backdrop of the history of ideas about the human mind, evolutionary psychology is a definitely nativist enterprise and needs to be understood accordingly.

One must not be misled about the nativist nature of evolutionary psychology by the fact that most of its practitioners hardly use terms like ‘innate’ or ‘nativist’ in describing their posits and their position. Nor even, more importantly, by the fact that theoretical writings of evolutionary psychologists sometimes explicitly disavow commitment to ‘innateness’, carefully scare-quoted. These are decisions of terminological strategy, not of substance. While I side with cognitive scientists like Chomsky or Pinker who cheerfully employ the term ‘innate’ in a considered and enlightened fashion, it is not difficult to understand the motives of researchers who wish to distance themselves from the innateness concept as they understand it. By disavowing a belief in ‘innateness’, evolutionary psychologists seek to distance themselves from a few interrelated, biologically absurd notions about development. Above all, these involve the incoherent doctrine of genetic preformationism—and the naive developmental dichotomy between the innate as that which is present at birth and the learned as everything that is acquired later on.

The notion of innateness I take to be assumed by cognitive scientists implies some guidance of learning by a specialized mechanism which results in knowledge (representational–computational structures) *heavily underdetermined* by the information available from the environment. Since Chomsky (1980a), cognitive capacities thus acquired have often been described as acquired in the face of a ‘poverty of the stimulus’ (Thomas 2002; Laurence and Margolis 2001). This feature of innate capacities being acquired in spite of impoverished environmental stimuli has been widely viewed as placing a substantive constraint on cognitive scientists’ concept of innateness. In fact, it can be viewed as going all the way towards providing a definition. For cognitive structures to be acquired in spite of being heavily underdetermined by environmentally available information plausibly just *is* what cognitive scientists usually mean who describe such structures as ‘innate’. On this, I agree with Khalidi (2002, 2007; cf. Stich 1975).

Theories that take substantial areas of psychological development to be guided by content-specific acquisition mechanisms have been influential already in topically specific subdisciplines of cognitive science. Such research communities have often been outspokenly nativist, endorsing 'innate knowledge' of language (Chomsky 1965, Chap. 1; 1967, 1975), systems of 'initial' or 'core knowledge' in the development of reasoning abilities (Spelke 1994; Spelke and Kinzler 2007), or pursuing an advanced, richly nativist science of vision (with foundations laid by Marr 1982). However, nativist cognitive scientists did not explicitly and systematically embed their research in selectionist evolutionary theory, and they tended to explicitly endorse nativism only for parts of the mind. Most notably, Chomsky, the most influential nativist cognitive scientist of the twentieth century, was not an adaptationist.[31] Even where he extended nativism to cognitive domains other than language, conceiving of the cognitive mind as a system of organs (1980b), adaptationist-evolutionary considerations remained strikingly absent from his argumentation.[32]

Evolutionary psychology's extension of the nativist perspective to *all* areas of the mind includes extending applications of the idea of innate knowledge beyond fields of cognitive science in which it had a strong foothold for decades (Chomskyan linguistics, cognitive developmental psychology, vision science, other sensory modalities), to the vast psychology of motivation. Given the computational–representational perspective together with its attention to biological function, evolutionary psychology can naturally view motivational adaptations as based on machinery that embodies 'innate knowledge' of behaviours that tended to be reproduction-promoting within the environments in which our ancestors evolved. Straightforward examples for this would include incest aversion (innate knowledge about whom not to copulate with) or disgust towards contaminated food (innate knowledge about what not to ingest).

The empiricist approach, in contrast, takes the human mind to develop on the basis of no more initial equipment than the senses, some primitive drives, and very few general-purpose learning mechanisms that treat all kinds of content equally (such as association, statistical induction, or analogy). This leaves the developing child open to be moulded by its social surroundings through learning that is unconstrained in any specific, functional ways. Put differently, the empiricist assumes that the human brain is, functionally, an initially largely undifferentiated organ which acquires content and functional differentiation purely through exposure to experience, 'developing on the basis of uniform principles of growth or learning […] that are common to all domains' (Chomsky 1980b, p. 3). This is also known as the blank slate doctrine, environmentalism, or the 'Standard Social Science Model' (Tooby and Cosmides 1992, pp. 23, 31 f., and passim), in light of its consensus status in the social sciences. Like more confined versions of computational nativism, evolu-

---

[31] Never mind different senses of 'adaptationist' in this connection—Chomsky was not an adaptationist in any sense of the term.

[32] Pinker in particular pointed out the peculiarity of this fact (Pinker and Bloom 1990, Sect. 1; Pinker 1994, 1997) and attempted to rectify it in his work, synthesizing the Chomskyan-nativist view of language with evolutionary adaptationism.

tionary psychology is highly sceptical about the powers of a few uniform learning mechanisms to account for the development of all typically human mental traits.

8. Evolution Through Duplication and Modification Regarding the evolutionary origins of the many hypothesized human computational adaptations, we must be careful not to think of them in developmentally naive terms that make it seem more implausible than it is that they evolved in the available time span. Probably one of the most pernicious ideas to be widely associated with evolutionary psychology is that all computational adaptations, even the potentially highly complex ones, would have had to *evolve from scratch*. It would be putting things mildly to say that this cannot be reasonably expected in the general case. And nothing in evolutionary psychology requires it. What we can more reasonably expect is that many human computational adaptations have roots in the duplication and subsequent modification of genes underlying the neural development of antecedently existing computational adaptations (Marcus 2006, esp. 450 f.; Marcus 2004, Chap. 7; Barrett 2012; Carruthers 2006, Chap. 1).

Inspiration for this idea comes from the relatively young science of evolutionary developmental biology, which took shape in the 1980s and 1990s. Commonly referred to as 'evo-devo', this field has been concerned with elements of the mechanisms of embryonic development qua objects acted upon by evolutionary change (Kirschner and Gerhart 2005; Arthur 2002; Minelli 2009; Carroll et al. 2005). Now, the type of evolutionary change which evo-devo has tended to focus on is the generation of new and novel *forms* in the animal kingdom. But the idea of the duplication and divergence of genes and, through genetic cascades, of entire complex body parts, is one that also seems applicable to mechanisms in the brain. Accordingly, it appears to be applicable to *learning mechanisms* and other computational adaptations. The computational adaptations characteristically posited by evolutionary psychologists may very often rest on mutated and subsequently divergent copies of neurodevelopmental pathways that already earned their keep in more ancient computational adaptations of hominin and hominid ancestors.

9. Shared Brain Areas The computational mechanisms resulting from the maturation of the mind's developmental adaptations can be expected to share many components: many brain areas and circuits. A cognitive scientist who has made a point of advocating this as a fundamental organizational principle of the brain is Michael Anderson (2007, 2010; also see already Marcus 2004; Bechtel 2003). As he states what he calls the redeployment hypothesis, '(i) a typical cognitive function requires the participation of more than one brain area, and (ii) each brain area may be a participant—may be redeployed—in support of other cognitive functions' (2007, p. 148). It is the second subclaim that is intended to be distinctive of redeployment. In later work, Anderson has more expressly emphasized the evolutionary nature of the redeployment claim. It is presented as a thesis about how existing brain areas, originally evolved through natural selection for certain functions, are co-opted for new functions in the further course of evolution: 'The core of the massive redeployment hypothesis is the simple observation that evolutionary considerations might often favour reusing existing components for new tasks over developing new circuits de novo' (2010, p. 246).

It is plain to see that this idea is closely related to the evo-devo idea just endorsed. If new learning mechanisms do not usually evolve from scratch, but rather by the duplication and divergence of genes for existing adaptations, then we cannot be surprised to hear that new adaptations overlap considerably with the older ones, especially in evolutionarily earlier stages of divergence. We cannot reasonably expect, let alone require, that the newly evolving adaptations rest on a wholly different set of brain areas than the pre-existing ones. If neural redeployment is a fundamental organizational principle of the brain, evolutionary psychology is entirely at ease with it.[33]

Nothing in the nature of adaptationist-computational evolutionary psychology requires that each functionally delimitable brain area can serve only one function. Various mechanisms in biology are distinct in spite of sharing functionally relevant components. Why should naturally selected computational mechanisms in the brain not be among them? In fact, as other commentators have already pointed out (Ritchie and Carruthers 2010), a limiting case can even be conceived in which two computational mechanisms share all of their component brain areas, while being nonetheless distinct in virtue of different patterns of connectivity that coordinate the workings of these brain areas differently.

It is only if one is in the grip of a simplistic conception of mental adaptations, implying the discreteness suggested by 'Swiss Army knife' metaphors or talk of mental 'modules', that one can take evolutionary psychology to be threatened by the idea of different acquisition mechanisms or their mature results sharing many brain areas. I take the suggestion of discreteness and neat dissociability to be the single most important reason why it is misleading to refer to mature computational mechanisms posited by evolutionary psychologists as *modules*. This use of the term 'module' is best avoided altogether. Here, I disagree at least terminologically with authors like Carruthers (2006), Clark Barrett (see the papers cited earlier), or Kurzban (2010b). They are happy to accept large overlaps among the mature circuits resulting from different acquisition mechanisms, but nonetheless use the term 'module' for them.

Discussing the applicability of potential modularity concepts to mature systems in the brain is beyond the scope of this chapter. What matters here is that the mental adaptations of evolutionary psychology, qua adaptations, only entail *modularity in development*. They must be based on separate developmental pathways that natural selection could have independently acted upon.

2.8 Evolutionary Psychology's Disciplinary Adaptationism

Having addressed central aspects of evolutionary psychology's ontological adaptationism, let me now turn to the additional component of its adaptationism I have distinguished before: what I call its *disciplinary adaptationism*. Whether or not one

---

[33] A crucial assumption underlying critiques such as those by Bechtel (2003), Prinz (2006) and others is therefore false.

qualifies as adaptationist in this sense depends upon a somewhat more qualitative distinction than ontological adaptationism, rather than upon the more-or-less question of how many mental adaptations one posits.

Three Aspects Evolutionary psychologists are adaptationist in the disciplinary sense in virtue of taking a certain view of how psychology ought to be done, given that ontological adaptationism is true—and in virtue of doing it accordingly. With regard to this practice, it may be reasonable to distinguish the following three aspects of evolutionary psychology:

1. It is *heuristically guided* by hypotheses about persisting selection pressures faced by our hominin ancestors, leading to testable *predictions* about the human mind.
2. It *reorganizes* psychology according to relevant *content domains:* classes of typically environmental constellation and stimulus types within which it hypothesizes selection pressures to have clustered and computational adaptations to have accumulated.
3. It aims at an *explanatory unification* of psychology, based on the jointly pursued goals of identifying the human mind's many computational adaptations and explaining mental/behavioural phenomena through:

   a. the respective operations of these adaptations,
   b. their interplay and interference with each other,
   c. their placement in modern environments,
   d. their by-products, and
   e. their occasional malfunctioning.

In the remainder of this chapter, I take these three aspects up in turn. I conclude by distinguishing disciplinary adaptationism from some other potentially relevant appeals to natural selection.

Heuristic Guidance Towards Predictions Experimental research in evolutionary psychology relies on hypotheses about past selection pressures to generate *non-trivial predictions* about the human mind's adaptations, ranging from the most tentative hypotheses one might consider in specific cases, to heavily confirmed predictions that are directly derivable from general selectionist theories.[34] Evolutionary psychologists have produced significant amounts of confirming evidence for various non-trivial predictions (for one list of examples, see Buss and Reeve 2003, p. 850, Table 1; cf. Buss et al. 1998, 544, Table 1—not all of these examples are actually psychological, but other psychological examples could still be added). The heuristic guidance by modern selectionist theory includes the sometimes nonobvious constraining of hypothesis generation within the discipline.

Many pages in many works of the founding theorists of evolutionary psychology have been filled with variations on the theme that psychological research needs to be guided by questions about biological function. It appears to me that too much

---

[34] Or what Buss (1995, 2008/1999, Chap. 1) calls 'middle-level evolutionary theories'—a term to be understood with caution: It is meant to refer to theories which are still very encompassing and fundamental to the activity of sizeable evolutionary research communities.

ink has been spilled on this. What I take to be fundamentally relevant is this. If you are not an ontological adaptationist to begin with, and place your bets on highly domain-general learning mechanisms instead, then no amount of touting the importance of considering biological function will convince you to pursue psychological research under the heuristic guidance of adaptationist predictions anyway. If on the other hand, you are an ontological adaptationist already, then it is hard to see why it should require a lot of persuading to make you let adaptationist thinking be your guide.

An organization of biological research around the products of natural selection is reasonable because natural selection provides the only explanation for the existence of intricately organized, complex structures, otherwise vanishingly improbable, that are fitted in a vast variety of ways to their specific environments and assembled anew from generation to generation without any external constructor. An understanding of the causal order of the living world needs to be built around the identification and mechanical decomposition of the organized complexity of these structures. What else could it possibly be built around? As Tooby, Cosmides, and Buss routinely point out, the evolutionary process generates only three kinds of biological products: adaptations, by-products of adaptations, and random effects (noise). But identifying traits that are by-products of adaptations presupposes a prior identification of the adaptations they are by-products of. On the other hand, the hypothesis that a given trait is a random product becomes worth considering only once all potential explanations in terms of natural selection or its concomitants have failed. Invoking chance is not an hypothesis from which to start out from in explaining the origin of a trait (Mayr 1983, p. 326). Starting from complex adaptations, one is able to put everything else in its place: simpler adaptations, non-functional by-products of adaptations,[35] and evolutionary products that are random from the viewpoint of natural selection (i.e., noise).

How are we to identify complex adaptations? Criteria such as precision, economy, efficiency, etc. are standardly mentioned in this connection (Williams 1966; Tooby and Cosmides 1992, p. 62; cf. Sommerhoff 1950). But as Andrews et al. (2002, p. 496) point out, we cannot expect to be able to work with such criteria as if with a checklist to identify if some trait is an adaptation or not: ‘Different traits may require satisfaction of different criteria’. In deciding that a psychological capacity is based on a dedicated adaptation, it will be crucial to determine that its acquisition is underdetermined by available stimuli, that the effects it contributes to behaviour would have been reproduction-promoting in ancestral environments, and that it is not better explained as a developmental by-product of a different adaptation known to exist.

---

[35] Also called ‘spandrels’, after Gould and Lewontin (1979). While I do list simple adaptations here separately from complex adaptations, I am adding no further category to the usual triad of adaptations, by-products, and noise. Buss et al. (1998) provide a more in-depth treatment, including discussion of the putatively separate category of adaptations co-opted for other purposes (widely called ‘exaptations’, since Gould and Vrba 1982).

Reorganization by Content Domains The many computational adaptations posited in evolutionary psychology can be arranged according to different content domains they belong to—domains of constellations, properties, and entities whose causal potentials and regularities would have led, over many generations, to the natural selection of causal regularities in information-processing adaptations that meshed with those contents in reproduction-promoting ways. These content domains were traditionally not recognized in the disciplinary subdivisions of mainstream psychology. Evolutionary psychologists endorse the reorganization of psychology according to selectionally relevant content domains.

A concrete idea of a potential reorganization of psychology along such lines was provided already two decades ago by Steven Pinker, buried deep inside his book on the language faculty. Presenting a list that has aged well enough to provide a good starting base for additions, modifications, further subdivisions and perhaps fusions, Pinker (1994, p. 418) speculated that psychology might fall into the following topical domains:[36] perception; language (these two vast areas being the only domains already recognized in standard psychology); intuitive mechanics; intuitive biology; numerical cognition; mental maps for large territories; habitat selection; responses to danger, including various fears or phobias and 'a motive to learn the circumstances in which each is harmless'; food recognition: 'what is good to eat'; contamination avoidance; monitoring of current well-being, 'including the emotions of happiness and sadness, and moods of contentment and restlessness'; intuitive psychology; a database for individuals; a 'self-concept: gathering and organizing information about one's value to other people, and packaging it for others'; justice, understood as involving a 'sense of rights, obligations, and deserts, including the emotions of anger and revenge'; kinship-related behaviour, 'including nepotism and allocations of parenting effort'; and finally mating, 'including feelings of sexual attraction, love, and intentions of fidelity and desertion'.

This assortment of domains and adaptations at least gives a general idea of the type of alternative disciplinary organization that evolutionary psychologists envisage. It is easily conceivable that this kind of organization according to selectionally relevant content domains would seem unremarkable to beginners in psychology who come to the academic field without any preconceptions about it. But it is crucial to appreciate the content-specific organization against the background of the traditionally established subdivisions of psychology. Again, Pinker describes matters aptly:

> To see how far standard psychology is from this conception, just turn to the table of contents of any textbook. The chapters will be: Physiological, Learning, Memory, Attention, Thinking, Decision-Making, Intelligence, Motivation, Emotion, Social, Development, Personality, Abnormal. I believe that with the exception of Perception and, of course, Language, not

---

[36] I am deliberately dropping Pinker's descriptions of these on pp. 419–421 as 'modules'. This term is inappropriate to Pinker's own treatment (apart from other problematic aspects that attach to it anyhow). For apart from a few examples which might be described as corresponding to single modules, he is talking about entire domains of reality for which large bundles of such adaptations would be expected to have evolved. Of course, perception in toto is not a single mental 'module'.

> a single curriculum unit in psychology corresponds to a cohesive chunk of the mind. [37] Perhaps this explains the syllabus-shock experienced by Introductory Psychology students. It is like explaining how a car works by first discussing the steel parts, then the aluminium parts, then the red parts, and so on, instead of the electrical system, the transmission, the fuel system, and so on. (Pinker 1994, p. 421)

Explanatory Unification The unification of psychology envisaged by evolutionary psychology consists in the fact that all causal explanations of psychobehavioural phenomena are grounded in explanations that reach into the evolutionary past, appealing to the same evolutionary-selectionist principles. Thus, in explanation, as in the heuristic guidance of predictions, adaptation is the organizing concept for everything else in psychology. Complete explanations of course include specifying the acquisition mechanisms on which natural selection acted so that they would mesh with developmental environments to produce reproduction-promoting results.

Contrary to popular beliefs encouraged by its detractors, evolutionary psychology's adaptationist explanations are not limited to offering *direct* explanations for psychobehavioural phenomena by hypothesizing adaptations which evolved to produce those very effects as their proper functions. Nor even are evolutionary psychology's adaptationist explanations restricted to direct explanations, plus explanations in terms of developmental by-products of adaptations. As indicated in the list offered above, highly nonobvious, deep explanations standardly examined by evolutionary psychologists also invoke: information-processing adaptations malfunctioning in certain ways (with distinctive effects); the placement of such adaptations in modern environments differing significantly from the environments the adaptations evolved to mesh with (with distinctive effects); and the effects of nonobvious interplays and interferences between different information-processing adaptations. My main focus here, however, is not with the specifics of these different kinds of adaptationist explanation, but with the unification of psychology under the adaptationist method that offers these shared explanatory resources.

As with the aspect of disciplinary reorganization, to appreciate the aspect of explanatory unification in evolutionary psychology's disciplinary adaptationism, one needs to be aware of the nature of standard experimental psychology with which it contrasts. Anyone unfamiliar with the curriculum and research practice of standard psychology can easily underestimate how rare it has traditionally been for deep and principled explanations to be put forth in this vast field, and how often experimentation and data collection proceed without substantive theory to guide them (for a lucid and concise critique, see Gigerenzer 1998, 2010). Where local theories exist, they tend to be unconnected with each other. This explanatory disunity of standard psychology was a source of great dissatisfaction for evolutionary psychologists of

---

[37] One must not be misled by the suggestive words 'cohesive chunk of the mind' into thinking of domain-dedicated spatial sectors of the brain, such as those familiar from nineteenth-century phrenology and its fantasy-based brain maps. See, by comparison, Pinker (1997 p. 30 f.) for some often-cited comments on the non-necessity of spatial contiguity for mental adaptations.

the first generation.[38] To quote the opening sigh of Buss's most widely cited programmatic paper on evolutionary psychology:

> Anyone familiar with the broad field of psychology knows that it is in theoretical disarray. The different branches—such as cognitive, social, personality, and developmental—proceed in relative isolation from one another, at most occasionally borrowing like a cup of sugar a concept here and a method there from a neighbor. Within each branch, psychologists also fail to reach consensus. Mini-theories proliferate unconnected, each conceived to account for a particular set of phenomena, such as obedience to authority, children's concepts of mind, or the effects of priming on categorization tasks. […] [N]o metatheory subsumes, integrates, unites, or connects the disparate pieces that psychologists gauge with their differing calipers. (1995, p. 1)

Pinker (2005b), recollecting how experimental psychology was taught to him in his student days, bemoans the same disunity. In a passage illuminating enough to warrant one final further quotation of some length, he describes his frustration as a beginning psychology student about his chosen field, which avoided 'any of the topics that attracted [students] to the subject'. As he writes:

> When I proceeded to more advanced courses, they only deepened the disappointment by revealing that the psychology canon was a laundry list of unrelated phenomena. The course on perception began with Weber's Law and Fechner's Law and proceeded to an assortment of illusions and aftereffects familiar to readers of cereal boxes. […] Cognitive psychology, too, consisted of laboratory curiosities analyzed in terms of dichotomies such as serial/parallel, discrete/analog, and top-down/bottom-up […]. To this day, social psychology is driven not by systematic questions about the nature of sociality in the human animal but by a collection of situations in which people behave in strange ways.
>
> But the biggest frustration was that psychology seemed to lack any sense of explanation. […] [P]sychologists were content to 'explain' a phenomenon by redescribing it. (2005b, p. xi)

The frustration about disunity described here is of a piece with the 'syllabus shock' of students referred to in Pinker's earlier criticism of standard psychology's content-neutral organization. The passage just quoted also criticizes standard psychology for being explanatorily shallow to the point of vacuity. These misgivings about explanatory disunity and lack of explanatory depth are connected, as I hope to have shown. Evolutionary psychology's disciplinary adaptationism addresses them both.

Distinction from Other Appeals to Natural Selection Having explained disciplinary adaptationism and the ontological thesis it entails, let me contrast it with other potential ways of appealing to natural selection in connection with the mind. The following distinctions are clearly not exhaustive. The first two 'adaptationisms' are defined by me through features that could be recombined to yield intermediate types. In any event, the generalized descriptions involve inevitable oversimplifications.

1. The disciplinary adaptationism of evolutionary psychology is obviously as different as can be from what might be called *dilettantic adaptationism*, if a term for it be needed. This is not a methodological doctrine, but an intellectual tendency.

---

[38] See also Silverman (2003, p. 2).

Dilettantic adaptationism, in its pure form, is uninformed by evolutionary biology (and, typically, it is not strong on information processing either). It occasionally offers explanations of human psychobehavioural phenomena in terms of intuitive notions of function and naive conceptions of group selection—appeals to 'the good of the species'—but never explanations based on population thinking along the broad lines of modern selectionist theory. Often, the absence of a scientifically informed adaptationist perspective is not due to an actual endorsement of anti-adaptationism, but simply to a lack of exposure to evolutionary science. One of my favourite, or least favourite, examples of dilettantic adaptationism in psychology is the reliance on an intuitive notion of function in the idea that 'happiness' or 'well-being' are somehow intrinsically 'functional'—as though natural selection cared to maximize human happiness.

2. We might next distinguish a more serious tendency to appeal to natural selection with regard to human psychology: what one might call *sporadic adaptationism*. Psychologists and other researchers grouped under this heading sometimes take selectionist evolutionary theory into consideration, but they do not allow it to guide their research heuristically. Crucially, they do not use selectionist theory to generate and test adaptationist predictions. Moreover, the sporadic adaptationist continues to adhere to the traditional organization of psychological science along the lines of content-neutral faculties, rather than along the lines of typically ecological domains within which natural selection accumulated information-processing adaptations. As far as I can see, the only interesting justification someone might have for not being anything more than a sporadic adaptationist is that she does not endorse ontological adaptationism to a sufficient degree. The connection is clear enough. The fewer content-specific mental adaptations somebody believes to exist, the less sense it makes to organize one's psychological research around the search for them.

3. Thirdly, let me set aside the idea of a '*merely methodological*' version of a research program, as opposed to an ontologically committed version. Foundational analyses of scientific or philosophical research programs often contrast ontological and methodological varieties of a research program. Both varieties work within the framework of an overarching theory and both endorse similar research methodologies based on that theory. What distinguishes them are their ontological commitments. The ontological adherents to the theory in question take its apparent existence commitments at face value (which evidently mandates pursuing research accordingly). The methodological adherents to the theory in question only think it is a good idea to pursue research as if it were true, whether it really is or not. It will be clear by now that this sense of 'methodological', used for ontologically attenuated endorsements of research-guiding hypotheses, is emphatically not the sense in which evolutionary psychologists are adaptationists about the mind.

For this reason, even though the aspects I distinguished under the heading of disciplinary adaptationism very much concern psychology's methodology, I am not using the label 'methodological' to describe disciplinary adaptationism here. The term 'methodological adaptationism', given what I just said, is naturally understood as designating a weaker thesis than ontological adaptationism. In contrast, evolutionary psychologists' disciplinary adaptationism necessarily presupposes a robust degree of ontological adaptationism about the mind. Without the assumption that there are many mental adaptations waiting to be discovered in the first place, it would be incoherent for evolutionary psychologists to endorse the research agendas they do, and to be revisionist about the way that psychology is to be organized and practiced. These activities would be rendered as pointless as any systematic search for something one does not expect to exist.

4. One last set of remarks will be necessary regarding a trichotomy of kinds of adaptationism due to Peter Godfrey-Smith (2001), which has been hugely influential among philosophers of biology. Godfrey-Smith, inspired by similar distinctions applied to behaviourism in psychology, recognizes three kinds of adaptationism: ‘empirical’ (in effect: ontological), ‘explanatory’, and ‘methodological’ adaptationism. He claims that each of these could be consistently held without holding any of the other two (pp. 338–342), though he acknowledges various relations of support between them. Discussing Godfrey-Smith’s justification for his tripartition of adaptationism is beyond the scope of this chapter, but at least a few remarks are in order as to why it is not adopted here.

For one thing, I would argue that his definition of the ontological thesis is not only too strong to capture a relevant notion among researchers in evolutionary theory, but way too strong to capture the notion of ontological adaptationism relevant to evolutionary psychology. For another thing, I would argue that Godfrey-Smith’s separation of a position of explanatory adaptationism is altogether indefensible. Ultimately, the notion either collapses into ontological adaptationism, or it reduces to a quasi-aesthetic sense of subjective impressedness.

Methodological adaptationism as defined by Godfrey-Smith makes the most sense among the three ideas he distinguishes, and it is useful for biology in general: ‘The best way for scientists to approach biological systems is to look for features of adaptation and good design. Adaptation is a good “organizing concept” for evolutionary research’ (2001, p. 337). The notion of methodological adaptationism thus defined is consistent with what I have called the disciplinary adaptationism of evolutionary psychology. But it does not capture all aspects of the latter. Nor is it tailored to the specific concerns of evolutionary psychology and to how it contrasts with standard psychology. Moreover, whatever sense might be made of Godfrey-Smith’s mutual separation of methodological and ontological adaptationism for biology in general, it is not applicable to psychology. The relevant ontological thesis for evolutionary psychology, I maintain, is simply the ontological adaptationism defined and fleshed out earlier. And, as explained before, this is a necessary part of disciplinary adaptationism.

Acknowledgments Supported by grant DA 1282/1-1 from the German Research Foundation (DFG) for the author’s project *Philosophy of evolutionary psychology*. I thank Steve Pinker for email correspondence on computationalism.

References

Alcock, J. (2001). *The triumph of sociobiology*. New York: Oxford University Press.

Anderson, M. L. (2007). The massive redeployment hypothesis and the functional topography of the brain. *Philosophical Psychology, 20,* 143–174.

Anderson, M. L. (2010). Neural reuse: A fundamental organizational principle of the brain. *Behavioral and Brain Sciences, 33,* 245–266.

Andrews, P. W., Gangestad, S. W., & Matthews, D. (2002). Adaptationism—How to carry out an exaptationist program. *Behavioral and Brain Sciences, 25,* 489–504.

Angell, J. R. (1907). The province of functional psychology. *Psychological Review, 14,* 61–91.

Arthur, W. (2002). The emerging conceptual framework of evolutionary developmental biology. *Nature, 415,* 757–764.

Barkow, J. H., Cosmides, L., & Tooby, J. (Eds.). (1992). *The adapted mind: Evolutionary psychology and the generation of culture*. New York: Oxford University Press.

Barrett, H. C. (2005). Enzymatic computation and cognitive modularity. *Mind and Language, 20,* 259–287.

Barrett, H. C. (2006). Modularity and design reincarnation. In P. Carruthers, S. Laurence, & S. Stich (Eds.), *The innate mind, vol. II: Culture and cognition* (pp. 199–217). New York: Oxford University Press.

Barrett, H. C. (2012). A hierarchical model of the evolution of human brain specializations. *Proceedings of the National Academy of Sciences, 109,* 10733–10740.

Barrett, H. C., & Kurzban, R. (2006). Modularity in cognition: Framing the debate. *Psychological Review, 113,* 628–647.

Baumeister, R. F., Vohs, K. D., & Funder, D. C. (2007). Psychology as the science of self-reports and finger movements: Whatever happened to actual behavior? *Perspectives on Psychological Science, 2,* 396–403.

Bechtel, W. (2003). Modules, brain parts, and evolutionary psychology. In S. J. Scher & F. Rauscher (Eds.), *Evolutionary psychology: Alternative approaches* (pp. 211–227). Dordrecht: Kluwer.

Bechtel, W., & Abrahamsen, A. (1991). *Connectionism and the mind: An introduction to parallel processing in networks*. Oxford: Blackwell.

Bechtel, W., & Richardson, R. C. (2010/1993). *Discovering complexity: Decomposition and localization as strategies in scientific research* (2nd ed.). Cambridge: MIT Press.

Bechtel, W., Abrahamsen, A., & Graham, G. (1998). The life of cognitive science. In W. Bechtel & G. Graham (Eds.), *A companion to cognitive science* (pp. 1–104). Oxford: Blackwell.

Block, N. (1995). The mind as the software of the brain. In E. E. Smith & D. N. Osherson (Eds.), *An invitation to cognitive science* (2nd ed.). *Volume 3: Thinking* (pp. 377–425). Cambridge: MIT Press.

Boakes, R. (1984). *From Darwin to behaviourism: Psychology and the minds of animals*. Cambridge: Cambridge University Press.

Bolhuis, J. J., & Giraldeau, L.-A. (2005). The study of animal behaviour. In J. J. Bolhuis & L.-A. Giraldeau (Eds.) *The behaviour of animals: mechanisms, function, and evolution* (pp. 1–9). Oxford: Blackwell.

Bowler, P. J. (2003/1983). *Evolution: The History of an Idea* (3rd ed., completely revised and expanded). Berkeley: University of California Press.

Boyer, P., & Barrett, H. C. (2005). Domain-specificity and intuitive ontology. In D. M. Buss (Ed.), *The handbook of evolutionary psychology* (pp. 96–118). Hoboken: Wiley.

Buller, D. J. (2005). *Adapting minds: Evolutionary psychology and the persistent quest for human nature*. Cambridge: MIT Press.

Buss, D. M. (1989). Sex differences in human mate preferences: Evolutionary hypotheses tested in 37 cultures. *Behavioral and Brain Sciences, 12,* 1–14.

Buss, D. M. (1995). Evolutionary psychology: A new paradigm for psychological science. *Psychological Inquiry, 6,* 1–30.

Buss, D. M. (2008/1999). *Evolutionary psychology: The new science of the mind* (3rd ed.). Boston: Allyn and Bacon.

Buss, D. M., & Barnes, M. (1986). Preferences in human mate selection. *Journal of Personality and Social Psychology, 50,* 559–570.

Buss, D. M., & Reeve, H. K. (2003). Evolutionary psychology and developmental dynamics: Comment on Lickliter and Honeycutt. *Psychological Bulletin, 129,* 848–853.

Buss, D. M., Haselton, M. G., Shackelford, T. K., Bleske, A. L., & Wakefield, J. C. (1998). Adaptations, exaptations, and spandrels. *American Psychologist, 53,* 533–548.

Carroll, S. B., Grenier, J. K., & Weatherbee, S. D. (2005). *From DNA to diversity: Molecular genetics and the evolution of animal design* (2nd ed.). Oxford: Blackwell.

Carruthers, P. (2006). *The architecture of the mind: Massive modularity and the flexibility of thought*. New York: Oxford University Press.

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge: MIT Press.

Chomsky, N. (1967). Recent contributions to the theory of innate ideas. *Synthese, 17,* 2–11.

Chomsky, N. (1975). *Reflections on language*. New York: Pantheon.

Chomsky, N. (1980a). *Rules and representations*. New York: Columbia University Press.

Chomsky, N. (1980b). Rules and representations. *Behavioral and Brain Sciences, 3,* 1–15.

Clark, A. (1989). *Microcognition: Philosophy, cognitive science, and parallel distributed processing*. Cambridge: MIT Press.

Cosmides, L. (1989). The logic of social exchange: Has natural selection shaped how humans reason? Studies with the Wason selection task. *Cognition, 31,* 187–276.

Cosmides, L., & Tooby, J. (1987). From evolution to behaviour: Evolutionary psychology as the missing link. In J. Dupré (Ed.), *The latest on the best: Essays on evolution and optimality* (pp. 277–306). Cambridge: MIT Press.

Cosmides, L., & Tooby, J. (1994). Beyond intuition and instinct-blindness: Toward an evolutionarily rigorous cognitive science. *Cognition, 50,* 41–77.

Cosmides, L., & Tooby, J. (1997). Dissecting the computational architecture of social inference mechanisms. *Characterizing human psychological adaptations. Ciba Foundation Symposium, Vol. 208* (pp. 132–161). Chichester: Wiley.

Cosmides, L., Tooby, J., & Barkow, J. H. (1992). Introduction: Evolutionary psychology and conceptual integration. In J. H. Barkow, L. Cosmides, & J. Tooby (Eds.), *The adapted mind: Evolutionary psychology and the generation of culture* (pp. 3–15). New York: Oxford University Press.

Crawford, C., Smith, M., & Krebs, D. (Eds.). (1987). *Sociobiology and psychology: Ideas, issues, and applications*. Hillsdale: Lawrence Erlbaum.

Cummins, R. (1989). *Meaning and mental representation*. Cambridge: MIT Press.

Daly, M., & Wilson, M. (1983/1978). *Sex, evolution, and behavior* (2nd ed.). Belmont: Wadsworth.

Daly, M., & Wilson, M. (1988). *Homicide*. New Brunswick: Transaction.

Daly, M., & Wilson, M. (1995). Evolutionary psychology: Adaptationist, selectionist, and comparative. [Comment on Buss 1995.] *Psychological Inquiry, 6,* 34–38.

Daly, M., & Wilson, M. (1999). Human evolutionary psychology and animal behaviour. *Animal Behaviour, 57,* 509–519.

Daly, M., & Wilson, M. (2005). Human behaviour as animal behaviour. In J. J. Bolhuis & L.-A. Giraldeau (Eds.), *The behaviour of animals: Mechanisms, function, and evolution* (pp. 393–408). Oxford: Blackwell.

Daly, M., & Wilson, M. (2008). Is the "Cinderella effect" controversial? A case study of evolution-minded research and critiques thereof. In C. Crawford & D. Krebs (Eds.), *Foundations of evolutionary psychology* (pp. 383–400). New York: Lawrence Erlbaum.

Darwin, C. (1859). *On the origin of species by means of natural selection* (1st ed.). London: John Murray. http://darwin-online.org.uk Accessed 24 Nov 2013.

Dawkins, R. (1985). Review of Lewontin, Rose, & Kamin, *Not in our genes: Biology, ideology, and human nature*. *New Scientist*, 24 January 1985.

Dawkins, R. (1989). Preface to the 2nd edition of *The selfish gene* (1976/1989) (pp. xv–xviii). Oxford: Oxford University Press.

Dawkins, R. (2003). Unfinished correspondence with a Darwinian heavyweight. In R. Dawkins (Ed.), *A devil's chaplain: Selected essays* (pp. 256–261). London: Orion Books.

Dawkins, R. (2005). Afterword. In D. M. Buss (Ed.), *The handbook of evolutionary psychology* (pp. 975–979). Hoboken: Wiley.

Dobzhansky, T. (1937). *Genetics and the origin of species*. New York: Columbia University Press.

Dunbar, R., & Barrett, L. (2007). Evolutionary psychology in the round. In R. Dunbar & L. Barrett (Eds.), *The Oxford handbook of evolutionary psychology* (pp. 3–9). New York: Oxford University Press.

Evans, N., & Levinson, S. C. (2009). The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and Brain Sciences, 32,* 429–448.

Fodor, J. A. (1975). *The language of thought*. Cambridge: Harvard University Press.

Fodor, J. A. (1980). Methodological solipsism considered as a research strategy in cognitive psychology. *Behavioral and Brain Sciences, 3,* 63–73.

Fodor, J. A. (1981). *Representations: Philosophical essays on the foundations of cognitive science*. Cambridge: MIT Press.

Fodor, J. A. (1983). *The modularity of mind: An essay on faculty psychology*. Cambridge: MIT Press.

Fodor, J. A. (2000). *The mind doesn't work that way: The scope and limits of computational psychology*. Cambridge: MIT Press.

Fodor, J. A., & Pylyshyn, Z. W. (1988). Connectionism and cognitive architecture: A critical analysis. Reprinted in C. Macdonald & G. Macdonald (Eds. 1995), *Connectionism: Debates on psychological explanation, Vol. II* (pp. 90–163). Oxford: Blackwell.

Gigerenzer, G. (1998). Surrogates for theories. *Theory and Psychology, 8,* 195–204.

Gigerenzer, G. (2010). Personal reflections on theory and psychology. *Theory and Psychology, 20,* 733–743.

Gigerenzer, G., & Goldstein, D. G. (1996). Mind as computer: The birth of a metaphor. *Creativity Research Journal, 9,* 131–144.

Godfrey-Smith, P. (2001). Three kinds of adaptationism. In S. H. Orzack & E. Sober (Eds.), *Adaptationism and optimality* (pp. 335–357). Cambridge: Cambridge University Press.

Gould, S. J., & Lewontin, R. C. (1979). The spandrels of San Marco and the Panglossian paradigm: A critique of the adaptationist program. *Proceedings of the Royal Society of London B, 205,* 581–598.

Gould, S. J., & Vrba, E. S. (1982). Exaptation: A missing term in the science of form. *Paleobiology, 8,* 4–15.

Griffiths, P. E. (2002). What is innateness? *The Monist, 85,* 70–85.

Hagen, E. (2005). Controversial issues in evolutionary psychology. In D. M. Buss (Ed.), *The handbook of evolutionary psychology* (pp. 145–173). Hoboken: Wiley.

Hamilton, W. D. (1964a). The genetical theory of social behaviour, I. *Journal of Theoretical Biology, 7,* 1–16.

Hamilton, W. D. (1964b). The genetical theory of social behaviour, II. *Journal of Theoretical Biology, 7,* 17–52.

Hauser, M. D., Chomsky, N., & Fitch, W. T. (2002). The faculty of language: What is it, who has it, and how did it evolve? *Science, 298,* 1569–1579.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences, 33,* 61–83.

Heyes, C. (2000). Evolutionary psychology in the round. In C. Heyes, & L. Huber (Eds.), *The evolution of cognition* (pp. 3–22). Cambridge: MIT Press.

Heyes, C., & Huber, L. (Eds.). (2000). *The evolution of cognition*. Cambridge: MIT Press.

Holcomb, H. R. (2005). Review of D. J. Buller, *Adapting minds*. *Evolutionary Psychology, 3,* 392–401.

Holland, J. H., Holyoak, K. J., Nisbett, R. E., & Thagard, P. R. (1986). *Induction: Processes of inference, learning, and discovery*. Cambridge: MIT Press.

Huxley, J. (1942). *Evolution: The modern synthesis*. London: Allen and Unwin.

James, W. (1890). *Principles of psychology*. Cambridge: Harvard University Press.

Kenrick, D. T. (1995). Evolutionary theory versus the confederacy of dunces. *Psychological Inquiry, 6,* 56–62.

Ketelaar, T., & Ellis, B. J. (2000). Are evolutionary explanations unfalsifiable? Evolutionary psychology and the Lakatosian philosophy of science. *Psychological Inquiry, 11,* 1–21.

Khalidi, M. A. (2002). Nature and nurture in cognition. *British Journal for the Philosophy of Science, 53,* 251–272.

Khalidi, M. A. (2007). Innate cognitive capacities. *Mind and Language, 22,* 92–115.

Kirschner, M. W., & Gerhart, J. C. (2005). *The plausibility of life: Resolving Darwin's dilemma*. New Haven: Yale University Press.

Kurzban, R. (2002). Alas, poor evolutionary psychology: Unfairly accused, unjustly condemned. *Human Nature Review, 2,* 99–109.

Kurzban, R. (2010a). Grand challenges of evolutionary psychology. *Frontiers in Psychology, 1,* 1–3.

Kurzban, R. (2010b). *Why everyone (else) is a hypocrite: Evolution and the modular mind*. Princeton: Princeton University Press.

Laland, K. N., & Brown, G. R. (2011/2002): *Sense and nonsense: Evolutionary perspectives on human behaviour* (2nd ed.). Oxford: Oxford University Press.

Laurence, S., & Margolis, E. (2001). The poverty of the stimulus argument. *British Journal for the Philosophy of Science, 52,* 217–276.

Macdonald, C., & Macdonald, G. (Eds.). (1995). *Connectionism: Debates on psychological explanation, Vol. II.* Oxford: Blackwell.

Machery, E., & Barrett, H. C. (2006). Debunking *Adapting minds*. *Philosophy of Science, 73,* 232–246.

Mameli, M., & Bateson, P. (2006). Innateness and the sciences. *Biology and Philosophy, 21,* 155–188.

Mandler, G. (2007). *A history of modern experimental psychology: From James and Wundt to cognitive science*. Cambridge: MIT Press.

Marcus, G. F. (2001). *The algebraic mind: Integrating connectionism and cognitive science*. Cambridge: MIT Press.

Marcus, G. F. (2004). *The birth of the mind: How a tiny number of genes creates the complexities of human thought*. New York: Basic Books.

Marcus, G. F. (2006). Cognitive architecture and descent with modification. *Cognition, 101,* 443–465.

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. San Francisco: Freeman.

Mayr, E. (1982). *The growth of biological thought: Diversity, evolution, and inheritance*. Cambridge: Harvard University Press.

Mayr, E. (1983). How to carry out the adaptationist program? *American Naturalist, 121,* 324–334.

Mayr, E. (1991). *One long argument: Charles Darwin and the genesis of modern evolutionary thought*. Cambridge: Harvard University Press.

Mayr, E., & Provine, W. B. (Eds.). (1980). *The evolutionary synthesis: Perspectives on the unification of biology*. Cambridge: Harvard University Press.

McClelland, J. L., Rumelhart, D. E., & the PDP Research Group. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition, Vol. 2: Psychological and biological models*. Cambridge: MIT Press.

McDougall, W. (1908). *An introduction to social psychology*. London: Methuen.

Miller, G. (2000). How to keep our metatheories adaptive: Beyond Cosmides, Tooby, and Lakatos. [Comment on Ketelaar & Ellis 2000.] *Psychological Inquiry, 11,* 42–46.

Minelli, A. (2009). *Forms of becoming: The evolutionary biology of development*. Princeton: Princeton University Press.

Mundale, J. (2002). Concepts of localization: Balkanization in the brain. *Brain and Mind, 3,* 313–330.

Neisser, U. (1967). *Cognitive psychology*. East Norwalk: Appleton-Century-Crofts.

Newell, A. (1980). Physical symbol systems. *Cognitive Science, 4,* 135–183.

Newell, A., & Simon, H. A. (1976). Computer science as empirical inquiry: Symbols and search. *Communications of the Association for Computing Machinery, 19,* 113–126.

Pigliucci, M. (2007). Do we need an extended evolutionary synthesis? *Evolution, 61,* 2743–2749.

Pigliucci, M., & Müller, G. B. (Eds.). (2010). *Evolution: The extended synthesis*. Cambridge: MIT Press.

Pinker, S. (1994). *The language instinct*. London: Penguin.

Pinker, S. (1997). *How the mind works*. London: Penguin.

Pinker, S. (2002). *The blank slate: The modern denial of human nature*. London: Penguin.

Pinker, S. (2003). Language as an adaptation to the cognitive niche. In M. H. Christiansen & S. Kirby (Eds.), *Language evolution: States of the art* (pp. 16–37). New York: Oxford University Press.

Pinker, S. (2005a). So how *does* the mind work? *Mind and Language, 20,* 1–24.

Pinker, S. (2005b). Foreword. In D. M. Buss (Ed.), *The handbook of evolutionary psychology* (pp. xi–xvi). Hoboken: Wiley.

Pinker, S., & Bloom, P. (1990). Natural language and natural selection. *Behavioral and Brain Sciences, 13,* 707–726.

Pinker, S., & Jackendoff, R. (2005). The faculty of language: What's special about it? *Cognition, 95,* 201–236.

Platek, S. M., & Shackelford, T. K. (Eds.). (2009). *Foundations in evolutionary cognitive neuroscience*. Cambridge: Cambridge University Press.

Prinz, J. J. (2006). Is the mind really modular? In R. J. Stainton (Ed.), *Contemporary debates in cognitive science* (pp. 22–36). Oxford: Blackwell.

Rauscher, F., & Scher, S. J. (2003a). Alternative approaches to evolutionary psychology: Introduction. In S. J. Scher & F. Rauscher (Eds.), *Evolutionary psychology: Alternative approaches* (pp. xi–xviii). Dordrecht: Kluwer.

Ritchie, J. B., & Carruthers, P. (2010). Massive modularity is consistent with most forms of neural reuse. [Comment on Anderson 2010.] *Behavioral and Brain Sciences, 33,* 289–290.

Rumelhart, D. E., McClelland, J. L., & the PDP Research Group. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition, Vol. 1: Foundations*. Cambridge: MIT Press.

Scher, S. J., & Rauscher, F. (2003a). Nature read in truth or flaw: Locating alternatives in evolutionary psychology. In S. J. Scher & F. Rauscher (Eds.), *Evolutionary psychology: Alternative approaches* (pp. 1–26). Dordrecht: Kluwer.

Scher, S. J., & Rauscher, F. (Eds.). (2003b). *Evolutionary psychology: Alternative approaches*. Dordrecht: Kluwer.

Schmitt, D. P. (2008). Evolutionary psychology research methods. In C. Crawford & D. Krebs (Eds.), *Foundations of evolutionary psychology* (pp. 215–236). New York: Lawrence Erlbaum.

Schmitt, D. P., & Pilcher, J. J. (2004). Evaluating evidence of psychological adaptation: How do we know one when we see one? *Psychological Science, 15,* 643–649.

Segerstråle, U. (2000). *Defenders of the truth: The battle for science in the sociobiology debate and beyond*. Oxford: Oxford University Press.

Selfridge, O. G. (1959). Pandemonium: A paradigm for learning. In *Mechanisation of Thought Processes,* proceedings of National Physical Laboratory symposium no. 10. London: HM Stationery Office.

Selfridge, O. G., & Neisser, U. (1960). Pattern recognition by machine. *Scientific American, 203,* 60–68.

Silverman, I. (2003). Confessions of a closet sociobiologist: Personal perspectives on the Darwinian movement in psychology. *Evolutionary Psychology, 1,* 1–9.

Simons, K. (Ed.). (1993). *Early visual development: Normal and abnormal*. New York: Oxford University Press.

Simpson, J. A., & Campbell, L. (2005). Methods of evolutionary sciences. In D. M. Buss (Ed.), *The handbook of evolutionary psychology* (pp. 119–144). Hoboken: Wiley.

Smocovitis, V. B. (1996). *Unifying biology: The evolutionary synthesis and evolutionary biology*. Princeton: Princeton University Press.

Sommerhoff, G. (1950). *Analytical biology*. Oxford: Oxford University Press.

Spelke, E. S. (1994). Initial knowledge: Six suggestions. *Cognition, 50,* 431–445.

Spelke, E. S., & Kinzler, K. D. (2007). Core knowledge. *Developmental Science, 10,* 89–96.

Sperber, D. (1994). The modularity of thought and the epidemiology of representations. In L. A. Hirschfeld & S. A. Gelman (Eds.), *Mapping the mind: Domain-specificity in cognition and culture* (pp. 39–67). Cambridge: Cambridge University Press.

Stich, S. P. (1975). The idea of innateness. In S. P. Stich (Ed.), *Innate ideas* (pp. 1–22). Berkeley: University of California Press.

Symons, D. (1979). *The evolution of human sexuality*. New York: Oxford University Press.

Symons, D. (1987). If we're all Darwinians, what's the fuss about? In C. Crawford, M. Smith, & D. Krebs (Eds.), *Sociobiology and psychology: Ideas, issues, and applications* (pp. 121–146). Hillsdale: Lawrence Erlbaum.

Symons, D. (1989). A critique of Darwinian anthropology. *Ethology and Sociobiology, 10,* 131–144.

Thomas, M. (2002). Development of the concept of "the poverty of the stimulus". *Linguistic Review, 18,* 51–71.

Tinbergen, N. (1963). On aims and methods in Ethology. *Zeitschrift für Tierpsychologie, 20,* 410–433.

Tomasello, M. (1995). Language is not an instinct. [Review of Pinker, *The Language Instinct*.] *Cognitive Development, 10,* 131–156.

Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Cambridge: Harvard University Press.

Tomasello, M. (2009). Universal grammar is dead. [Comment on Evans & Levinson 2009.] *Behavioral and Brain Sciences, 32,* 470–471.

Tooby, J., & Cosmides, L. (1989). Evolutionary psychology and the generation of culture, Part I: Theoretical considerations. *Ethology and Sociobiology, 10,* 29–49.

Tooby, J., & Cosmides, L. (1990a). On the universality of human nature and the uniqueness of the individual: The role of genetics and adaptation. *Journal of Personality, 58,* 17–67.

Tooby, J., & Cosmides, L. (1990b). The past explains the present: Emotional adaptations and the structure of ancestral environments. *Ethology and Sociobiology, 11,* 375–424.

Tooby, J., & Cosmides, L. (1992). The psychological foundations of culture. In J. H. Barkow, L. Cosmides, & J. Tooby (Eds.), *The adapted mind: Evolutionary psychology and the generation of culture* (pp. 19–136). New York: Oxford University Press.

Tooby, J., & Cosmides, L. (2005). Conceptual foundations of evolutionary psychology. In D. M. Buss (Ed.), *The handbook of evolutionary psychology* (pp. 5–67). Hoboken: Wiley.

Tooby, J., Cosmides, L., & Barrett, H. C. (2003). The second law of thermodynamics is the first law of psychology: Evolutionary developmental psychology and the theory of tandem, coordinated inheritances: Comment on Lickliter and Honeycutt. *Psychological Bulletin, 129,* 858–865.

Tybur, J. M., Miller, G. F., & Gangestad, S. W. (2007). Testing the controversy: An empirical examination of adaptationists' attitudes toward politics and science. *Human Nature, 18,* 313–328.

Waddington, C. H. (1957). *The strategy of the genes*. London: Allen & Unwin.

Williams, G. C. (1966). *Adaptation and natural selection: A critique of some current evolutionary thought*. Princeton: Princeton University Press.

Wilson, D. S. (1999). Tasty slice—but where is the rest of the pie? [Review of Buss, *Evolutionary Psychology: The New Science of the Mind*, 1st ed.] *Evolution and Human Behavior, 20,* 279–287.

Wilson, D. S. (2003). Evolution, morality, and human potential. In S. J. Scher & F. Rauscher (Eds.), *Evolutionary psychology: Alternative approaches* (pp. 55–70). Dordrecht: Kluwer.

Wilson, E. O. (1975). *Sociobiology: The new synthesis*. Cambridge: Harvard University Press.

Chapter 3
Towards a Cognitive Philosophy of Science

Ulrich J. Frey

3.1 Introduction

Epistemology is an important discipline of philosophy because it scrutinizes the very instruments we use to gain knowledge. In analogy, philosophy of science is well-advised to know about the scope, limits, and context dependencies of our cognitive abilities to be able to assess our instruments when employing them. In this sense, this chapter aims for a *more complete* approach to the philosophy of science. Research on influences on scientific work, such as historical and social aspects (e.g., Kuhn 1962/1976), should be *complemented* by taking cognitive aspects into account. If central concepts of philosophy of science like "explanation" are riddled with cognitive biases such as overconfidence (Bishop and Trout 2002), then philosophers of science should be aware of these problems, e.g., when making recommendations on what scientists ought to do.

This presupposes that cognitive aspects do play a vital part in scientific activities (activities in a very broad sense). As attested by renowned philosophers of science (e.g., Laudan 1977; Giere 1987), this is indeed the case—science is at its heart a cognitive activity: "Moreover I shall be concerned to treat the growth of science as a process in which cognitively limited biological entities combine their efforts in a social context" (Kitcher 1993, p. 9). Empirical investigations concerning these abilities are therefore necessary to obtain valid descriptions of the generation of scientific knowledge. Thus, we are well advised to consider what cognitive psychology has to say about our "instruments," that is our cognitive abilities and their roots.

But there is more: If we are biological beings, and therefore subject to the evolutionary process, then it also applies to our cognitive abilities (being an adaptation, see Buss 2004; Sober 1993; Pinker 1998). This means that it is important to research their primordial function in the evolutionary history of our species. "Considering the source," as Cosmides and Tooby put it (Cosmides and Tooby 2000; but see Gould

---

U. J. Frey (✉)
University of Halle, Halle (Saale), Germany
e-mail: ulifrey@yahoo.com

T. Breyer (ed.), *Epistemological Dimensions of Evolutionary Psychology,*
DOI 10.1007/978-1-4939-1387-9_3

and Lewontin 1979 for criticism), is vital for cognitive psychology. Evolutionary psychology produces knowledge about causes, structures, and particularities of our cognitive abilities. There is a broad empirical basis for such claims, ranging from countless experiments in cognitive psychology to research in phylogeny, ontogeny, and comparative behavioral biology. Cognitive psychology hereby answers the "How?" and evolutionary theory the "Why?" Evolutionary psychology is therefore important for any science concerned with human behavior. Although some philosophers judge evolutionary-based theories as unproven or irrelevant, neither is the case. On the contrary, today, theories distancing themselves from evolutionary theory are the ones that need to justify themselves.

This chapter aims to refute the second charge, the alleged *irrelevancy,* and will focus on *cognitive errors*. Although there is abundant literature about cognitive errors (Kühberger 1998; Gilovich et al. 2002), this evidence has, to my knowledge, never been applied to actual case studies in the philosophy of science. Section 3.2 describes some of these errors known from cognitive psychology. Section 3.3 provides examples for these cognitive errors in actual scientific research programs, focusing on the individual level of the scientist. Section 3.4 explains their evolutionary background and possible function. Section 3.5 describes the changes towards a naturalized philosophy of science, while Sect. 3.6 provides a summary of the results.

3.2 Cognitive Errors in Science

To err is human—and scientists are no exception. Among these errors is one group of errors—cognitive errors—whose form and peculiarities can be attributed to their evolutionary history. These errors are frequent, difficult to notice, very hard to remediate and will be at the center of the following analysis. There have been some attempts to bring together the famous results of Kahneman and Tversky on cognitive biases with the history of science (Trout 2002), but no claim for a general applicability of cognitive errors to case studies in the history of science, as I will demonstrate in the following sections.

First, it is necessary to precisely define the controversial term "error." The meaning of error that is used here is rather psychological than philosophical. *Errors* miss a purpose, they do not miss the truth. This analysis of errors is about failures in thinking patterns, about malfunctions of cognitive mechanisms. This more psychological approach rates cognitive errors as inevitable consequence of the fallibility of biological beings. It will be shown that cognitive errors have a significant impact on the generation of knowledge in science, despite the existence of a number of measures to avoid or detect errors in order to ensure "good" science. These measures range from statistical methods, controls in experiments, randomization, and double-blind designs to honest and open behavior towards the community (raw data, explanation of methods, etc.). However, these checks are not always sufficient.

3.2.1 Cognitive Errors in Complex Systems

It is well known that humans have great difficulties in dealing with complex systems. This has been extensively documented, for example, by computer-based simulations of complex systems including hundreds of subjects (Dörner 1983, 2004) as well as case studies in the history of science (Bechtel and Richardson 1993; Wimsatt 1986). Since some computer-based simulations show very good ecological validity (Dörner 2004), it suggests itself that these results can be applied to real scientific research. As it happens, it can be demonstrated that the exact same errors that are known to be frequent across large samples of participants from laboratory experiments do occur in real scientific research (see Sects. 3.2 and 3.3). These case studies in scientific research, whose errors can be traced back to similar results in the laboratory, prove that there is indeed no wide gap between cognitive science and its applicability to the history of science (Frey 2007).

There is much evidence for cognitive errors. However, we limit our scope here to errors in complex systems. Complex systems are dynamic, intransparent, nonlinear, and cross-linked systems. Often, knowledge about their workings is limited or incomplete. Subjects in experimental studies uniformly show problems in three areas:

1. Managing complexity (oversimplification, etc.)
2. Difficulties with time-related tasks (neglect of long-term effects, etc.)
3. Shortcomings in problem solving (wrong choice of problems, etc.)

As a sidenote: There are of course subjects who perform very well in managing complex systems. Also, subjects can be trained to some extent (Dörner 2004).

3.2.1.1 Oversimplification

One prominent error is the tendency of subjects to radically reduce complex systems to a simple linear model. Although sometimes this tendency is attributed to a typical male way of thinking, experimental evidence shows that women and men do not differ significantly in their cognitive strategies, as many typical actions reveal (Dörner 2004). The following list describes some typical cognitive errors and presents a scientific example for each one of them:

Cognitive error 1: A network is seen as an assortment of unconnected single systems, and problems are solved in a linear way.

Scientific example: Macquarie Island near Antarctica (Bergstrom et al. 2009). Rabbits, introduced in 1878, severely damaged the vegetation of the island. The officials responsible consulted scientific experts and decided in 1968 to contain the rabbit plague by introducing a virus. A few years later, the once 130,000 rabbits were decimated to a mere 20,000. In consequence, the vegetation recovered and in 2006, these measures were terminated. However, as early as 1985, it had become clear that cats—introduced in 1828—had to be killed, too, as they switched to birds instead of the scarcer rabbits. Therefore, from 1985 to 2000, every year some

120–200 cats were killed until there was no cat left. This in turn led to a massive explosion of rabbits, resulting in a complete displacement of the endemic vegetation by invasive plants. The example shows *linear problem solving* (first rabbits, then cats, then plants) *with seemingly unconnected single systems* (rabbits, cats, plants) in a pure form, even if the effects in nature triggered multiple multilayered impacts and feedback loops.

Cognitive error 2: Feedback loops and interactions are almost always cut out of the analysis of a complex system and consequences of own interventions are rarely controlled.

Scientific example: Construction of the Aswan Dam in Egypt. No scientific expert foresaw that the millennia-old inundation of fertile Nile soil would cease, leading to sharply reduced revenues from agriculture and fishery because of missing nutrients. Other unforeseen problems included erosion, increased salinity, and the frequent appearance of bilharzia. It can be argued that these errors—and in fact similar ones in any other case study—are not cognitive errors, but are due to political considerations. Here, the political prestige of building such a dam could have outweighed any other considerations, even though the dam was actually built to control floods and thus improve agricultural harvests. But in fact, since the exact opposite effects occurred—deteriorating harvests due to missing nutrients—this is a good argument against prestige-based decisions. A second argument against it is the level of analysis: In almost all other case studies or examples, the analysis is centered around the errors of scientific experts, not errors on the political level. On this level, scientists should be worried about their reputation and even liability, which speaks against making decisions based on prestige considerations.

Cognitive error 3: Even if subjects construct a mental representation of the system (which most do not do), even the best attempts single out one specific aspect to be responsible *for all other* effects.

Scientific example: Introduction of the Nile perch in Lake Victoria, 1954. Considered as a cure-all for economic problems, this large predator wiped out some 50 % of the endemic fish population (Ogutu-Ohwayo 1990). The basic idea was that the money earned by exporting this fish would lead to a restructuring and rebuilding of the infrastructure, greater wealth for the population, better medical care, and so on. Although the Nile perch became an export hit, it definitely proved not to be an economic cure-all—on the contrary, as a documentary aptly named “Darwin’s Nightmare” demonstrated.

Cognitive error 4: Test subjects virtually do not correct their errors, one’s own errors are almost never noticed and learning from errors is very infrequent.

Scientific example: Nuclear power plant problems and simulations. Prominent examples are the accidents of *Three Mile Island* or *Oyster Creek*: The errors of both operating teams were discovered *only* through new team members (*fresh eyes*). Simulations corroborate this: Twenty-three experienced operating teams of nuclear power plants did not discover *a single* diagnostic error (e.g., what is the impact of the valve on the control circuit?) in 99 emergency simulations. Only 50 % of all execution errors were discovered by the originators (Reason 1990).

Cognitive error 5: Rethinking interventions and inquiring about new information is declining rapidly (under 10 % of all actions) in decision-making situations.

Scientific example: Introduction of snails (see also Sect. 3.3.2). When scientific experts introduced the predatory snail *Euglandina rosea* in Hawaii, there was not a single control whether the measures taken were successful or not (they were not): "However, the lack of follow-up, both in terms of determining its impact on crop pests and in ensuring that it does not cause damage to native species, is striking" (Griffiths et al. 1993, p. 80). Experts recommended the introduction, although they knew that there was no empirical data about the predatory snail and its supposed prey, and although they were very well aware of the potential dangers of new introductions (Mead 1961).

This list is by no means complete, but many tendencies should ring a familiar tone. Of course, reducing complex systems into isolated subsystems can be a valid and fruitful research strategy (Bechtel and Richardson 1993), but *only* if one is aware of the potential risks.

3.2.1.2 Shortcomings with Time-Related Tasks

The second problem area is related to time. Human beings are comparably apt in solving spatial tasks, but do significantly worse in time-related tasks. Subjects have massive problems even with very simple test simulations, e.g., the task to regulate a temperature by adjusting a thermostat. Whereas visual feedback and possibilities to adjust were permanently given in this experimental design, the only difficulty was a slight, but realistic time lag from regulative action to effect (Reichert, as cited in Dörner 1985). Even under optimal conditions, subjects did poorly in similar tasks. Typical error-prone reactions are:

Overgeneralization: Some or even one instance suffices to infer per induction to all instances (logically incorrect, see Hume 1748/1964).

Scientific example: Decimation of cactuses. The success story in Australia in fighting one invasive species (cactuses) with another (the insect *Cactoblastis cactorum*) turned into a disaster in other countries, as scientists inferred its effectiveness to also be valid in totally different climates, contexts, and conditions (Low 1999).

Ad hoc explanations to salvage early hypotheses: The first is kept even after massive evidence to the contrary is available. This is especially interesting for researchers, as this psychological behavior in particular endangers methodically sound science. Huxley put it eloquently as "the great tragedy of Science—the slaying of a beautiful hypothesis by an ugly fact" (Huxley 1870/1970, p. 244).

Example: Experiments with feedback systems (Dörner 2004). Subjects had to find rules for controlling a feedback system via a dial with numbers on it. Subjects formulated rules like "95 is a good number." When this rule turned out to be wrong, they salvaged it with conditional rules like "but only, if the number 80 came before," instead of simply giving up the initial hypothesis.

Neglect of side effects and long-term effects: Once an action has been taken, its side- and long-term effects are not considered.

Scientific example: Management of the Blue Mountain Forests in Oregon. The *National Forest Service*[1] (NFS) was founded in 1905 to guarantee sustainability in forest management. The inability to master complex systems can be demonstrated best with examples where measures taken lead in fact to the *exact opposite* of the desired results. "In the process of trying to manage extremely complex landscapes, foresters set into motion a chain of events that increasingly swung out of control" (Langston 1995, p. 8).

To decrease fires, the NFS cleaned up dead wood to minimize the fuel for forest fires. In fact, controllable middle-sized fires did indeed decrease, but unfortunately gave way to an increasing proportion of uncontrollable major fires. The associated damages were thus multiplied: "a kind of worst-case scenario come true" (Little 1995, p. 84).

Current problems are considered more important than more crucial problems in the future: This is known as the discount factor, which has been experimentally researched for different species.

Scientific example: Management of the Blue Mountain Forests in Oregon. Already in the 1930s and 1940s, the NFS knew that their management was not sustainable. They could even correctly predict a major collapse in the 1980s due to their nonsustainability, but even with this knowledge put present considerations before future problems. Again, this list is by no means complete.

3.2.1.3 Methodical Shortcomings

The third problem area contains general weaknesses in problem solving. Apparently, (unsuccessful) subjects are not able to separate important from unimportant problems. Instead, they choose to solve those problems first that seem solvable or prominent, or interesting for them (Dörner 2004). Furthermore, those problems are considered first that belong to their area of competence. Worse, the central problem is often considered as if it was of their own area of competence, even if this is not the case. This spills over to problems not pertaining to the area of competence, as they are incorrectly treated with the known, subject-specific methods (Schönwandt 1986).

Additionally, nearly all subjects choose the same methods to solve problems. This is as true for the content (Johnson-Laird and Byrne 1991) as it is for the methods themselves. For example, practically no one plans backwards from the goal (Newell and Simon 1972), although this can be a highly successful strategy. Other errors include wrong priorities, absent controls, and insufficient structuring. Most subjects practice a "muddling through" approach, tackling always the next problem in line and starting with the most prominent, but not the most important problem. In doing so, the central issue is often forgotten (Dörner 2004). Another interesting

---

[1] Based on scientific experts, the NFS has an output of 1200 publications in peer-reviewed journals per year.

weakness is the assumption that a complex system is in truth easy to analyze and to master.

It is also important to note that humans are boundedly rational. They are not a perfectly rationalizing machine like the *homo oeconomicus,* a concept still used in economics. On the contrary, bounded rationality implies that humans use *heuristics,* which are error-prone in their own way (Czerlinski et al. 1999). Optimal solutions cannot be attained due to time limitations, capacity constraints, and incomplete information. Heuristics use filtering and conditionals to terminate the search for solutions. As a consequence, much of the information related to a specific decision is never considered at all.

Scientific example: Challenger catastrophe 1986 (Hilton 2002). The day before launch, the engineers knew that in four out of eight cases, the O-rings broke in cold conditions. Was the low temperature the decisive factor? The engineers rejected this hypothesis, since in 50 % of the cases the O-rings stayed intact. They did not look at the figures of shuttle launches in warm weather, although those were readily available: sixteen shuttle launches and only one case of defect O-rings. This demonstrates how fast a search for hypotheses is closed: If the first test is not meaningful, the standard heuristic rejects that hypothesis and starts to look for other explanations, even if a simple test (lookup) would have been sufficient for verification.

This summary of findings of the research on managing complex systems concludes the first step of establishing that psychology is important for all human-behavior-related disciplines, including the branch of philosophy of science dealing with the description of scientific progress. We used experimental evidence to show that generation, use, and evaluation of scientific knowledge depend on our cognitive abilities, including peculiarities and error-prone heuristics. Many different examples, ranging from biology and ecology to physics, document that—contrary to conventional wisdom—this evidence is pertinent to science. The following section develops and strengthens this argument with additional case studies. A further step (see Sect. 3.4) consists in the evolutionary explanation of these errors. With this background, a new perspective on errors opens up: Errors should not just be reduced to their negative aspects, but could actually be necessary trade-offs of mechanisms allowing humans to cope with a complex world.

3.3 Management of Complex Systems

Undoubtedly, ecosystems are complex systems. A closer look at efforts by scientists trying to manage ecosystems shows that indeed the same errors known from experimental cognitive psychology occur in their work. The next section establishes that cognitive strategies, peculiarities, and errors exhibit a huge impact on scientific thinking and problem-solving strategies. The errors identified in the following case studies are in italics and identical to those explained in detail in Sect. 3.2.

3.3.1 Example 1: Yellowstone National Park

The management of the Yellowstone National Park (YNP) and other national parks in the USA exhibits nearly identical errors (cf. Chase 1987). YNP was created in 1872 and has been managed with the explicit goal of nature conservation. Still, it is easy to find failures or even effects reverse to the ones originally intended: "They were all decent people, well intentioned and devoted to wildlife conservation. It seemed nearly inconceivable […] [that] the destruction of Yellowstone—could happen when so many people were dedicated to its preservation" (Chase 1987, p. 193).

Well-intentioned attempts of experts and scientists lead to catastrophic results for the complex ecosystem of YNP, as the above mentioned error-prone problem-solving strategies were applied. Examples for *linear problem solving* can be found in the attempts to regulate animal populations. To increase the elk population, coyotes, mountain lions, and wolves were shot. Although this measure clearly did not help at all, it was nevertheless continued. However, the real reason for the decline of large plant-eaters was in fact overgrazing. In other cases, only one animal population was regulated, but not the interdependent interplay between populations. Thus, the results were unsatisfactory: massive, undesired fluctuations in populations, underlining human's general *difficulties with time-related tasks*. Here it becomes most apparent how uniformly humans solve certain problems: Subjects in the laboratory typically produced massive fluctuations (zigzag charts) when trying to manipulate populations—as did the scientific experts in YNP for decades. This in turn is symptomatic for another cognitive peculiarity: not learning from one's own errors. Although precisely these regulating errors had been well known and published since 1930, there is *no indication of a learning effect* up to today (Chase 1987).

Another example shows evidence of several other cognitive errors. Through regulation measures, the population of deer had been surging in the 1970s and 1980s. An unwanted consequence was the decimation of beavers and grizzly bears, as *feedback loops and mutual interactions had been ignored*. The problem consisted in *unforeseen side effects* of one of these interactions: Resources, common to beavers, grizzly bears, and deer, had been overused by the thriving deer population. As a result, not only beavers and grizzly bears were negatively affected but the deer population meant to proliferate completely collapsed shortly thereafter due to overuse. Likewise, after 8 years of a "grizzly bear protection program," experts had to account the same program for a *decrease* of over 50 % due to protection measures. This is sad evidence that the opposite of the desired results often occurred.

3.3.2 Example 2: Introduction of New Species

Other current examples for man's failure to cope adequately with complex systems are introductions of new species in existing ecosystems—mostly with devastating results.

The snail *Achatina fulica* was introduced as food source for the local population to Hawaii around 1955 and to Tahiti in 1967. Instead of its supposedly positive effect, it soon led to the worst snail pest in the tropics (Cowie 2001; Mead 1961). The response of scientific experts was to introduce another snail—the predatory snail *Euglandina rosea*. However, this snail failed to prey on *Achatina* and instead wiped out nearly all other endemic snail species. From originally around 750 species, today only 10–35 % have not yet become extinct (Cowie 2001). A very similar situation occurred in Tahiti, where *Euglandina* eradicated many *Partula* snail species (Cook 1989).

The usual reaction to such incidents is: Even scientific experts sometimes commit errors. These errors, however, are individual errors and there is not much to do about it. Yet, we will learn from these errors and as we will continue to know more about ecological systems and accumulate more scientific knowledge about them, such errors will become scarcer and introductions of new species will be handled more carefully. Thus, we will see some sort of *progress*.

Remarkably, none of these assumptions are right. Many of these deplorable errors are not individual mistakes, but follow certain patterns. These patterns occur again and again—not only in ecology but in physics, psychology, medicine, and other disciplines as well (Frey 2007). Scientists repeatedly make certain cognitive errors, in spite of their education and in spite of the corrective methodology of science.

A short discussion shows more evidence of these—by now known—errors. The *radical simplification* is particularly evident in this case, as 12 other snail species were introduced with the same intention even after *Euglandina* (Cowie 2001): The model of the ecosystem shrinks to two species and one interaction (predator and prey). In addition, a second error is apparent—*linear problem solving*. If two introductions of snails fail, then introducing another 12 snail species shows clearly that a specific problem is attempted to be solved over and over again with the same (unsuccessful) strategy.

The damages caused to ecosystems are irrevocable and cannot be compensated again. The *economic damage alone* due to newly introduced species in the USA is estimated at about US$130 billion per year (USDA Forest Service 2005; Pimentel et al. 2005): "What makes our pest problems unique is the way in which so many of them were *deliberately created*. No one ever set out to pollute our rivers or erode our soils" (Low 1999, p. 41, emphasis added).

The sad truth is that scientific experts *do not learn from their errors:* Although it is known that only 6 % of all deliberate introductions of new species were successful and about 18 % had some positive effects, the overwhelming majority of 76 % were total failures (Low 1999). However, the introductions continue unabated. Virtually thousands of other examples could be cited (cf. www.issg.org for the "Top 100" of the world's worst alien species; see also Kowarik 2003; Low 1999), many of them within the past few years.

Still, these are just some errors among many others, and errors are just one aspect of cognitive abilities in general. The point that should be made here is that the way scientists and humans in general think, i.e., their cognitive abilities, are relevant to obtain valid descriptions of the generation of scientific knowledge.

3.3.3 Example 3: The Chernobyl Catastrophe

One of the great catastrophes in the history of mankind took place in Chernobyl on April 26, 1986. Unlike the two examples of biological complexity, the accident of Chernobyl is an example of a highly complex *technical* system which humans (scientists and highly skilled experts) were unable to manage. The following analysis (based on Medwedew 1991 and Reason 1987) shows that this accident is no isolated case. Instead, "the question becomes not 'Could it happen here?' but 'When and where will it happen, and how bad will it be?'" (Reason 1987, p. 206). Unfortunately, this warning has now become reality in Fukushima Daiichi. It is furthermore a prime example of cognitive errors, rule violations, and wrong decisions rather than technical problems. Again, the errors of the analysis above (Sect. 3.2) will be in italics. For space reasons, the most prominent errors are presented in form of a list:

1. The team responsible for the accident was under heavy time pressure to complete their assigned task of testing an emergency procedure.
2. The team had successfully completed numerous other tasks, was therefore overconfident and sloppy in executing safety procedures. This led to violations of safety protocols. Moreover, their area of competence did not include emergency situations.
3. Most team members were neither physicists nor nuclear power specialists and had little or no concept or visualization of the workings of a nuclear power plant. This led to wrong decisions in critical moments.
4. Expert opinion was subordinated to authority. Blatantly wrong decisions were executed despite better judgment.

The chronology of the accident dates back to long-standing warnings concerning the inherently unsafe design of the widespread reactor type Reaktor Bolshoy Moshchnosti Kanalniy (RBMK). One of the well-known problems with RBMK reactors is that they become easily unstable when run with low-power load. Consequently, safety procedures forbid operating the power plant with power loads below 20 %. The test itself should therefore have been run at a power load of 25 %.

Error 1—failure to regulate the load properly, a *typical problem of time-related tasks:* Regulating the reactor to about 25 % load failed spectacularly; instead, the power dropped to a mere 1 %. After hours of struggling, the team managed to stabilize the power plant at 7 %. The experiment should definitely have been abandoned at this point, as the reactor was now "poisoned." However, they continued to run the experiment (*blindness to own errors*).

Errors 2 and 3—wrong activation of cooling pumps, based on *a wrong concept and visualization of the system:* For safety reasons, all eight water cooling pumps were activated. This is a clear violation of safety procedures. Contrary to the intuitive belief that more pumps are safer (*linear thinking*), the opposite is the case. Thus, the complex interdependent forces of the reactor system were not understood properly but *replaced by a simplified and wrong model*. As an automatic consequence, safety graphite rods were pulled out of the reactor core. This is clearly *an unforeseen side effect* and was not even noticed by the team.

Errors 4 and 5—blind regulation of the reactor, *no control of own measures taken:* The water throughput was increased threefold. Again, this was done for "safety reasons." However, as a result, still more graphite rods were pulled out from the core. The reactor was thus regulated "intuitively." Finally, someone ordered a printout of the number of graphite rods still in the core. Although only 6–8 rods remained—the minimum number of rods allowed was 12—the experiment continued. Again, this is clear evidence for *linear problem solving* and *underestimation of future developments*. This was just 2 minutes prior to the catastrophe.

Error 6—turning off safety systems, based on an *inadequate model of the complex system:* The valves of generator eight were then closed, turning off another automatic safety system. The combination of all of these errors finally led to the explosion. The last minutes, which cannot be spelled out in detail here, are characterized by *linear problem solving, wrong extrapolations, missing controls of the consequences of own interventions,* and *no corrections for errors*.

One more error from the aftermath shall be pointed out: *Errors are not corrected* and *learning from errors is very infrequent*. This is evident from the reaction from the West: "It could not happen here," and also from the Soviet Union, as these reactor types are still in use today.

These three case studies and the short examples show that cognitive errors seem to be structurally identical, even in disciplines as diverse as biology or physics. The argument is that errors are representative of cognitive abilities in general, which in turn are an essential building block if one aims for a more complete understanding of how science works. Of course, there are also examples of successful management of complex systems. Thus, resignation is not in order. However, the base rate of errors seems to remain constant, often in spite of technical support or scientific expertise. It therefore seems advisable to investigate what *causes* these kinds of failures, with the evolutionary background of these errors being one promising path to explore.

Another argument supporting these claims comes from evolutionary psychology: There is a difference in complexity between the postindustrial environment of modern humans and the so-called environment of evolutionary adaptedness (EEA). Modern humans have a range of options which were not available to our ancestors: dispersing invasive species, managing complex ecosystems, and technical devices amplifying or extending abilities by orders of magnitude.

3.4 Evolutionary Explanation of Cognitive Errors

One basic tenet of evolutionary theory is that adaptations have to be analyzed in regard to the environment they developed in. To therefore understand how humans handle complexity, we have to look at the environment of our past. It had certain features: First, its more complex features had to be made more manageable. Second, many situations required fast actions. Third, erroneous or incomplete information should not negatively interfere with decision making. The described "errors" or what could better be called *mechanisms* do just that: Complicated estimations are reduced to simple extrapolations; causal nets are reduced to one single cause; alternatives are not considered at all; measures, once taken, are never controlled; and a few singular cases suffice to produce a general rule.

On the one hand, these ways of behavior are shortcomings, malfunctions in regard to scientific standards and norms, and should therefore rightly be classified as errors. From an evolutionary point of view, however, these "errors" are apt methods that enable humans to cope with complex situations under time pressure, incomplete information, and limited capacity. This is why humans use certain heuristics that are just precise enough for matters at hand while using up minimal brain resources. Here, natural selection does not favor precision, but speed and simplicity—which are primarily achieved by reduction on all levels. In fact, this has to be the case, as our environment is polymorphic to such an extent that not all information can be analyzed: Only a very small percentage of the already heavily filtered sensory stimuli are analyzed at all, not to mention the still fewer sensory inputs actually used or remembered. In addition, humans use some surprisingly efficient lexicographical heuristics (cf. Gigerenzer et al. 1999; Payne et al. 1993) and follow strategies that can be described as satisficing within a bounded rationality: "To satisfice is to choose a path that satisfies your most important needs, even though the choice may not be ideal or optimal" (Plous 1993, p. 95).

The mentioned mechanisms of handling complex situations under suboptimal conditions can be specified as follows:

1. *Goal:* Use reduction, selection, and quick termination in order to simplify complex situations and to be able to react fast.
2. *Search:* Use only the "most important" (most conspicuous) cues and ignore the rest. Compare cues only in yes-or-no decisions, not in a quantitative manner. Proceed serially. Treat causal nets as isolated systems. Reduce many causes to a single one. Use routine procedures and repeat them. Use induction to generalize rules from one or few single examples. If there are further limiting factors: Speed up the processing, filter, and simplify more.
3. *Assessment of cues:* Prefer concrete and conspicuous cues. Treat cause and effect always as if they were linear. Ignore side effects and long-term effects, disregard alternatives. Do not control your actions.
4. *Termination:* Terminate your heuristic search as soon as the target set by yourself is reached. If the target is not reachable with the above mentioned heuristic measures, simplify still more by ignoring other specifications (e.g., precision). If there is no hint of a solution, then guess.

Due to the listed principles of action, it is possible to "mirror" the mentioned errors, thus discovering a *function*. This function is the result of an adaptation. The (ultimate) main function of this particular adaptation could be named as "the ability to act fast, achieved through reduction," thus being able to cope with suboptimal conditions (capacity and time constraints, incomplete information, etc.). The "errors" are not adaptations by themselves, but inevitable trade-offs of this adaptation (for more details, see Frey 2007).

Accordingly, the tendency to think mostly in a linear way and to neglect side effects is a *trade-off* between the ability to decide and act fast and to assess problems correctly in their complexity. The use of very robust but nevertheless successful heuristics (Payne et al. 1993) is an adaptation to the very broad and diverse range of environmental problems. As a *by-product,* we find the error of radical abortion criteria leading sometimes to suboptimal solutions.

3.5 Towards a Naturalized Philosophy of Science

It is surprising that even today, philosophy disregards empirical evidence as mostly irrelevant, although it can be demonstrated that empirically based philosophy of science is the most suited approach for describing and explaining patterns in the history of science (e.g., Thagard 1998a, b; Bechtel and Richardson 1993; Giere 1988; Donovan et al. 1988b). As Casebeer and Churchland put it in the field of ethics: "This means that moral philosophers will need to think the unthinkable, namely that the empirical results in the evolutionary and neural sciences have an unavoidable impact on moral philosophy." (Casebeer and Churchland 2003, p. 189). In fact, in the field of ethics, the groundbreaking work of Knobe (2003) has demonstrated the importance of empirically grounded philosophy. A naturalized approach criticizes mainstream philosophy of science, because empirical evidence is neglected in three ways:

1. Most analyses concentrate on *social* and *historical* influences on scientific change as the most important forces. *Cognitive* aspects are often neglected (cf. Rubinstein 1984).
2. Studies in the history of science are not considered to back up theories.
3. Explaining phenomena in the philosophy of science is done without consulting natural scientific theories.

Such criticism is not entirely new, but a systematic use of evidence from the natural sciences and a more encompassing cognitive and evolutionary framework for analyses in the philosophy of science has seldom been applied. Especially, Tweney and Dunbar argue for integrating cognitive sciences and philosophy of science for a mutual benefit (Tweney 2001; Tweney et al. 1981; Dunbar 1995, 1997). Both authors demonstrate how historical or actual case studies can be more profitably analyzed by considering cognitive strategies of scientists. In the same vein, their examples from the history of science illustrate both functioning and applicability of the approach described in this chapter.

I will now discuss the three points of criticism mentioned. The former two concern the analysis of a change in scientific theory, the latter one argues for a framework grounded in the natural sciences.

To point 1—*neglect of cognitive aspects:* Many studies have focused on social or historical influences to account for and explain many phenomena in the philosophy of science. Little attention has been paid to cognitive aspects, although science can actually be described at its core as follows: “Science is a cognitive activity, which is to say it is concerned with the generation of *knowledge*” (Giere 1987, p. 139).

Even when cognitive strategies are mentioned as important (Solomon 1994), they are eventually cut out in the final analysis and replaced by social factors and group decision processes. Exceptions are the research of Dunbar on the influence of cognitive mechanisms on scientific activities in laboratory studies (Dunbar 1995, 1997, 2002) and work on the generation of hypotheses and the use of heuristics (Klahr et al. 1993). However, such studies show that individual cognitive strategies, like the use of analogies and apt representations of the problem, are decisive for scientific success, as well as the strategies of the research group. It is even safe to say that scientific activities are *at their core* cognitive activities, or, as Giere puts it: “Science provides arguably the best example of a higher cognitive activity” (Giere 2002, p. 2).

Some building blocks of research as cognitive activity are, for example, the tasks of developing models, thinking in analogies, and daily problem solving in general. Even if the impact of group communication, group proceedings, and career thinking would indeed be the decisive factor, as is often assumed, the *content* of all these social forces is based on cognitive scientific activities. It is therefore important to consider psychological aspects of cognitive activities. Indeed, there is a strong argument against a one-sided social interpretation:

> More specifically, cognitive sociology of science is predicated on the existence of determinable correlations between the social background of a scientist and the specific beliefs about the physical world which he espouses. Despite decades of research on this issue, cognitive sociologists have yet to produce a single general law which they are willing to evoke to explain the cognitive fortunes of any scientific theory, from any past period. (Laudan 1977, pp. 217–218, emphasis in the original)

Or, to put it another way: There is neither a Jewish branch of physics, nor a bourgeois mathematics, nor a proletarian theory of relativity.

To point 2—*neglect of historical case studies:* The neglect of these studies in the history of science applies even to philosophers like Thomas Kuhn and Imre Lakatos, who are known—in particular Kuhn—to back up their theories with examples from the history of science, some of which have become classics in their own respect. Nevertheless, many assumptions and hypotheses of these two great scholars are not tenable (see Donovan et al. 1988a), since many thorough historical case studies revealed a rather unexpected result: Some of the most cherished concepts of Kuhnian followers did not survive a closer look, e.g., the incommensurability of paradigms or the alleged inability of scientists to communicate when belonging to different paradigms. This is also true for most Lakatosian concepts. Moreover, the majority of case studies in philosophy of science are about old events, although around 90 % of all scientists are living today (Bishop and Trout 2002).

In contrast, there are minute and precise accounts of historical conceptual revolutions (see, e.g., Gooding 1990; Thagard 1998a, b; Rheinberger 1998). Taking as an example how bacteria can cause gastritis demonstrates how conceptual revolutions may start with a “crazy idea,” are subsequently reinforced by an interplay between theory and experiments, demonstrate causal links by experiments, and thus go from “preposterous” to commonly accepted in only a decade (Thagard 1998a, b).

To point 3—*hypotheses are not grounded in natural sciences:* In general, researchers should check personal intuitions and assumptions against the current state of research. However, in philosophy of science, this seems to be done infrequently. Just as an example, it is claimed that the logical capabilities of animals are unknown (Papineau 2000), even though they are actually well known (e.g., Papini 2002). It is also claimed that there is no way to deduce specific details of heuristics, despite the fact that there are several (e.g., Krebs and Davis 1996). These wrong intuitive assumptions lead to false speculation about decision making and optimization, a notable exception being Giere (1988, 1992).

These problems add to the more general argument of shifting the framework of philosophy of science that is concerned with descriptions of scientific research, towards the natural sciences. This approach of philosophy of science would profit if it were more *naturalistic* (Bishop and Trout 2005). More precisely, one way to define naturalistic philosophy is with regard to *methodology* ("Use findings from the natural sciences to the greatest extent possible"), to *ontology* ("Take materialism as foundation"), and to *epistemology* ("Work with hypothetical realism"; Sukopp 2006). It is suggested that some long-lasting problems could be resolved this way. I agree with other philosophers of science that a naturalistic philosophy of science should itself employ naturalistic methods, e.g., using base rates and relative frequencies in comparing case studies (Bishop and Trout 2002).

Many philosophers of science have judged the mentioned scientific results as relatively unimportant (the above mentioned *irrelevancy* argument). The case studies above however demonstrate that they are not. A more valid criticism is the apparent gap between evolutionary psychology and philosophy of science. Again, the case studies demonstrate that this gap is surprisingly narrow.

Cognitive aspects in particular play no role in many analyses. Reasons mentioned are missing coherent psychological theories (but this argument is outdated since the 1970s, particularly since the advent of evolutionary psychology, see, e.g., Barkow et al. 1992), reductionism in biological explanations (but these explanations are legitimate in their claims, see Ruse 1988), and the problem of generalizing from individual thinking patterns to common rules (but such analyses are not interested in individual, but only general patterns that can be found in the majority of people). For a more detailed refutation of these objections, see Frey (2007).

3.6 Conclusion: Cognitive Aspects Are Relevant for Science

This chapter argued for the proposal to incorporate empirical evidence (in particular on cognition) into philosophy of science concerned with describing scientific research (and other disciplines related to human behavior) in three steps. First, by integrating results from natural sciences into research (here: experimental psychology and its subsection of heuristics in complex systems). Second, by providing evidence from case studies that these results matter (here: that these experimentally demonstrated heuristics with their errors can be found in the history of science). Third, by providing an evolutionary explanation why this may be so (here: why humans use these and not others heuristics?).

Of course, the case studies used can be interpreted in a noncognitive way, but there are many more examples of the significance of cognitive aspects to be found in other case studies (Frey 2007). These case studies are taken from disciplines as diverse as medicine, physics, biology, and psychology. Moreover, they span a time period of three centuries (seventeenth to twenty-first century). The more diverse and temporally separate these historical examples are, the more problematic historical interpretations necessarily become. The striking similarity, the invariance over time, the transcultural and transdisciplinary occurrence of cognitive influences makes a “historical-only” interpretation hardly viable.

In fact, there is a surprising isomorphic structure, in some cases even identity between documented experimental errors and real errors in the history of science. As cognitive errors are very specific, it seems impossible to mistake them for something else. Both scientists and laymen commit the same errors, which can be observed independent of epoch or special field, as they are common and systematic. They seem to be invariant to time or field of research. An interplay of historical, social, and cognitive influences is to be expected. However, it has not been researched yet how these factors weaken or strengthen each other in mutual ways.

To my knowledge, research in philosophy of science has so far only used cognitive or evolutionary findings to explain *singular* facts. Examples are Ronald Giere and Nancy Nersessian. These approaches ask questions like “What is the cognitive basis for modeling in science?” or “How do analogies work in scientific research?” (Nersessian 1987, 2002). This kind of research is without doubt a very valuable contribution to a more cognitive-focused philosophy of science I am arguing for. But it does not take into account in a *systematic* way the frameworks of evolutionary biology and evolutionary psychology to broaden the approach to an encompassing naturalistic one. In the same vein, Ronald Giere limits his “cognitive studies of science” mainly to two specific cognitive mechanisms, i.e., representation and judgment (Giere 1987), but does not consider the whole spectrum of our cognitive abilities and their evolutionary roots.

The argument presented here demonstrates how such an analysis may proceed and go beyond existing approaches. First, hypotheses for an investigation in the philosophy of science interested in describing scientific research are derived from the natural sciences. Second, empirical proof for these hypotheses is presented. Third, the hypotheses are tested by applying them to case studies from science and technology. Fourth, evolutionary theory is used to explain their structure. A fifth step, missing here, is to analyze the interplay of various other influences on the hypothesis in question. To conclude, the main results can be summed up as follows:

Concerning scientific change:

1. Cognitive abilities of scientists are important for almost all analyses in the philosophy of science.

2. Roots, development, strengths, weaknesses, and limits of cognitive abilities can be explained by evolutionary theory.
3. Cognitive influences complement other impact factors on scientific change, like social and historical aspects.

Concerning the methodology of philosophy of science in general:

1. Philosophy of science has to be complemented by the history of science, e.g., through the use of (more than one, systematic, and comparable) case studies (see also Bishop and Trout 2002, for a similar claim).
2. Premises and assumptions in the philosophy of science describing research processes have to be anchored in (or at least cross-checked with) natural sciences.

This has not been worked out in a *systematic* way yet, although there have been many steps in that direction (e.g., Ruse 1995; Callebaut 1993; Rubinstein 1984; Dunbar 1995; Giere 1987).

An evolutionary-cognitive approach therefore asks philosophers of science for two modifications: On the one hand, a more empirical methodical approach should be applied, including a precise and in-depth verification of theories with case studies and grounding of hypotheses in natural sciences. In fact, there has been an increasing tendency to do so—at least for the first part—in recent years. Moreover, these seem to be modest modifications, as no modern philosopher of science would ever oppose them. But, and this is an important "but," this is only true for the first part—the use of case studies—not for the systematic grounding of hypotheses in natural sciences. There may be a consensus about how naturalistic philosophy of science should proceed (see Callebaut 1993), but too often this remains a lip service (Rubinstein 1984; Ruse 1995).

On the other hand, philosophers should consider cognitive influences on the generation of scientific knowledge in a more systematic way. This implies taking into account the evolutionary functions of our cognitive abilities. Again, this may be consensus already, but has very rarely been implemented in a systematic way. The strong influence of the cognitive dimension in the philosophy of science has so far often been underrated.

References

Barkow, J. H., Cosmides, L., & Tooby, J. (Eds.). (1992). *The adapted mind: Evolutionary psychology and the generation of culture*. Oxford: Oxford University Press.

Bechtel, W., & Richardson, R. C. (1993). *Discovering complexity: Decomposition and localization as strategies in scientific research*. Princeton: Princeton University Press.

Bergstrom, D. M., Lucieer, A., Kiefer, K., Wasley, J., Belbin, L., Pedersen, T. K., & Chown, S. L. (2009). Indirect effects of invasive species removal devastate World Heritage Island. *Journal of Applied Ecology, 46,* 73–81.

Bishop, M. A., & Trout, J. D. (2002). 50 years of successful predictive modeling should be enough: Lessons for philosophy of science. *Philosophy of Science, 69,* 197–208.

Bishop, M. A., & Trout, J. D. (2005). *Epistemology and the psychology of human judgment*. Oxford: Oxford University Press.

Buss, D. M. (2004). *Evolutionary psychology: The new science of the mind*. Boston: Pearson.

Callebaut, W. (1993). Turning naturalistic: An introduction. In W. Callebaut (Ed.), *Taking the naturalistic turn or how real philosophy of science is done: Conversations with William Bechtel, Robert N. Brandon, (...)* (pp. 1–11). Chicago: University of Chicago Press.

Casebeer, W. D., & Churchland, P. S. (2003). The neural mechanisms of moral cognition: A multiple-aspect approach to moral judgment and decision-making. *Biology and Philosophy, 18*(1), 169–194.

Chase, A. (1987). *Playing God in yellowstone: The destruction of America's first national park*. San Diego: Harcourt Brace.

Cook, A. (1989). Factors affecting prey choice and feeding technique in the carnivorous snail *Euglandina rosea* (Ferussac). *Journal of Molluscan Studies, 55*(4), 469–477.

Cosmides, L., & Tooby, J. (2000). Consider the source: The evolution of adaptations for decoupling and metarepresentation. In D. Sperber (Ed.), *Metarepresentations: A multidisciplinary perspective* (pp. 53–115). Oxford: Oxford University Press.

Cowie, R. H. (2001). Can snails ever be effective and safe biocontrol agents? *International Journal of Pest Management, 47*(1), 23–40.

Czerlinski, J., Gigerenzer, G., & Goldstein, D. G. (1999). How good are simple heuristics? In G. Gigerenzer, P. M. Todd, & ABC Research Group (Eds.), *Simple heuristics that make us smart* (pp. 97–118). Oxford: Oxford University Press.

Donovan, A., Laudan, L., & Laudan, R. (Eds.). (1988a). *Scrutinizing science: Empirical studies of scientific change*. Baltimore: John Hopkins University Press.

Donovan, A., Laudan, L., & Laudan, R. (1988b). Testing theories of scientific change. In A. Donovan, L. Laudan, & R. Laudan (Eds.), *Scrutinizing science: Empirical studies of scientific change* (pp. 3–44). Baltimore: John Hopkins University Press.

Dörner, D. (1983). Heuristics and cognition in complex systems. In R. Groner, M. Groner, & W. F. Bischof (Eds.), *Methods of heuristics* (pp. 89–109). Hillsdale: Lawrence Erlbaum Publishers.

Dörner, D. (1985). Wie Menschen eine Welt verbessern wollten, und sie dabei zerstörten. *Bild der Wissenschaft, 2*, 48–53.

Dörner, D. (2004). *The logic of failure: Recognizing and avoiding error in complex situations*. New York: Basic Books.

Dunbar, K. (1995). How scientists really reason: Scientific reasoning in real-world laboratories. In R. J. Sternberg & J. E. Davidson (Eds.), *The nature of insight* (pp. 365–395). Cambridge: MIT Press.

Dunbar, K. (1997). How scientists think: On-line creativity and conceptual change in science. In T. B. Ward, S. M. Smith, & S. Vaid (Eds.), *Conceptual structures and processes: Emergence, discovery and change* (pp. 461–493). Washington: APA Press.

Dunbar, K. (2002). Understanding the role of cognition in science: The science as category framework. In P. Carruthers, S. Stich, & M. Siegal (Eds.), *The cognitive basis of science* (pp. 154–170). Cambridge: Cambridge University Press.

Frey, U. J. (2007). *Der blinde Fleck – Kognitive Fehler in der Wissenschaft und ihre evolutionsbiologischen Grundlagen*. Heusenstamm: Ontos.

Giere, R. N. (1987). The cognitive study of science. In N. J. Nersessian (Ed.), *The process of science: Contemporary philosophical approaches to understanding scientific practice* (pp. 139–159). Dordrecht: Kluwer.

Giere, R. N. (1988). *Explaining science: A cognitive approach*. Chicago: University of Chicago Press.

Giere, R. N. (1992). Introduction: Cognitive models of science. In R. N. Giere & H. Feigl (Eds.), *Cognitive models of science: Minnesota studies in the philosophy of science: Vol. XV. Cognitive models of science* (pp. XV–XXVIII). Minneapolis: University of Minnesota Press.

Giere, R. N. (2002). Scientific cognition as distributed cognition. In P. Carruthers, S. Stich, & M. Siegal (Eds.), *The cognitive basis of science* (pp. 285–299). Cambridge: Cambridge University Press.

Gigerenzer, G., Todd, P. M., & ABC Research Group (Eds.). (1999). *Simple heuristics that make us smart*. Oxford: Oxford University Press.

Gilovich, T., Griffin, D., & Kahneman, D. (Eds.). (2002). *Heuristics and biases: The psychology of intuitive judgment*. Cambridge: Cambridge University Press.

Gooding, D. C. (1990). *Experiment and the making of meaning*. Dordrecht: Kluwer.

Gould, S. J., & Lewontin, R. C. (1979). The spandrels of San Marco and the Panglossian paradigm: A critique of the adaptationist programme. *Proceedings of the Royal Society of London, 205*(1161), 581–598.

Griffiths, O., Cook, A., & Wells, S. M. (1993). The diet of the introduced carnivorous snail *Euglandina rosea* in Mauritius and its implications for threatened island gastropod faunas. *Journal of the Zoological Society of London, 229,* 79–89.

Hilton, D. J. (2002). Thinking about causality: Pragmatic, social and scientific rationality. In P. Carruthers, S. Stich, & M. Siegal (Eds.), *The cognitive basis of science* (pp. 211–231). Cambridge: Cambridge University Press.

Hume, D. (1748/1964). Eine Untersuchung über den menschlichen Verstand. In R. Richter (Ed.), *Philosophische Bibliothek* (Vol. 35). Hamburg: Meiner.

Huxley, T. H. (1870/1970). *Biogenesis and Abiogenesis*. Hildesheim: Georg Olms.

Johnson-Laird, P. N., & Byrne, R. M. J. (1991). *Deduction*. Hillsdale: Lawrence Erlbaum Publishers.

Kitcher, P. (1993). *The advancement of science: Science without legend, objectivity without illusions*. Oxford: Oxford University Press.

Klahr, D., Fay, A. L., & Dunbar, K. (1993). Heuristics for scientific experimentation: A developmental study. *Cognitive Psychology, 25*(1), 111–146.

Knobe, J. (2003). Intentional action in folk psychology: An experimental investigation. *Philosophical Psychology, 16*(2), 309–324.

Kowarik, I. (2003). *Biologische Invasionen: Neophyten und Neozoen in Mitteleuropa*. Stuttgart: Ulmer.

Krebs, J. R., & Davis, N. B. (1996). *Einführung in die Verhaltensökologie*. Oxford: Blackwell.

Kühberger, A. (1998). The influence of framing on risky decisions: A meta-analysis. *Organizational Behaviour and Human Decision Processes, 75,* 23–55.

Kuhn, T. S. (1962/1976). *Die Struktur wissenschaftlicher Revolutionen*. Frankfurt a. M.: Suhrkamp.

Langston, N. (1995). *Forest dreams, forest nightmares: The paradox of old growth in the Inland West*. Seattle: University of Washington Press.

Laudan, L. (1977). *Progress and its problems*. Berkeley: University of California Press.

Little, C. E. (1995). *The dying of the trees: The pandemic in America's forests*. New York: Viking Penguin.

Low, T. (1999). *Feral future: The untold story of Australia's exotic invaders*. Chicago: University of Chicago Press.

Mead, A. R. (1961). *The Giant African snail: A problem in economic malacology*. Chicago: University of Chicago Press.

Medwedew, G. (1991). *Verbrannte Seelen: Die Katastrophe von Tschernobyl*. München: Hanser.

Nersessian, N. J. (1987). A cognitive-historical approach to meaning in scientific theories. In N. J. Nersessian (Ed.), *The process of science: Contemporary philosophical approaches to understanding scientific practice* (pp. 161–177). Dordrecht: Kluwer.

Nersessian, N. J. (2002). The cognitive basis of model-based reasoning in science. In P. Carruthers, S. Stich, & M. Siegal (Eds.), *The cognitive basis of science* (pp. 133–153). Cambridge: Cambridge University Press.

Newell, A., & Simon, H. A. (1972). *Human problem solving*. New Jersey: Prentice Hall.

Ogutu-Ohwayo, R. (1990). The decline of the native fishes of lakes Victoria and Kyoga (East Africa) and the impact of introduced species, especially the Nile perch, *Lates niloticus*, and the Nile tilapia, *Oreochromis niloticus*. *Environmental Biology of Fishes, 27*(2), 81–96.

Papineau, D. (2000). The evolution of knowledge. In P. Carruthers & A. Chamberlain (Eds.), *Evolution and the human mind: Modularity, language and meta-cognition* (pp. 170–206). Cambridge: Cambridge University Press.

Papini, M. (2002). *Comparative psychology: Evolution and development of behavior*. New Jersey: Prentice Hall.

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1993). *The adaptive decision maker*. Cambridge: Cambridge University Press.

Pimentel, D., Zuniga, R., & Morrison, D. (2005). Update on the environmental and economic costs associated with alien-invasive species in the United States. *Ecological Economics, 52,* 273–288.

Pinker, S. (1998). *Der Sprachinstinkt*. München: Droemer Knaur.

Plous, S. (1993). *The psychology of judgment and decision making*. New York: McGraw-Hill.

Reason, J. (1987). The Chernobyl errors. *Bulletin of the British Psychological Society, 40,* 201–206.

Reason, J. (1990). *Human error*. Cambridge: Cambridge University Press.

Rheinberger, H.-J. (1998). Kurze Geschichte der Molekularbiologie. In I. Jahn (Ed.), Geschichte der Biologie (pp. 642–663). Jena: Gustav Fischer.

Rubinstein, R. A. (1984). *Science as a cognitive process: Toward an empirical philosophy of science*. Philadelphia: University of Pennsylvania Press.

Ruse, M. (1988). *Philosophy of biology today*. New York: State University of New York.

Ruse, M. (1995). *Evolutionary naturalism: Selected essays*. London: Routledge.

Schönwandt, W. (1986). *Denkfallen beim Planen*. Braunschweig: Vieweg.

Sober, E. (1993). *Philosophy of biology*. Oxford: Oxford University Press.

Solomon, M. (1994). Social empiricism. *Noûs, 28*(3), 325–343.

Sukopp, T. (2006). *Naturalismus – Kritik und Verteidigung erkenntnistheoretischer Positionen*. Heusenstamm: Ontos.

Thagard, P. (1998a). Ulcers and bacteria I: Discovery and acceptance. *Studies in History and Philosophy of Biological and Biomedical Sciences, 29*(1), 107–136.

Thagard, P. (1998b). Ulcers and bacteria II: Instruments, experiments, and social interactions. *Studies in History and Philosophy of Biological and Biomedical Sciences, 29*(2), 317–342.

Trout, J. D. (2002). Scientific explanation and the sense of understanding. *Philosophy of Science, 69,* 212–233.

Tweney, R. D. (2001). Scientific thinking: A cognitive-historical approach. In K. D. Crowley, C. D. Schunn, & T. Okada (Eds.), *Designing for science: Implications from everyday, classroom, and professional settings* (pp. 141–176). Hillsdale: Lawrence Erlbaum Publishers.

Tweney, R. D., Doherty, M. E., & Mynatt, C. R. (Eds.) (1981). *On scientific thinking*. New York: Columbia University Press.

USDA Forest Service (2005). *Projects and policies: four threats*. http://www.fs.fed.us/projects/four-threats/. Accessed 6th August 2014.

Wimsatt, W. C. (1986). Heuristics and the study of human behavior. In D. W. Fiske, & R. A. Shweder (Eds.), *Metatheory in social science: Pluralisms and subjectivities* (pp. 293–314). Chicago: University of Chicago Press.

Chapter 4
The ‘Meme’ Meme Revisited

Alan Costall

[…] neo-Darwinism involves a breach between organism and nature as complete as the Cartesian dualism of mind and matter.
(Waddington 1957, ix)

4.1 Introduction

In a lecture to the *British Association of Science* in 1874, Thomas Huxley attempted to reconcile animal awareness with the Cartesian notion of animals as mere automata. He proposed that the consciousness of animals should be regarded as a “collateral product” of the working of the body, in much the same way that the whistle on a steam locomotive is incidental to how the engine itself functions (Huxley 1874/1896, p. 240).[1] Huxley went on to argue that this also holds for people: The contents of consciousness are simply symbols of events occurring quite automatically within the body. Thus “the feeling we call volition is not the cause of a voluntary act, but the symbol of that state of the brain which is the immediate cause of that act”. For those who remained sceptical—or who were, at least, prepared to take their feeling of scepticism seriously—Huxley offered the following proof:

> Any one who cares to run a pin into himself may perform a sufficient demonstration of the fact that a mode of motion of the nervous system is the immediate antecedent of a state of consciousness. All but the adherents of “Occasionalism”, or of the doctrine of “Pre-established Harmony” (if any such now exist) must admit that we have as much reason for regarding the mode of motion of the nervous system as the cause of the state of consciousness, as we have for regarding any event as the cause of another. (Huxley 1874/1896, p. 239)

---

[1] The analogy of the locomotive engine has given way to the computer metaphor: “Consciousness may be like the heat or the hum or the smell of the computer. It is a side effect of the particular hardware and software being used, but is not of any particular importance in understanding that hardware or software” (Thagard 1986, p. 311).

A. Costall (✉)
Department of Psychology, University of Portsmouth, Portsmouth, UK
e-mail: alan.costall@port.ac.uk

T. Breyer (ed.), *Epistemological Dimensions of Evolutionary Psychology,*
DOI 10.1007/978-1-4939-1387-9_4

Darwin himself was not convinced. When the lecture was later reprinted, he simply congratulated Huxley on providing such a good example: "I wish to God there were more automata in the world like you" (Darwin 1882/1887, p. 358). However, to another correspondent he confessed that "If I were as well armed as Huxley I would challenge him to a duel on this subject" (1882/1903, p. 399).

Somewhat later, Morgan (1990) turned Huxley's argument neatly on its head. Morgan raised the perfectly reasonable question of what would prompt anyone to stick a pin into themselves in the first place—other than the wish, for example, to demonstrate a dubious theoretical point. So, from this same example of the pinprick, one could, Morgan argued, just as well conclude that consciousness was the antecedent cause of the so-called physical stimulus (Morgan 1900, pp. 308–309).

Now all this certainly goes to show that there is more than one way to take a pinprick. But it also points to an enduring tension within Darwinian theory. Huxley's revival of the Cartesian reflex theory was a crucial element in the neo-Darwinian attempt to explain evolution—including mental evolution—in strictly mechanistic terms. Indeed, as William McDougall noted, "the theory of the evolution of mind was established at the same time that the theory of evolution by, or through the agency of, mind was sapped and rejected" (McDougall 1925, p. 328).

According to McDougall, Huxley's automaton theory was as important a challenge to Lamarckian theory as Weismann's doctrine of the separation of the reproductive cells from the rest of the body.[2] Yet the turn towards mechanistic explanation did not have an impact on evolutionary thought solely through its attack on the Lamarckian principle of the inheritance of acquired characteristics.[3] It also ran counter to Darwin's own approach to natural selection. This is why Romanes coined the term "*Neo*-Darwinians" to distinguish the revisionists from those who, like himself, continued to adhere "to the unadulterated doctrines of Darwin" (Romanes 1895, p. 9).

This was also the point of Lloyd Morgan's criticism of Huxley's example of the pinprick. For although Morgan came to reject the Lamarckian principle of inheritance, he did not see why he had to accept the automaton theory in its place. In fact, Morgan, in association with James Mark Baldwin, developed an extensive theoretical account of the place of intentionality, intelligence, and sociality in natural selection (e.g., Baldwin 1902, 1910; Morgan 1900; see Weber and Depew 2003). Evolutionary theory, they argued, involves psychology and sociology as much as biology. Furthermore, evolutionary theory, properly thought through, challenges the

---

[2] It is curious that Weismann's doctrine of the continuity of the germ plasm was so widely accepted. As Hill (1893, p. 84) noted, "we must remember that the doctrine of the germ-plasm is itself purely theoretical".

[3] Lamarck's principle of the inheritance of acquired characteristics was a supplement to his orthogenetic theory of evolutionary transformation: the continuing spontaneous emergence of organisms and their subsequent spontaneous differentiation into more and more complex organic forms. The inheritance of acquired characteristics was invoked to explain adaptation to the specific environmental conditions. Before Darwin, adaptation and transformation were widely regarded as contradictory on the grounds that any change would surely destroy the adaptive relations between organisms and their circumstances.

existing disciplinary boundaries, based as they are on dualisms, such as mind and matter, biology and culture, individual and society. Their main objection to *neo*-Darwinism lay in its revival of Cartesian psychology:

> The pendulum swing of opinion has, under the teaching of Professor Weismann, swung so far in the direction of the non-acceptance of the hereditary transmission of characters individually acquired through intelligent adjustment or otherwise; that the part played by consciousness in the evolution of the higher and more active animals is apt to pass unnoticed or unrecorded. It is well, therefore, to put in a reminder that a great number of animals would never reach the adult state in which they pass into the hands of the comparative anatomist save for the acquisition of experience, and the effective use of the consciousness to which they are heirs; that their survival is due, not only to their possession of certain structures and organs, but, every whit as much, to the practical use to which these possessions are put in the give and take of active life. (Morgan 1900, pp. 310–311; see Costall 1993)

4.2 Memes

In his book, *The Selfish Gene,* Richard Dawkins drew an analogy between biological and cultural evolution. He coined the term "meme" for what he called "a unit of cultural transmission or a unit of imitation" (1978, p. 206):

> Just as not all genes which can replicate do so successfully, so some memes are more successful in the meme-pool than others. This is the analogue of natural selection […] [Qualities] which make for high survival values among memes […] must be the same as those discussed for [genes]: longevity, fecundity, and copying-fidelity. (Dawkins 1978, p. 208)

Dawkins's formulation of the analogy is *neo*-Darwinian. People, on his view, are merely *passive* vehicles for both genes and memes. Dawkins' intention was to displace intentionality and agency from biological discourse, except, that is, for those scheming, selfish genes![4]

Dawkins identified Darwinism itself as an example of the meme. He conceded that Darwin would "scarcely recognize his own original theory" in *The Selfish Gene,* but then continued: "there is something, some essence of Darwinism, which is present in the head of every individual who understands the theory" (1978, p. 210). There is no indication in Dawkins' account that Darwin had also drawn analogies between organic and cultural change. Indeed, apart from a reference to Popper's comparison of cultural and genetic evolution as Darwinian processes (see Popper 1982), Dawkins leaves us with the impression that he has come up with a pretty bright new idea (cf. Campbell 1974).

Perhaps the earliest attempt to define a cultural analogue of the *gene* is to be found in a paper by Gerard et al. (1956), which actually refers to a "cultural genotype"

---

[4] "R. Dawkins […] is, of course, aware that his title is simply a clever gimmick—for only an extreme 'mentalist' could assume that a gene has the experience of being selfish. […] His restriction of memes to man amounts to saying that, from the evolutionary point of view, we can afford to ignore all the evidence of consciousness in animals; and the consequent development of primitive proto-cultures, as in primates" (Thorpe 1978, p. 75).

(see Ingold 1986, p. 362). But the selection analogy as such—framed in terms of unit ideas and unit phenotypic traits—was being deployed even before Mendel's work was rediscovered at the beginning of the twentieth century and before the gene concept was eventually reconciled with Darwinian theory in the "new synthesis".

Here, for instance, is a passage from William James's *Principles of Psychology* of 1890, where he repeats an argument he had made some 10 years before:

> Every scientific conception is, in the first instance, a "spontaneous variation" in someone's brain. For one that proves useful and applicable there are a thousand that perish through their worthlessness. Their genesis is strictly akin to that of the flashes of poetry and sallies of wit to which the instable brain-paths equally give rise. But whereas the poetry and wit […] have to run the gauntlet of no further test, the "scientific" conceptions must prove their worth by being "verified". This test, however, is the cause of their preservation, not of their production. (James 1990, p. 636)

James's friend, Charles S. Peirce, also considered the selection analogy and identified *three* evolutionary models of human thought and science: development "by the formation of habits" (Lamarckism), "by the violent breaking up of habits" (mutationism), and "by the action of innumerable fortuitous variations of ideas combined with differences in the fecundity of different variations" (Darwinism). He argued that creative science is mutationist "through the violent breaking up of certain habits, combined with the action of other habits not broken up". There is something distinctly "Kuhnian" not only in Peirce's treatment of revolutionary science but of normal science as well. He claimed that the "ordinary successful prosecution of scientific inquiring" is Lamarckian in that "it is growth by exercise, or by the direct efforts in the direction of growth". Thus Peirce considered the Darwinian model, but only to reject it. As far as the history of science was concerned, he concluded that Darwinism "made […] no figure at all" (Peirce 1892/1966, p. 259).

Although James Mark Baldwin applied the concept of selection to both biological evolution and what he called "imitative selection", he emphasized the *differences* in the nature of inheritance in the two cases:

> In so-called "imitative selection" […] the imitative propagation of ideas in society—we have a phenomenon which biology shows us no analogies. What survives in this case is not individuals, but *ideas*, and these do not survive in the form in which the first thinker conceives them, but in the form in which society applies them. Again, their fitness is not in any sense a fitness for struggle, it is a fitness for imitative reproduction and application. And, finally, they are not physically inherited, but handed down by 'social heredity' as accretions to the store of traditions. (Baldwin 1897, pp. 182–183)

The selection analogy has been applied to aspects of culture other than intellectual development. A striking example comes from the work of the marine biologist and anthropologist Alfred Haddon on the "life-histories of designs". Haddon explored the analogy in some detail. He argued that the determination of the provenance of a decorative form should, as in biology, be informed by careful classification of its structure, complemented by an analysis of its geographical distribution (Haddon 1895, pp. 319–320). Furthermore, in order to clarify the processes of transformation, the anthropologist should again follow the biologist's lead and conduct "breeding" experiments where changes in cultural forms are recorded as they occur across different "generations" (Haddon 1895, p. 311). Haddon cites the experiments conducted by Pitt-Rivers and Balfour in which they asked one person to copy a design and then pass it on to another person who, in turn, copied and passed it on, and so on. These experiments demonstrated a process of serial transformation. Haddon's basic point, however, concerned the analogy of selection, the idea that the fate of a design is to be "subject to vicissitudes very similar to those which beset the existence of any organism" (Haddon 1895, p. 310).

Haddon acknowledged that it would be absurd "to endeavour to make the evolution of decorative art run on all fours with that of animals". Yet his discussion is resolutely *neo*-Darwinian in its suppression of intentionality:

> [It may be objected] that the decorated objects […] are merely pieces of wood or stone, and that they are therefore not to be compared with living beings. It is perfectly obvious that ethnological objects cannot change themselves or develop themselves into anything else. On the other hand, though animals are alive they also have no voluntary power to alter themselves, nor can they develop themselves in any direction. They are almost as passive as fabricated objects. (Haddon 1895, p. 315)

So what about human beings? After some dismissive comments about children and "savages", Haddon considers the "civilized races":

> [With] increased complexity comes augmented mental power, and it may be argued that this may, so to speak, take the helm; but I would venture to ask, Is there much evidence in support of this view? The mind of man is subject, like his body, to the ordinary operations of the universe, his individuality is apparent rather than real, […] [Mind] cannot escape from the forces which act on the body. (Haddon 1895, p. 317)

Strangely, it is only at the end of his argument that Haddon notes that the question of intentionality might well arise not just for living beings *but for fabricated objects as well*. Almost as an afterthought, he adds:

> To return to our immediate subject, consciousness of purpose has extremely little to do with human evolution, nor has it much more to say to the evolution of patterns among primitive peoples. The selection of one design instead of another, or of a particular part of a design, is a conscious act, but probably in the great majority of cases an unreasoning one […]. While there is a certain amount of conscious selection, the variation as a whole of any design is an entirely unguided operation so far as the intelligence of the human units is concerned. (Haddon 1895, pp. 317–318; emphasis added)

So far, the examples I have chosen have shared two important features. First of all, there is an asymmetry in the way the selection analogy is formulated: The understanding of cultural change is to be *derived* from biological explanation rather than the reverse. Secondly, intentionality, whether in relation to the biological or the cultural, is dismissed as irrelevant to the analogy. We are indeed dealing with "fortuitous variations" (Peirce 1892/1966, p. 259), *blind* not only to their ultimate effects but also to their immediate consequences. Yet, when we consider examples closer to Darwin's time, the selection analogy undergoes an important reversal.

Darwin's son, George, in a contribution to the September 1872 issue of *Macmillan's Magazine,* sought to illustrate "the almost infinite ramifications to which natural selection […] may be applied". As he put it:

> A new invention bears a kind of analogy to a new variation in animals; there are many such inventions, and many such variations; those that are not really beneficial die away, and

> those that are really good become incorporated by 'natural selection', as a new item in our system. (Darwin 1872, p. 410)

George Darwin took as his example the development of the dress. Just by considering a single article of clothing, he managed to bring out the principles of selective advantage, continuity, vestigial structures, and even sexual selection. Here, for example, is his account of the evolution of the busby:

> It would now appear to be merely a fancy head-dress, but on inspection it proves not to be so. The hussar was originally a Hungarian soldier, and he brought his hat with him to our country. I found the clue to the meaning of the hat in a picture of a Hungarian peasant. He wore a red night-cap […], but the cap was edged with so broad a band of fur, that it made in fact a low "busby". And now in our hussars the fur has grown enormously, and the bag has dwindled into a flapping ornament, which may be detached at pleasure. Lastly, in the new "busby" of the Royal Engineers the bag has vanished, although the top of the cap (which is made of cloth and not of fur) is still blue, as was the bag formerly; the top cannot, however, be seen, except from a bird's-eye view. (Darwin 1872, p. 412)

George Darwin, here, is not generalizing the biological principle of natural selection to artefacts. Rather, he is arguing *from* plausible examples of *cultural change* for the credibility of a similar process in relation to *organic* evolution. And, at the time he was writing, such indirect "evidence" was sorely needed. There was no direct evidence of *gradual* organic change to support Darwinian theory (Costall 1985).[5]

4.3 Darwin's Metaphors

In the examples I have been considering, there are some cases where the meme concept is not only based on an analogy with natural selection but it also denies a role for intentionality in cultural change. Sometimes, however, the direction of the argument is reversed. Cultural change is invoked to argue for the plausibility of organic change, and in these cases the cultural change does unquestionably entail intentionality.

Charles Darwin invoked two metaphors based on cultural practices. The first is not well known. In *The Origin,* he presents the example of language to help explain the notions of descent and modification, and extensive use of the example of historical linguistics was made shortly afterwards by Lyell, Müller, and Schleicher to enhance the credibility of Darwin's account of organic change (see Richards 1987).[6] Now, *if* the denial of intentionality were crucial to the Darwinian account of natural selection, why did Darwin use the example of language to illustrate not merely the ideas of transformation and phylogenetic relationship *but also the process of selection itself?* After all, the use of language is both intentional and intelligent.

---

[5] This was the great attraction of analogies between evolution and individual development, especially the theory of recapitulation.

[6] Lyell, however, was not convinced that natural selection could explain the origin of our capacity for language, and Müller was convinced that it could not.

When Darwin and the early Darwinians drew upon examples of cultural transformation, it was not as though culture itself had not already been theorized by, among others, the political economists (see Schweber 1985; Young 1985).[7] Darwin himself explicitly mentioned the role of Malthus in his thinking, but Adam Smith is perhaps more relevant here.

The problem many later commentators have had with Darwin’s appeal to intentionality is that it seems at odds with his insistence that the process of longer-term transformation is not itself purposeful. Consider, then, the following passage from a letter Darwin wrote to the American philosopher, Chauncey Wright, where he specifically raises the question of the intentionality of language:

> As your mind is so clear, and as you consider so carefully the meaning of words, I wish you would take some incidental occasion to consider when a thing may properly be said to be effected by the will of man. I have been led to the wish by reading an article by your Professor Whitney versus Schleicher. He argues, because each step of change in language is made by the will of man, the whole language so changes; but I do not think that this is so, as man has no intention or wish to change the language. It is a parallel case with what I have called “unconscious selection”, which depends on men consciously preserving the best individuals, and thus unconsciously altering the breed. (Darwin 1872/1887, p. 164)

According to Darwin, linguistic change should therefore be regarded as an unintended effect of *intentional* activity. Darwin’s second metaphor of selection is much more familiar. Darwin presented his concept of *natural* selection in a remarkably anthropomorphic way:

> It may be said that natural selection is daily and hourly scrutinising, throughout the world, every variation, even the slightest; rejecting that which is bad, preserving and adding up all that is good; silently and insensibly working, whenever and wherever opportunity offers, at the improvement of each organic being in relation to its organic and inorganic conditions of life. (Darwin 1859, p. 84)

Of course, we should not take Darwin’s language too literally here. Yet, the question of intentionality certainly does arise in connection with *artificial* selection, the basis of Darwin’s fundamental metaphor. As in the example of language, we are once again presented with an analogy based on a *cultural practice,*[8] although in this case the activity would appear to involve not just intentionality but intentions aimed specifically at the process of *selection*.

However, as Secord (1981) has explained, Darwin insisted upon a distinction between *methodical* selection and *unconscious* selection, and it was the latter that was relevant to his analogy.[9] This distinction involved two different considerations.

---

[7] Such cultural theory could take a distinctly evolutionist turn, invoking notions of progress, stages, and “parallels” with biological transformation (see Greene 1981). In fact, Darwinism needs to be carefully distinguished from evolutionism.

[8] The critical importance of animal and plant breeding to the Victorian economy has been stressed by Secord (1985) in a paper on Darwin’s relation to the breeders. As he points out, even after the Industrial Revolution, England remained “a largely agricultural nation with an important proportion of the populace engaged in the production of food, and most wealth and power securely anchored in the land” (Secord 1985, p. 521).

[9] Of course, the evidence that selective breeding, deliberate or incidental, could effect transformations was a useful “existence proof” that living beings could undergo transformation. The implica-

First of all, at the level of the individual breeder, selection might occur as an incidental effect:

> At the present time, eminent breeders try by methodical selection, with a distinct object in view, to make a new strain or sub-breed, superior to anything existing in the country. But, for our purpose, a kind of Selection, which may be called Unconscious, and which results from every one trying to possess and breed from the best individual animals, is more important. Thus, a man who intends keeping pointers naturally tries to get as good dogs as he can, and afterwards breeds from his own best dogs, but he has no wish or expectation of permanently altering the breed. (Darwin 1859, p. 35)

The additional consideration is that the effects of the breeders cannot be understood solely at an individual level. Each breeder plays a part in a wider process that transcends their individual concerns. For example:

> In man's methodical selection, a breeder selects for some definite object, and free intercrossing will wholly stop his work. But when many men, without intending to alter the breed, have a nearly common standard of perfection, and all try to get and breed from the best animals, much improvement and modification surely but slowly follow from this unconscious process of selection, not withstanding a large amount of crossing with inferior animals. (Darwin 1859, p. 102)

According to Secord, Darwin had to downplay the role of methodical selection if natural selection were not, after all, to imply design. Thus, "the element in artificial selection of conscious, careful planning by the breeder had to be *minimized*" (Secord 1981, p. 185; emphasis added). But what also has to be noted in relation to Darwin's discussion of artificial selection and linguistic change is that he did not *eliminate* intentionality. Intentionality figures as an essential feature of the accounts Darwin presents. And, as we shall now see, this is also true, to an important extent, of Darwin's explanation of *natural* selection.

4.4 Darwin's Intentions

The modern proponents of a mechanistic interpretation of Darwinism can hardly deny that there is much in Darwin's own writings compatible with a "psychological" or "intentionalist" account. However, they usually argue that we should not take Darwin's 'anthropomorphic' talk seriously at all. His references to "intentionality" were unintentional! For example, Gillian Beer has argued that Darwin had little choice when developing his own ideas but to formulate them in terms of the earlier "discourse" of natural theology (Beer 1983, pp. 24, 53, 69). Furthermore, there was also his continued appeal to the inheritance of acquired characteristics as a supplement to his own theories of natural and sexual selection. Thus, it would

---

tion, however, of Darwin's own argument concerning unconscious selection was that, since such selection was unwitting and based on very minute variations (rather than striking novelties), the changes would not be noticed, and hence not be recorded, by the breeders. Thus, artificial selection could not, as Darwin himself recognized, provide the evidence for gradual transformation he really needed.

seem Darwin’s talk of “will” and “intention” could be dismissed as simply a lapse into Lamarckism or anthropomorphism.

However, Darwin made deliberate and systematic use of intentionalist language not solely in relation to the supplementary principle of “Lamarckism”.[10] The theory of sexual selection, for example, specifically refers to choice on the part of the female. Even in his researches on the movements of plants and the behaviour of earthworms, Darwin emphasized their non-reflexive, purposive character (Allen 1977; Reed 1982). Thus, on the basis of a series of remarkable experiments on the ability of earthworms to plug up their burrows, Darwin concluded that earthworms simply had to be credited with some degree of intelligence:

> They act in nearly the same manner as would a man, who had to close a cylindrical tube with different kinds of leaves, petioles, triangles of paper, &c. For they commonly seize such objects by their pointed ends. But with thin objects a certain number are drawn in by their broader ends. They do not act in the same unvarying manner in all cases, as do most of the lower animals; for instance, they do not drag in leaves by their foot-stalks, unless the basal part of the blade is as narrow as the apex, or narrower than it. (Darwin 1881, p. 313)

Darwin’s work on earthworms also has a wider implication. It provides a striking example of how animals do not merely *react* to circumstances, but actually create their own circumstances. After all, the environment that now surrounds earthworms did not, of course, precede their evolution (Ghilarov 1983).

In the first edition of *The Origin,* Darwin considered the implications of an observation by Samuel Hearne that the black bear swims along with its mouth wide open in order to “fish” for flies trapped on the surface of the water. The possible evolutionary implications did not escape Darwin’s attention:

> Even in so extreme a case as this, if the supply of insects were constant, and if better adapted competitors did not already exist in the country, I can see no difficulty in a race of bears being rendered, by natural selection, more and more aquatic in their structure and habits, with larger and larger mouths, till a creature was produced as monstrous as a whale. (Darwin 1859, p. 184)

Here, as in the earlier example of the behaviour of earthworms, we find Darwin formulating not a mechanistic, but a dialectical relation between the organism and its environment. In this case, the “initiative” of the bear in both taking to the water and fishing for flies radically changes the context of selection. And if this merely *sounds* like the inheritance of acquired characters, it does surely echo Lamarck’s more fundamental point that the “habits and mode of life” of an animal are an important factor in shaping “its body, the number and state of its organs and lastly, the faculties which it possesses” (Lamarck 1809/1983, p. 114).

Although the bear example makes no reference to the inheritance of acquired characters (it was formulated solely in terms of natural selection), Darwin’s friends urged him to remove it from later editions of *The Origin,* and he complied. Nevertheless, as Beer (1963, p. 176) put it, Darwin “always regretted that he had jettisoned

---

[10] Darwin was very aware that he needed to distance himself from Lamarck. In this respect, therefore, “Lamarckism” would have discouraged him from making too free use of intentionalist descriptions. For an important clarification of Darwin’s use of Lamarckism, see Montgomery (1985).

his bear", since he remained convinced that both his argument and the basic observations were sound (see Darwin 1860/1903, p. 162).

The explorer James Lamont enthusiastically acknowledged Darwin's bear example when he presented his own account of how the polar bear might have originated from individuals of *Ursus arctos* in Siberia, who, "finding their means of subsistence running short, and pressed by hunger, ventured on the ice and caught some seals". Once the bears had settled on the shore, Lamont argued, the palest among them would be selected because they would be less conspicuous to the seals. Thus, "the process of Natural Selection would do the rest, and *Ursus arctos* would in the course of a few thousands, or a few millions of years, be transformed into the variety at present known as *Ursus maritimus*" (Lamont 1861/1903, p. 179).

Darwin quickly got in touch with Lamont to congratulate him on having grasped his point:

> As you thought it out independently, it is no wonder that you so clearly understand Natural Selection, which so few of my reviewers do or pretend not to do. I never expected to see any one so heroically bold as to defend my bear illustration […] It is laughable how often I have been attacked and misrepresented about this bear. I am much pleased with your remarks, and thank you cordially for coming to the rescue. (Darwin 1861/1903, p. 179)

Thus, in his discussions of not only linguistic change and artificial selection but also of the adoption by animals of new modes of life, Darwin regarded intentional activity, through its unintended effects, as an important basis of evolutionary change.[11]

4.5 Interplay

To a remarkable extent, the analogies that have been drawn between the cultural and biological eventually bring us back to where we started from—without us noticing. This was the point of Friedrich Engels's famous protest against the rise of Social Darwinism:

> The whole Darwinist teaching of the struggle for existence is simply a transference from society to living nature of Hobbes's doctrine of bellum omnium contra omnes and of the bourgeois-economic doctrine of competition together with Malthus's theory of population. […] the same theories are transferred back again from organic nature into history and it is now claimed that their validity as eternal laws of human society have been proved. (Engels 1875, cited in Parsons 1977, pp. 142–143)

The same "conjurer's trick" (to use Engels's term) has also been played on us in relation to other key terms in evolutionary biology. The original sense of "inheritance" concerned the (usually) intentional passing on of money or property from parents to their descendants. It was an act of *will*. However, it came to refer to how "like-begets-like", the resemblance of parents and offspring (Mameli 2005). (Shakespeare was among the first to use the verb "inherit" in this sense.) This *social*

---

[11] Lewontin, an impressive critic of the mechanistic approach of neo-Darwinism, has also consistently misrepresented Darwin himself as a Cartesian (e.g., Lewontin 1982).

metaphor had already died by the time that Baldwin (1897, pp. 182–183) coined the metaphor of *social* heredity (see also Odling-Smee 2007, on *ecological* inheritance).

Similarly, "trial and error" has come to imply an essentially "blind" process, both within biology and psychology. Yet the original reference was to human practice: an intelligent and systematic approach to problem solving within mathematics and science (see Still 1988). Like the selection metaphor, it was taken *from* human practice into biology and then projected *back* into psychological discourse, stripped of all intelligence, if not intentionality.

4.6 Conclusion

The analogies drawn between the psychological and the biological, including the "meme" meme, can take either of two opposing directions. The Darwinian schema of variation–selection–retention, as represented, for example, by Dawkins, involves a projection from the biological to the psychological, where intentionality is dismissed as either non-existent or an epiphenomenon. In other cases, and most notably in the formulation of Darwin's own theory, the transfer of analogy takes the other direction. Human practices, such as language and artificial selection, are invoked to throw light on the nature of "biological" evolution. Such explanations *begin* with human intentions, even if *the long-term consequences* do not themselves count as intentional (see also Allen 1983; Oldroyd 1984; Smith 1978, pp. 261–262).

In some cases, there is a boomerang effect when analogies are "thrown" from the cultural and psychological to the biological, only to return distorted and to cause further confusion. However, the basic problem does not concern the particular forms these analogies have taken, but their common starting point: the assumption that the psychological and biological constitute quite separate realms. They are not mutually exclusive,—there is an interplay (Henriques et al. 1984, pp. 21–22; Gerbault et al. 2011). Rethinking the relation of the "biological" and the "psychological" in a non-dualistic way challenges "DNA-centric theory" (Mameli 2005; Oyama 2003). It should also open up a space for intentionality within biology—as Darwin himself had intended.

References

Allen, M. (1977). *Darwin and his flowers*. London: Faber.

Allen, G. E. (1983). The several faces of Darwinism: Materialism in nineteenth and twentieth century thought. In D. S. Bendall (Ed.), *Evolution from molecules to men* (pp. 81–102). Cambridge: Cambridge University Press.

Baldwin, J. M. (1902). *Development and evolution*. New York: MacMillan.

Baldwin, J. M. (1910). *Darwin and the humanities* (2nd ed.). London: Swan Sonnerschein.

Baldwin, J. M. (1897). *Social and ethical interpretations in mental development: A study in social psychology.* New York: Macmillan Co.

Beer, G. (1983). *Darwin's plots*. London: Routledge and Kegan Paul.

Campbell, D. T. (1974). Unjustified variation and selective retention in scientific discovery. In F. Ayala & T. Dobzhansky (Eds.), *Studies in the philosophy of biology*. London: Macmillan.

Costall, A. (1985). Specious origins: Darwinism and developmental theory. In G. Butterworth, J. Rutkowska, & M. Scaife (Eds.), *Evolution and developmental psychology* (pp. 30–41). Brighton: Harvester Press.

Costall, A. (1993). How Lloyd Morgan's canon backfired. *Journal of the History of the Behavioral Sciences, 29,* 113–124.

Darwin, C. (1859). *On the origin of species by means of natural selection or the preservation of favoured races in the struggle for life*. London: John Murray.

Darwin, C. (1860/1903). Letter to W. H. Harvey, August 1860. In F. Darwin & A. C. Seward (Eds.), *More letters of Charles Darwin* (Vol. 1). London: John Murray.

Darwin, C. (1861/1903). Letter to James Lamont, February 25th, 1861. In F. Darwin & A. C. Seward (Eds.), *More letters of Charles Darwin* (Vol. 1). London: John Murray.

Darwin, C. (1872/1887). Letter to Chauncey Wright, June 3rd, 1872. In F. Darwin (Ed.), *Life and letters of Charles Darwin* (Vol. 3). London: John Murray.

Darwin, C. (1881). *The formation of vegetable mould, through the action of worms with observations on their habits*. London: John Murray.

Darwin, C. (1882/1887). Letter to T. H. Huxley. In F. Darwin (Ed.), *Life and letters of Charles Darwin* (vol. 3). London: John Murray.

Darwin, C. (1882/1903). Letter to John Collier, February 16th, 1882. In F. Darwin & A. C. Seward (Eds.), *More letters of Charles Darwin* (Vol. 1). London: John Murray.

Darwin, G. (1872). Development in dress. *Macmillan's Magazine*, September, 410–416.

Dawkins, R. (1978). *The selfish gene*. London: Paladin Books.

de Beer, G. (1963). *Charles Darwin*. London: Nelson.

Gerard, R. W., Kluckhohn, C., & Rapoport, A. (1956). Biological and cultural evolution: Some analogies and explorations. *Behavioral Science, 1,* 6–34.

Gerbault, P., Liebert, A., Itan, Y., Powell, A., Currat, M., Burger, J., Swallow, D. M., & Thomas, M. G. (2011). Evolution of lactose persistence: An example of human niche construction. *Philosophical Transactions of the Royal Society, 366,* 863–877.

Ghilarov, M. S. (1983). Darwin's 'Formation of vegetable mould'—its philosophical basis. In J. E. Satchell (Ed.), *Earthworm ecology: From Darwin to vermiculture* (pp. 1–4). London: Chapman & Hall.

Greene, J. C. (1981). Biology and social theory in the nineteenth century: Auguste Comte and Herbert Spencer. In J. C. Greene (Ed.), *Science, ideology, and world view: essays in the history of evolutionary ideas*. Berkeley: University of California Press.

Haddon, A. C. (1895). *Evolution in art as illustrated by the life histories of design*. London: Walter Scott, Ltd.

Hill, D. J. (1893). *Genetic philosophy*. New York: Macmillan.

Henriques, J., Hollway, W., Urwin, C., Venn, C., & Walkerdine, V. (1984). *Changing the subject: Psychology, social regulation and subjectivity*. London: Methuen.

Huxley, T. H. (1874/1896). On the hypothesis that animals are automata, and its history. *Fortnightly Review, 22,* 555–580. (reprinted in: T. H. Huxley, Methods and results: essays. New York: D. Appleton, 1896. A shorter version was published in Nature, 10, 362–366).

Ingold, T. (1986). *Evolution and social life*. Cambridge: Cambridge University Press.

James, W. (1990). *The principles of psychology* (Vol. 2). New York: Henry Holt.

Lamont, (1861/1903). Seasons with the sea-horses; or, sporting adventures in the Northern Seas. London. Cited in F. Darwin & A. C. Seward (Eds.), *More letters of Charles Darwin* (Vol. 1, p. 179). London: John Murray.

Lamarck, J. B. (1809/1983) *Zoological philosophy: An exposition with regard to the natural history of animals* (trans. H. Elliot). Chicago: University of Chicago Press. (First published in 1809; this translation first published in 1914, by Macmillan).

Lewontin, R. C. (1982). Organism and environment. In: H. C. Plotkin (Ed.), *Learning, development, and culture*. Chichester: John Wiley.

Mameli, M. (2005). The inheritance of features. *Biology and Philosophy, 20,* 365–399.

McDougall, W. (1925). Mental evolution. In *Evolution in the light of modern knowledge: A collective work*. London: Blackie and Son.

Montgomery, W. (1985). Charles Darwin's thought on expressive mechanisms in evolution. In C. G. Zivin (Ed.), *The development of expressive behavior: Biology-environment interactions*. New York: Academic Press.

Morgan, C. L. (1900). *Animal behaviour*. London: Edward Arnold.

Odling-Smee, J. (2007). Niche inheritance: A possible basis for classifying multiple inheritance systems in evolution. *Biological Theory, 2*(3), 276–289.

Oldroyd, D. R. (1984). How did Darwin arrive at his theory? The secondary literature to 1982. *History of Science, 22*, 325–374.

Oyama, S. (2003). On having a hammer. In B. H. Weber & D. J. Depew (Eds.), *Evolution and learning: The Baldwin effect reconsidered* (pp. 169–191). Cambridge: MIT Press.

Parsons, H. L. (1977). *Marx and Engels on ecology*. Westport: Greenwood Press.

Peirce, C. S. (1892/1966). Conclusion of the [Lowell] History of science lectures. In C. S. Peirce (Ed.), *Selected writings*. New York: Dover Books.

Popper, K. R. (1982). *Objective knowledge: An evolutionary approach*. Oxford: Clarendon Press.

Reed, E. S. (1982). Darwin's earthworms: A case study in evolutionary psychology. *Behaviorism, 10*(2), 165–185.

Richards, R. J. (1987). *Darwin and the emergence of evolutionary theories of mind and behavior*. Chicago: University of Chicago Press.

Romanes, G. J. (1895). The Darwinism of Darwin and of the Post-Darwinian Schools. *Monist, 6,* 1–27.

Schweber, S. S. (1985). The wider British context in Darwin's theorizing. In D. Kohn (Ed.), *The Darwinian heritage* (pp. 35–69). Princeton: Princeton University Press.

Secord, J. A. (1981). Nature's fancy: Charles Darwin and the breeding of pigeons. *ISIS, 72,* 163–186.

Secord, J. A. (1985). Darwin and the breeders: A social history. In D. Kohn (Ed.), *The Darwinian heritage* (pp. 519–542). Princeton: Princeton University Press.

Smith, C. U. M. (1978). Charles Darwin, the origin of consciousness, and panpsychism. *Journal of the History of Biology, 11,* 245–267.

Still, A. (1988). Word meaning and historical change: The case of "trial and error". *Proceedings of the 7th European Cheiron conference* (pp. 590–592). Budapest: Hungarian Psychological Association.

Thagard, P. (1986). Parallel computation and the mind-body problem. *Cognitive Science, 10,* 301–318.

Thorpe, W. H. (1978). *Purpose in a world of chance*: *A biologist's view*. Oxford: Oxford University Press.

Young, R. M. (1985). *Darwin's metaphor: Nature's place in Victorian thought*. Cambridge: Cambridge University Press.

Waddington, C. H. (1957). *The strategy of the genes*. London: George Allen & Unwin.

Weber, B. H., & Depew, D. J. (Eds.) (2003). *Evolution and learning: The Baldwin effect reconsidered*. Cambridge: MIT Press.

Chapter 5
Evolutionary Psychology, Altruism, and Kin Selection

Robert C. Richardson

5.1 The Darwinian Background

Charles Darwin famously wrote in the final chapter of *On the Origin of Species* that the theory of the *Origin* would “open fields for far more important researches. Psychology will be based on a new foundation, that of the necessary acquirement of each mental power and capacity by gradation. Light will be thrown on the origin of man and his history” (Darwin 1859, p. 488). This was his only mention of human evolution in the *Origin*. It is natural to think, as many have, that he meant that *natural selection* would shed light on “the origin of man and his history” and put psychology on a “new foundation.” However, that would be wrong. The context for the remark is a discussion of common descent, which meant descent from a common ancestor, and the mutability of species. Darwin sees that embracing “common descent”—one of his more radical departures from his evolutionary predecessors, such as Lamarck and Geoffroy—will result in “a considerable revolution in natural history.” In this context, neither natural selection nor sexual selection is even mentioned. He presses, e.g., that living things have much in common “in their chemical composition, their germinal vesicles, their cellular structure, and their laws of growth” (Darwin 1859, p. 484), inferring from this that they all have a common ancestor. This was important to Darwin. He was very conscious that adaptation and, thus, natural selection could disguise common descent rather than reveal it. He knew that selection could yield evolutionary convergence. Indeed, that is the point to thinking about convergence as a mark of adaptation, though Darwin was also conscious that such convergence is imperfect. When natural selection is at work, it is capable of producing the same, or nearly the same, structure in quite different organisms, quite apart from issues of common descent. Wings of bats and birds are a typical example (though both are forelimbs). It is, Darwin thought, *homologies* rather than *analogies*

R. C. Richardson (✉)
Department of Philosophy, University of Cincinnati, Cincinnati, (OH), USA
e-mail: robert.richardson@uc.edu

T. Breyer (ed.), *Epistemological Dimensions of Evolutionary Psychology,*
DOI 10.1007/978-1-4939-1387-9_5

that reflect common descent. The forelimbs of bats and birds exhibit parallel bone structures, even though the adaptations that make them effective wings are quite different. So in pressing for common descent, he does not emphasize natural selection, though he never doubts its significance.

In opening his discussion in the *Descent of Man*, some 12 years after the *Origin,* Darwin returned to the topic of *human* evolution:

> He who wishes to decide whether man is the modified descendant of some pre-existing form, would probably first enquire *whether man varies*, however slightly, in bodily structure and in mental faculties; and if so, *whether the variations are transmitted* to his offspring in accordance with the laws which prevail with the lower animals; such as that of the transmission of characters to the same age or sex. Again, are the variations the result, as far as our ignorance permits us to judge, of the same general causes, and are they governed by the same general laws, as in the case of other organisms? (Darwin 1871, p. 9)

Notice that the reference here is to evolution, or descent, and not to natural selection. Also, natural selection (but not sexual selection) plays a much less prominent role in *Descent* as opposed to *Origin.* There is appeal to variations and to inheritance and to the "laws" governing each of them, which includes the "mental faculties," but there is not a hint of competition or the "struggle for existence," much less of natural selection. In *Descent,* he initially recapitulates the argument for common descent from *Origin,* extending it to what he calls the "mental faculties" of man, saying at the outset of Chap. II that his object "is solely to show that there is no fundamental difference between man and the higher mammals in their mental faculties" (Darwin 1871, p. 35).

Darwin is clear that this commitment to evolution is meant to include what he calls the "moral sense." This was crucial for Darwin. It meant, among other things, that our capacities for social interaction were meant to be within the purview of his evolutionary theory. For Darwin, sympathy is not the whole story behind our moral capacities or our social capacities, but it is nevertheless essential for it. Much like David Hume, he thinks that without sympathy, the rest of our moral capacities would have no traction. He therefore conjectures that an animal with "social instincts" would almost certainly acquire something like a developed "moral sense," given even a modicum of intelligence (even though it is not clear to me how he thinks this to be a natural development). The additional capacities that are gained with the acquisition of linguistic skills, he thought, would enhance their responsiveness to social pressures. All of this, though, is laid on the foundation of our natural sympathies, our social responsiveness.

Darwin was not the first to bring evolutionary insights to the discussion of our social sentiments and reasoning, and he was not the last either. Herbert Spencer had placed his discussion of psychology in an explicitly evolutionary setting; William James' psychology is inspired by Darwinian insights, as were many other important psychologists at the turn of the century. There have been other ventures into the evolution of human psychology that are in retrospect less well regarded (be it deservedly or not), such as Desmond Morris' *The Naked Ape* (Morris 1967). With the elaboration of models designed to capture social behavior in the middle of the last century, sociobiology took up the task of capturing animal behavior in evolutionary terms and, almost as an appendix, extended that project to the domain of human social behavior. It is clear that E. O. Wilson thought that human social behavior was, after all, just another form of social behavior (Wilson 1975). He was right. It is also clear that for Wilson, this was initially an afterthought, almost like an appendix to a comprehensive work on social behavior. Work in evolutionary psychology (EP) is the most recent variation in taking up the Darwinian idea that evolution should shed light on human psychology. Unlike Wilson, however, this work is almost exclusively concerned with the human case.

5.2 The Program of Evolutionary Psychology

Contemporary EP is not a homogeneous collection of views, even with respect to its evolutionary commitments, though it is possible to articulate a loose set of claims that are broadly endorsed and typical of contemporary adherents (cf. Barkow et al. 1992; Buss 1995, 2005). In large parts, these commitments are consistent with evolutionary theory as it was articulated during the "evolutionary synthesis" years in the first half of the twentieth century, updated by evolutionary models from the 1960s. Not every advocate of EP is committed to precisely the same set of claims, but it is possible to provide a rough portrait of research interests, which include a number of familiar themes:

- *Psychological mechanisms are the result of natural selection and sexual selection.* While it is generally acknowledged that some outcomes of evolution are due to chance or by-products of selection for other traits, the focus of EP is on traits that are subject to selection, e.g., finding a mate, cooperative activities like hunting, or the raising of offspring. The assumption is that natural selection will tend to efficiently "solve" problems like these. The tendency is to offer only selectionist models without including possible alternatives. In terms of attractiveness, for example, there is a widely touted view about a certain female waist to hip ratio which is supposedly attractive to males, as it is thought to be connected to female fertility; thus, male attraction is "explained" as an adaptation which enhances male fertility.
- *Psychological mechanisms can be thought of as computational mechanisms.* Among psychological mechanisms are narrowly cognitive processes (e.g., probabilistic reasoning or problem solving) and emotional responses (e.g., jealousy or fear). The idea that psychological mechanisms are computational is an assumption common among a range of cognitive scientists, though its prevalence has faded considerably in the last decade. Alternatively, these computational mechanisms can be thought of as exhibited in and causing behavioral *strategies* that respond to environmental challenges. In either case, the psychological mechanisms are supposed to identify risks and benefits of the environment, e.g., strangers pose a risk, thus a fear of strangers is thought to be an evolutionarily favored strategy. Likewise, cooperation within a group, as in sharing of food, tends to enhance fitness over the longer run, therefore being an evolutionarily favored strategy as well.

- *Psychological mechanisms evolved in response to relatively stable features of ancestral environments.* This ancestral environment is often referred to as the environment of evolutionary adaptedness (EEA). EP asserts that most of human evolution took part during the Pleistocene (roughly 2.6 million to 12 thousand years ago), and presumably in the later Pleistocene. Some more ancient adaptations are fundamentally primate, while others are more specifically human. We thus share more ancient adaptations with our primate kin, while more recent adaptions are specifically human. Often, the EEA is identified with the savanna of the African Pleistocene and with a hunter-gatherer lifestyle. Sometimes, the EEA is also identified with a kind of statistical aggregate of the total range of ancestral environments. It is unclear what exactly a statistical aggregate in this case might mean and it is even more unclear how this might play out in an evolutionary scenario.
- *Since psychological mechanisms are adaptations to ancestral environments, there is no reason to assume that they are adaptive in contemporary circumstances.* Social environments are a significant part of the environment and are obviously crucial to human evolution. If we assume within the approach of EP that our ancestral social environment consisted of small, nomadic bands of relatives, then the difference between that and our contemporary culture suggests that whatever strategies were adaptive for our ancestors may not be so for us. Likewise, if we assume that our distant ancestors lived in a sugar-deprived environment, then our fondness for sweets might be "natural," though no longer adaptive. In general, EP assumes that evolutionary responses are too slow to have had any significant effect in the last 12,000 years, the earlier advent of agriculture and sedentary life. The picture is clear enough. We are adapted for a life in relatively small groups and at least some human adaptations are selected for their usefulness in these small groups.
- *The human mind is a kind of mosaic of mechanisms, each with some specific adaptive function, rather than merely a general purpose learning machine.* Different adaptive problems will require different solutions, or different strategies, in order to be able to deal with them. So, e.g., a mechanism for mate selection, for example, is unlikely to be of much use in foraging. At least some of this machinery must be domain specific, specialized for particular tasks. Some of these may count intuitively as *instincts*. Some advocates of EP treat these mechanisms as *modules,* though others insist that all that is required is distinct domain-specific mechanisms.

5.3 Psychological Evidence

Evolutionary psychologists make use of an array of techniques to evaluate their psychological models, most of which do not specifically depend on the evolutionary assumptions. These methods include the use of questionnaires, controlled experiments, observational methods, and brain imaging, among others. They also make use of a variety of less standardized techniques, including ethnographic records, paleontological information, and life history data. As one would expect, the specific methods and the evidence obtained through them, are sometimes contested; however, this does not seem to be the place to enter into these psychological disputes (cf. Buller 2005; for an in-depth exploration of the psychological issues). Reliance on a broad array of evidence would seem, *prima facie,* to be an epistemic virtue. Evolutionary assumptions are used to advance and formulate hypotheses. Whether they are more than *merely* heuristic is sometimes not clear.

Evolutionary psychologists have articulated and tested a wide array of psychological hypotheses inspired by evolutionary thinking. These include our propensity for aspects such as cooperation, cheater detection, differences in spatial memory, short-term mating preferences, and many others. Some simple examples may be sufficient to illustrate the method. Assume that human memory will be sensitive to items that affected fitness among our ancestors. This might include food items, shelter, or possible mates. The standard memory probes in nineteenth- and twentieth-century psychology relied on random lists, often terms or signs without meaning, to test recall. Using standard memory probes in a test on recall and recognition for lists of words, James Nairne and his collaborators found that recall for survival-oriented terms was significantly better than recall for more neutral words. This suggests that recall and recognition are affected by their felicity towards fitness. Similarly, theories of parental investment, prevalent among sociobiologists and evolutionary psychologists (see Trivers 1974), suggest that females will tend to prefer mates that are more likely to invest in offspring. Males and females also differ in the pattern of jealousy, with females being more sensitive to emotional infidelity (as a risk of abandonment) and males more sensitive to actual sexual infidelity (as a risk to paternity). I do not propose to focus too much on these findings, but to see them as insightful examples on second-generation advantages to fitness. Females are concerned with the fitness of their offspring rather than merely the quantity of offspring, whereas males are more concerned with whether a partner's offspring is his offspring. The predictions obviously concern only average differences: both cues may trigger jealousy, but males and females should differ in their sensitivity to the various cues. These predictions have been supported by straightforward evaluations of preferences using questionnaires, spontaneous recall, and fMRI (functional magnetic resonance imaging).

One of the most striking and prominent examples concerning human judgment focuses on what is known as the "Wason task," a standard tool used by psychologists to study reasoning about conditionals. The initial test was developed by the psychologist Peter Wason in the 1960s, and was subsequently elaborated in conjunction with P. N. Johnson-Laird. The format is familiar and simple. Subjects are given simple conditional rules of the form *If P, then Q,* and are provided four cards with different information. The available values are *P, not-P, Q*, and *not-Q.* The hidden side can contain any of these values as well. The experimental subjects are then asked which of the four cards would need to be turned over in order to be sure that the four cards conform to the generalization. With a formal problem and a variety of informal variants, subjects regularly fail to exhibit the required choices, most often picking *P* and *Q,* despite the fact that the *Q* card is irrelevant; moreover, the *not-Q* card is rarely picked even though it is directly relevant to the task. Falsification is not recognized as immediately relevant. Descriptive and strictly formal rules tend to elicit this "failure" of rationality. It turns out that some variations of this task are considerably easier for subjects to solve, though they are in fact superficially more complicated. More specifically, when the conditional rules explain what is permissible or forbidden, the subjects can systematically elicit more appropriate choices. For example, suppose that subjects are given the rule "If they are drinking beer, they must be 21," and four individuals, one clearly over 21, one under, one drinking Coke, and one drinking beer. The task is to determine whether any of these individuals is violating the rule. Typically, subjects have little difficulty in realizing that they need to check what the person under 21 is drinking, and whether the person drinking beer is 21; they are not inclined to check the one who is clearly 21. The problem is formally identical to the more abstract problem above, but leads to a very different performance.[1]

Cosmides and Tooby (1992) suggest that the differences in performance, despite the formal identity of the problems, are best explained by appealing to *ecological rationality,* a view supported by Gigerenzer and Hug (1992). Within the concept of *ecological rationality,* the human mind has a variety of domain-specific mechanisms governing our inferences that were adaptive in the EEA. When applied to an appropriate content domain, they are both efficient and reliable, yielding the "normatively appropriate" choices. In particular, Cosmides and Tooby (1992) claim that reasoning about social exchange can explain at least a large part of the differences on the Wason selection task, which basically depend on task demands. Social exchange is of course an important part of human social life, and presumably has been so for our hominid ancestors as well. What Cosmides and Tooby call "social contracts" encompasses rules expressing the contingencies of accepting a benefit on some identifiable requirement. In cases of social exchange, reciprocity or status is important, and as a consequence it is advantageous to be able to detect cheaters. There should, in other words, be a "cheater detection algorithm," naturally mobilized in the face of social obligations. It is not necessary here to go into the controversies regarding the psychological data. What is important, however, is that the hypothesis of a cheater detection algorithm is suggested based on a theory of reciprocal altruism, and that the corresponding social contract theory captures significant patterns in the behavioral data. Whether the evolutionary background is supposed to provide additional credentials to the psychological model is not clear in this case. Perhaps the psychological evidence alone would suffice. In other cases, the evolutionary models play a more central role.

---

[1] This is not true, though it's often touted as true. It is clear that *deontic* contexts concerned with obligation and permission are logically much more nuanced and complex. These issues are contested. What this means for human performance is unclear.

5.4 Evolutionary Models for Altruism

Relying on work in paleoanthropology and ethnography relating especially to contemporary hunter-gatherers, evolutionary psychologists have elaborated a plausible portrait of ancestral social life (which is however controversial among anthropologists). In the typical case, EP assumes that ancestral hominids lived in relatively compact groups of no more than one hundred. These groups were kin-based and characterized by a sexual division of labor, with males more engaged in hunting and females more engaged with gathering, stable male–female bonds with long periods of biparental care, and cooperative foraging. In addition, much is known about the physical environment. We know, for example, that they were subject to a variety of predators and pathogens and had to deal with a considerable variance in resources. We also know that the range of environments within which ancestral hominids lived and died was quite variable.

This knowledge allows EP to construct a variety of evolutionary scenarios. Depending on the case, they use a variety of resources from evolutionary biology, including theoretical models concerning reciprocal altruism, parental investment, kin selection, and evolutionary game theory. Beginning with the relevant dimensions assumed to be typical in the EEA, evolutionary psychologists construct an account of the adaptive functions that must be satisfied. These can be thought of as a kind of design specification. The psychological task is then to reverse engineer a solution to the adaptive problem. In some crucial cases, the "problem" is the evolutionary source of altruistic tendencies. This is the modern analogue for Darwin's interest in the social instincts such as empathy.

An evolutionary emphasis can lead to problems, especially if the "adaptive problem" is not clear. If the adaptive problem is underspecified, then there are issues over whether the evolutionary solution really is the right one. Underspecification will inevitably lead to underdetermination. It also raises issues for connecting the psychological hypotheses and the evolutionary interpretations. As this concerns a centerpiece of EP relevant to evolutionary biology, and as similar issues affect the use of other evolutionary models within EP, it is worth illustrating in more detail. The following case is built on one of the seminal accomplishments of twentieth-century biology and contains a considerable amount of anthropological evidence and is therefore often appealed to by advocates of EP.

One of the seminal figures in the rise of Darwinism in the mid twentieth century was W. D. Hamilton, who strongly influenced sociobiology and evolutionary psychology. His development of the theory of *inclusive fitness, extended fitness,* or *kin selection* was fundamental to the evolutionary interpretation of a variety of important social phenomena, including not only "altruism" and cooperation, but also selfishness, aggression, and spite. I will focus on altruism, though the extension to other "problems" is very significant as well. The problem is straightforward. We widely observe cooperative behavior: Bats share food; monkeys alert others to predators, even though that draws attention to themselves. Female ants forego reproduction to raise the offspring of their queen. How is this possible in a Darwinian world in which self-interest should be the rule? Hamilton's idea was as simple as it is elegant and developed in exquisite mathematical detail, which will however not be presented here. He essentially proposed a conservative extension of Darwin's insight that variations in fitness, together with inheritance, tend to promote adaptations. In his models, fitness will tend to be optimized over the long run, though we only deal with shorter runs. Darwin originally focused on the individual and the consequences of variations in morphology or behavior for the individual's survival and reproduction. Hamilton extended fitness to include the effects on kin and factored them in organismic fitness. He recognized that it is not *only* the *direct* impact on the individual's fitness that matters for natural selection, but that the *indirect* effects on related individuals also contribute to one's fitness. My overall fitness does not depend exclusively on me, but on my brothers as well, as R. A. Fisher recognized early in the century. Hamilton took this one step further. My fitness does not depend only on me and my offspring but also on the fitness of my brother's offspring. *Inclusive fitness* is just the sum of these various components, suitably weighted by their degree of relatedness to the agent. If b is the benefit a recipient garners from another's behavior, and c is the cost of the behavior to the actor, then Hamilton's rule says that an altruistic behavior with cost c can be favored, provided rb is greater than c, with r being the relatedness coefficient. Thus, I should be willing to consider sacrificing one of my children for two of my brother's. Of course, the benefit may be distributed across a number of individuals (e.g., when an individual emits a warning call), so we need a more generalized rule, which is exactly what Hamilton provided (cf. Hamilton 1964).

When Hamilton originally formulated his account of kin selection, one key problem for evolutionary biologists was the presence of eusociality among the hymenoptera (ants, bees, and wasps). The problem is how sterile castes could possibly evolve when foregoing reproduction should not be favored. We need to explain eusociality in Darwinian terms. The key insight from Hamilton is that asymmetries related to haplodiploid mechanisms of sex determination can facilitate eusociality. Workers are more closely related to their sisters than to their offspring and so kin selection can, under some conditions, make it more advantageous to raise sisters than offspring. Eusociality has arisen numerous times in the hymenoptera as well as in a number of other lineages not characterized by haplodiploidy. This result was seen as a spectacular success for Hamilton's inclusive fitness (cf. Trivers and Hare 1976).

Kin selection and inclusive fitness have become staples for theories in human sociobiology and EP as well as for animal social behavior. They are also staples for evolutionary biology more generally. It is worth noting that, quite recently, evolutionary biologists have raised a number of difficulties in applying kin selection models (cf. Nowak et al. 2010; van Veelen 2010; Gadagkar 2010; Doebeli 2010). The debates arising out of this have been heated (cf. Abbot et al. 2011; Boomsma et al. 2011; Strassman et al. 2011; Ferriere and Michod 2011; Herre and Wcislo 2011; Nowak et al. 2011). These complications do not feature in EP's use of the models, though they are important to contemporary evolutionary biology and should feature in future work by evolutionary psychologists. I do not intend to enter this discussion here and will assume that inclusive fitness theory is on a solid footing (cf. Bourke 2011; for a useful discussion). I have deployed this strategy before (Richardson 2007). Since the use of inclusive fitness assumes the adequacy of the framework, assuming the theory is on a solid footing does not bias my ultimate case. I will suggest that their deployment of the theory is empirically inadequate. I am actually sympathetic with the use of inclusive fitness. But if it should turn out that the critics are right and inclusive fitness cannot even account for the iconic cases, this will hinder rather than help EP.

Let us now turn to a core case, which is a relatively recent application of Hamilton's model that has gained some prominence and that illustrates the use of inclusive fitness within EP. The use of inclusive fitness resembles the application exhibited in Wilson's *On Human Nature* (1978), though with slight variations and amendments. Prohibitions against incest are pervasive features in human cultures. Often, they are cited as cultural universals. Incest is a very interesting case for a social prohibition, since psychological studies show that its disapproval survives even the recognition that it will produce no actual harm. It has been a very significant case for both sociobiologists and evolutionary psychologists. There is a straightforward case against incest from an evolutionary perspective based on the biological costs of inbreeding. It is difficult to assess in detail the implications of inbreeding from a strictly biological perspective, but it is clear that inbreeding does often result in reduced fitness, referred to as *inbreeding depression*. There are different explanations of inbreeding depression. Inbreeding depression is sometimes explained by overdominance, with heterozygote superiority, and sometimes by partial dominance, each with the idea that inbred lines can become fixed for recessive and deleterious alleles. Assuming overdominance, the effects of deleterious alleles are entirely masked; with partial dominance, they are reduced. The former is one important contender and is the one EP appeals to, though either might be mobilized. Assuming recessive alleles are often deleterious, inbreeding would be more likely to couple these deleterious recessives with the consequence of offspring with reduced fitness. Since nearly everyone has some recessives that are likely to be deleterious, incest should lead to reduced fitness among one's offspring. This would be an evolutionary pressure against inbreeding, *which might be called* Fisher's influence: A reduction in my offspring's fitness is a reduction in my own fitness. This much has been recognized, though the exact mechanisms have been disputed for the last century.

The importance of inbreeding depression leads EP advocates to suggest that there is a natural tendency—sometimes a psychological "module"—for incest avoidance. Lieberman, together with Cosmides and Tooby (2003), suggested that humans have a specialized kin recognition system (there are such mechanisms in other mammals) and tries to bring this into line with Hamilton's model for kin selection. They also observe that it could facilitate an avoidance of any deleterious consequences associated with inbreeding depression. The connection to kin selection is indirect. The explanation offered is roughly the following. Aversion to incest, as expressed in incest taboos, encompasses not only oneself but also a negative assessment of third party behavior. This is important, as incest aversion is oriented toward one's own partners, but does not explain or underwrite the more general aversion to incestuous behavior among our kin. The things our relatives do that have a direct impact on their fitness also has an indirect impact on our fitness, given the theory of inclusive fitness. Thus, if incest comes with direct deleterious effects for my relatives, it carries indirect, but nonetheless real, deleterious effects for me. Lieberman et al. (2003) therefore suggest that we should expect a tendency to interfere with sexual unions among close relatives. Since EP assumes our tribe to be composed mostly of relatives, our aversion to incest should extend to them as well. We should have an aversion of incest not only for ourselves but also for our kin. This noticeably neglects the fact that, assuming inbreeding depression, allowing sexual unions among other relatives could actually improve extended fitness. Reducing the fitness of competitors, even relatives, indirectly improves our own since all that matters is *relative* fitness. Further, improving the fitness of some kin, in the presence of other kin, reduces the fitness of those kin and, indirectly, ours. Insofar as this is true, the most one can get is an aversion for ones' own incestuous relations, but not for a competitor's incestuous relations.

The Westermark hypothesis posits a psychological mechanism of the sort Lieberman et al. (2003) predict, suggesting that young children who are raised together develop a sexual disinterest, or even a sexual aversion, to each other. The proper function of this "natural" aversion is to avoid incest, since those who are raised together are most often closely related, or at least would have been in ancestral groups. Lieberman reasonably assumes that co-residence during periods of high parental investment should be a reliable indicator of kinship or at least would have been a reliable indicator in the EEA with small kinship bands. Together with Cosmides and Tooby, Lieberman shows considerable support for the conclusion that duration of co-residence is *psychologically* predictive of sexual avoidance. I do not know of any evidence that this extends to aversion to others' sexual encounters. The evolutionary interpretation above generalizes the aversion to the third person.

Even though this seems to be an attractive view, the evolutionary interpretation is nonetheless problematic. The association cannot be directly tested in ancestral populations, but that is hardly a significant objection. It does fit the patterns of some contemporary "hunter–gatherer" populations, all of which have incest taboos, though with varying scope and severity. It is, of course, true that siblings would typically be associated with one another during childhood; but the proper question is whether those one is typically associated with, once reproductively active, are likely to be siblings, or would have been resident in ancestral groups. This depends on the specific form of social organization that was present in these ancestral groups. We know that our Pleistocene ancestors did not simply have one lifestyle in one region, but lived on the African savannah, in deserts, next to rivers, by oceans, in forests, and even (somewhat later) in the Arctic, employing very different foraging methods, living off diverse diets, with technologies ranging from the simple chopping tools of *Homo habilis* to the rich and sophisticated stone, bone, and antler toolbox of late Pleistocene *Homo sapiens*. There is little reason to think that there was a single form of social structure associated with the full range of human physical environments, or that contemporary "hunter–gatherer" populations exhibit a social structure typical of ancestral groups. It is hard to know whether associations will be limited to siblings or more likely to be with siblings, without a fairly specific account of social organization, including the relative viscosity of the groups and issues such as group size. The application of inclusive fitness theory to the human case depends on knowing what we do not know, assuming without warrant. It is not just a matter of skepticism. It rests on widely accepted and embraced principles.

This concern with the EP model for the evolution of incest avoidance fits with the widely recognized thought among evolutionary biologists that kin selection is more effective within relatively "viscous" populations, i.e., populations that do not involve significant immigration and emigration. This has two underlying theoretical connections. On the one hand, kin selection is more likely to favor altruistic behavior in the presence of kin and social viscosity will tend to increase the number of kin. In this case, one specific concern regards immigration and emigration patterns. In many animal species, there is a tendency for animals to disperse prior to mating, to move away from their familial unit. This clearly has the effect of reducing inbreeding, although there does not seem to be any consensus on whether incest avoidance is particularly significant in supporting dispersal. For example, when male chimpanzees reach reproductive age, they tend to emigrate from the ancestral clan. There is therefore no need for incest aversion, since they move away from their siblings. In order to know how to apply the kin models to ancestral human groups, we would need to know, e.g., whether males and females both remain with the ancestral groups or emigrate. There is some evidence that among early *Homo sapiens,* the males tended to move out of their ancestral groups once they were reproductive, as is the case with chimpanzees. It does not matter whether this is correct. The important point is that without this information, the relevance of the evolutionary models of kin selection to incest aversion is not clear. There are two connections. First, emigration and immigration patterns directly affect the mean relatedness in the group. Mean relatedness impacts directly on the r values involved in inclusive fitness: As relatedness goes down, the corresponding benefit must go up in order to satisfy Hamilton's rule. Likewise, as genetic relatedness goes up, the effect of kin selection may be diluted. Much of this depends on population structure. Here we are in the dark. In addition, emigration and immigration rates are crucial for determining *effective population size*. This is also a key parameter in models of kin selection, rather than simply the size of the local kin group. With relatively low immigration rates, the effective population size increases dramatically. Again we are in the dark.

5.5 Evolutionary Alternatives

The general assumptions above, which form the background for EP, are characteristic for much of the research done in evolutionary biology. However, there is reason to believe that the assumptions are problematic. Recent developments in genetics, evolutionary biology, and in developmental biology suggest that rethinking these assumptions might improve EP considerably, bringing it more in line with more recent evolutionary thinking. Several points can be briefly noted (cf. Bolhuis et al. 2011; Laland and Brown 2011):

- Natural selection and sexual selection are doubtless potent evolutionary forces. There are alterative evolutionary factors that can, and do, affect evolutionary trajectories. EP acknowledges such factors as genetic drift (though it plays no role in their scenarios), but phylogenetic factors play no significant role, such that our primate kin do not typically feature in EP explanations. One salutary change would be to systematically take account of our debt to our primate kin. This would, in particular, downplay the commitment to natural selection acting on specifically human social capacities.
- EP typically assumes that the relevant selection forces are relatively ancient and that recent changes are insignificant. From the perspective of EP, we are as much Pleistocene relics as the condor or the grizzly bear. We do know, however, that there have been substantial changes in the human genome over even the last 10,000 years. Many of these are connected with the adoption of agriculture and the domestication of animals. More generally, evolutionary biologists have found that the rate of evolution can be much faster than EP tends to assume.
- The environment of the Pleistocene is known to have been highly variable, both over time and over space. The environment of the early Pleistocene was very different from, say, the upper Paleolithic. Moreover, humans came to be widely dispersed, occupying a variety of distinctive environments. Given what we know, it would also be reasonable to think that social structures are different in different physical environments; e.g., some would be more conducive to sedentary life styles and others to more mobile ones. Though humans are not as genetically diverse as many other animals, there is sufficient genetic variation to support genetic changes in relatively short amounts of time.
- Human behavior is both adaptive and malleable. When there is variation, a typical assumption of EP is that the strategies are conditional strategies, evoked in different conditions, or that the variants are simply abnormal. More recent developmental biology emphasizing such things as epigenetic inheritance, niche construction, and developmental plasticity, make the assumption of a universal form problematic. It is not that EP assumes some form of genetic determinism; rather, the point is that the kind of interplay we see among genetic factors, epigenetic influences, and learning, makes universals less likely.
- There are significant alternatives to the typical emphasis of EP on individual- and gene-centered models of evolution. This is an issue beyond the problems of applying their preferred modes of analysis. Gene-cultural co-evolution may be an important source of evolutionary changes. This is coming to be a well-developed alternative, emphasizing the role of cultural practices in modifying the human brain. In general, gene-cultural dynamics can enhance and accelerate rates of evolution. Multi-level selection models are also being developed. With distinctive groups, genetically isolated, competing against each other, it is possible to develop models for the evolution of social behavior that do not assume the typically individual- and gene-oriented perspective of EP.

There are alternatives which could enrich the work within EP, but which typically remain beyond its purview. Darwin was right to think that evolution should reshape our understanding of human psychology. There are many avenues to explore in seeing how it might be enriched.

References

Abbot, P., Abe, J., Alcock, J., Alizon, S., Alpedrinha, J. A., Andersson, M., et al. (2011). Inclusive fitness theory and eusociality. *Nature, 471*(7339), E1–E4.

Barkow, J., Cosmides, L., & Tooby, J. (Eds.). (1992). *The adapted mind: Evolutionary psychology and the generation of culture*. New York: Oxford University Press.

Bolhuis, J. J., Brown, G. R., Richardson, R. C., & Laland, K. N. (2011). Darwin in mind: New opportunities for evolutionary psychology. *PLoS Biology, 9*(7), e1001109. doi:10.1371/journal.pbio.1001109.

Boomsma, J. J., Beekman, M., Cornwallis, C. K., Griffin, A. S., Holman, L., Hughes, W. O., et al. (2011). Only full-sibling families evolved eusociality. *Nature, 471*(7339), E4–E5.

Buller, D. J. (2005). *Adapting minds: Evolutionary psychology and the persistent quest for human nature*. Cambridge: MIT Press.

Buss, D. M. (1995). Evolutionary psychology: A new paradigm for psychological science. *Psychological Inquiry, 6*(1), 1–30.

Buss, D. M. (Ed.). (2005). *The handbook of evolutionary psychology*. New York: Wiley.

Cosmides, L., & Tooby, L. (1992). Cognitive adaptations for social exchange. In J. H. Barkow, L. Cosmides, & J. Tooby (Eds.), *The adapted mind. evolutionary psychology and the generation of culture* (pp. 163–228). New York: Oxford University Press.

Darwin, C. (1859). *On the Origin of Species*. Cambridge: Harvard University Press.

Darwin, C. (1871). The descent of man and selection in relation to sex. In J. T. Bonner & R. M. May (Eds.). Princeton: Princeton University Press.

Doebeli, M. (2010). Inclusive fitness is just bookkeeping. *Nature, 467*(7316), 661.

Ferriere, R., & Michod, R. E. (2011). Inclusive fitness in evolution. *Nature, 471*(7339), E6–E8.

Gadagkar, R. (2010). Sociobiology in turmoil again. *Current Science (Bangalore), 99*(8), 1036–1041.

Gigerenzer, G., & Hug, K. (1992). Domain specific reasoning: Social contracts, cheating and perspective change. *Cognition, 43*(2), 127–177.

Herre, E. A., & Wcislo, W. T. (2011). In defence of inclusive fitness theory. *Nature, 471*(7339), E8–E9.

Laland, K. N., & Brown, G. R. (2011). *Sense and nonsense: Evolutionary perspectives on human behaviour* (2nd ed.). Oxford: Oxford University Press.

Lieberman, D., Tooby, J., & Cosmides, L. (2003). Does morality have a biological basis? An empirical test of the factors governing moral sentiments relating to incest. *Proceedings of the Royal Society B: Biological Sciences, 270*(1517), 819–826.

Nowak, M. A., Tarnita, C. E., & Wilson, E. O. (2010). The evolution of eusociality. *Nature, 466*(7310), 1057–1062.

Nowak, M. A., Tarnita, E. E., & Wilson, E. O. (2011). Nowak et al. reply. *Nature, 471*(7339), E9–E10.

Richardson, R. C. (2007). *Evolutionary psychology as maladapted psychology*. Cambridge: MIT Press.

Strassmann, J. E., Page, R. E., Robinson, G. E., & Seeley, T. D. (2011). Kin selection and eusociality. *Nature, 471*(7339), E5–E6.

Trivers, R. L. (1974). Parent-offspring conflict. *American Zoologist, 14*(1), 249–264.

Trivers, R. L., & Hare, H. (1976). Haplodiploidy and the evolution of the social insects. *Science, 191*(4224), 249–263.

Van Veelen, M., Garcıá, J., Sabelis, M. W., & Egas, M. (2010). Call for a return to rigour in models. *Nature, 467*(7316), 661.

Wilson, E. O. (1975). *Sociobiology: The new synthesis*. Cambridge: Harvard University Press.

Wilson, E. O. (1978). *On human nature*. Cambridge: Harvard University Press.

Chapter 6
Like Me: A Homophily-Based Account of Human Culture

Daniel Haun and Harriet Over

6.1 Introduction

In many ways, the stability of human cross-cultural variation is surprising, since high rates of migration (Hill et al. 2011) and visitation (Chapais 2008) should, over time, reduce differentiation across groups (Yeaman et al. 2011). Assuming a long enough period, any difference between human groups should inevitably fade by means of these processes (Boyd and Richerson 2005, 2009; Henrich and Boyd 1998).

Independent of inter-group migration, there is another parallel migration into any group at any given point in time: newborns. Every new generation of children confronts the group with a number of individuals that do not act according to the group-specific behavioural repertoire (Harris 2012). Thus, in addition to immigrants entering the community with conflicting behaviours and norms, there is also a constant influx of individuals entering the community either without any established behavioural patterns or sometimes even with predispositions that are counter to the local cultural variant of a particular behaviour (e.g. Haun et al. 2006).

How then do children acquire the appropriate group-specific beliefs and behaviours? Previous accounts of cultural transmission have emphasized the role of learning mechanisms such as high-fidelity imitation (Lyons et al. 2007; Whiten et al. 2009) or cognitive abilities such as perspective taking (Tomasello 1999) and sensitivity to ostensive cues (Gergely and Csibra 2006). In contrast to these accounts,

---

This chapter was originally published in Richerson, Peter J., and Morten H. Christiansen, eds., *Cultural Evolution: Society, Technology, Language, and Religion,* chapter: “Like Me: A Homophily-Based Account of Human Culture”, © 2013 Massachusetts Institute of Technology and Frankfurt Institute of Advanced Studies, by permission of The MIT Press.

D. Haun (✉)
University of Jena, Jena, Germany
e-mail: haun@eva.mpg.de

H. Over
University of York, York, UK
e-mail: harriet_over@eva.mpg.de

T. Breyer (ed.), *Epistemological Dimensions of Evolutionary Psychology,*
DOI 10.1007/978-1-4939-1387-9_6

we emphasise the importance of more social processes, in particular, homophily (a preference for others we perceive as similar to ourselves). The homophily account is based on two closely related claims. First, children preferentially affiliate with and learn from similar others. Second, and more importantly, the homophilic preferences of the group in general create a feedback loop that ensures that children engage in high-fidelity copying of the group's behavioural repertoire. This allows them to reap the benefits of others' homophilic preferences and to therefore maintain their position within the group. This homophily-based account thus unites research on the social functions of imitation (e.g. Carpenter and Call 2009; Over and Carpenter 2012; Nadel 2002; Nielsen 2009; Užgiris 1981) with that on group membership (e.g. Dunham et al. 2011; Kinzler et al. 2007; Turner 1991) and normative behaviour (Kallgren et al. 2000; Rakoczy et al. 2008).

We do not claim that the homophily account provides an exhaustive description of how social motivations influence cultural transmission. Other social motivations and preferences (for example, for prestigious others and competent others) and the interactions between them are also important in explaining humans' social learning (Laland 2004). We simply wish to highlight that the preference for similar others is one key factor in explaining cultural transmission and that species-specific differences in this tendency might be one factor in explaining the origins of species-typical features of human cultural transmission.

Below, we outline our homophilic account in more detail. We begin by discussing the importance of homophilic assortment from an evolutionary perspective. After that, we review the available evidence that, from the early stages of development, humans have had a strong preference for similar others. We then move on to the main thrust of our argument and present evidence that this preference for similar others has transformed a number of pre-existing cognitive mechanisms, such as emulation learning and majority-biased transmission, into a suite of human-unique traits that includes social imitation, conformity, and a norm-psychology.

6.2 Homophilic Social Preferences from an Evolutionary Perspective

For cooperation to be maintained within a group, it is essential for group members to be able to distinguish cooperators from defectors. In stable, personalized groups, familiarity serves to reduce aggression and to create a tolerant context; the foundation of any cooperative exchange. As groups increase in size, so does the frequency with which individuals have to interact with less and less familiar others. Eventually, the personal interaction history can no longer be relied on to estimate the reliability of a partner.

At some point during human evolution, social networks increased to a size where group members were more and more likely to encounter others that were only vaguely familiar. For instance, even the most mobile extant forager groups live in networks that typically exceed several hundred individuals (Hill et al. 2011; Apicella et al. 2012). Under such conditions, familiarity itself remains important, but is no longer as effective as before. Thus, a proxy measure for familiarity is required that reliably correlates with familiarity. Similarity in aspects of the phenotype (morphology and behaviour) is one such measure. Individuals who grow up within the same community are likely to be similar on a number of dimensions, thus making phenotypic similarity an honest signal of group membership.

We argue that a preference for similar others allowed humans to assort between strangers, identifying in-group members who were not personally known to them. Choosing to interact and cooperate with more similar strangers maximised the chance of successful cooperative interactions because similar individuals were more likely to share relevant behavioural tendencies (McElreath et al. 2003; Cohen 2012). In consequence, humans were able to function within qualitatively different forms of social organisation compared to other primates, thus tapping into the cooperative potential of strangers. Formal models have shown that such a pattern of cultural transmission, in which individuals are disproportionally influenced by those who are similar to themselves, is adaptive, since a homophilic preference causes sub-populations to become culturally isolated. This in turn allows the mean value of locally adaptive traits to converge to the optimum. A transmission strategy based, for example, on success would only adapt much more slowly to a variable habitat (Boyd and Richerson 1987). In other words, "the preference to interact with people with markers like one's own may be favored by natural selection under plausible conditions" (McElreath et al. 2003, p. 123).

In the next section, we shift our focus to empirical evidence supporting this hypothesis. We centre our discussion on the developmental and comparative data demonstrating that the human preference for similar others is much stronger than that seen in other primate species.

6.3 Like Me? Homophilic Social Preferences from a Comparative Perspective

6.3.1 Homophilic Social Preferences in Non-human Primates

Interpersonal relations in chimpanzee groups are characterised by tolerance of in-group members and hostility towards out-group members (Wrangham 1999; Wilson et al. 2012). Members of other groups detected within the home range are typically killed (with the one exception being migrating females; Kahlenberg et al. 2008). This preference for in-group members over out-group members is almost certainly based on familiarity rather than similarity, as chimpanzees typically encounter all the members of their own group on a fairly regular basis.

A recent study however raises the possibility that some non-human primates also use similarity as a means to assort between others. Paukner et al. (2009) reported that capuchin monkeys who were presented with two human experimenters, one who imitated them and another who just performed monkey-like movements, sat closer to the imitator and exchanged more tokens with him. Hence, a transient increase in behavioural similarity (social mimicry) made capuchins prefer one human to the other.

There are thus some hints that non-human primates utilise similarity in their social judgments (at least to some extent) and, in consequence, that the common ancestor of humans and other primates had rudimentary preferences for similar others. This may have provided the evolutionary starting point from which homophilic social preferences in humans could emerge. As we will see below, however, the evidence for homophilic preferences in humans far exceeds that of any other primate.

6.3.2 Homophilic Social Preferences in Children

In contrast to non-human primates, the evidence that humans assort unfamiliar others based on similarity is quite substantial (e.g. Gruenfeld and Tiedens 2010; Jones et al. 2004; Tajfel et al. 1971). This preference for similar others appears to structure social interactions from early development. For example, 6-month-olds prefer to look at individuals who speak their own versus a different language, and 10-month-olds prefer to accept toys from speakers of their own language (Kinzler et al. 2007). This preference for native language speakers structures social interactions also later in development: 5-year-olds preferentially choose native language speakers over foreign-language speakers or foreign-accented speakers as friends (Kinzler et al. 2009). However, in all the above-mentioned studies with children it is not possible to separate a preference for similar others from a preference for individuals children find easier to understand.

Fawcett and Markson (2010) have provided evidence that young children's social preferences are, at least at times, based on self-similarity alone. Fawcett and Markson demonstrated that 3-year-old children prefer to play with a puppet who expresses the same food preference as them as opposed to a contrasting preference, and a puppet whose physical appearance matches rather than mismatches their own. Other evidence comes from research on the effects of being imitated. One of the consequences of being imitated is a momentarily increased level of perceived similarity between social partners (Chartrand and Bargh 1999). From early in development, children appear to prefer individuals who imitate them to individuals who engage in independent behaviour. For example, 14-month-old infants look towards and smile more at an experimenter who imitates them than at an experimenter who engages in equally contingent but non-imitative behaviour (Agnetta and Rochat 2004; Asendorpf et al. 1996; Meltzoff 1990). Furthermore, infants and toddlers are more likely to help an experimenter who has imitated them than an experimenter who has engaged in contingent but non-imitative behaviour (Rekers and Haun submitted).

Further evidence for children's preference for similar others comes from the so-called minimal group paradigm (Tajfel et al. 1971). In this paradigm, individuals are randomly allocated to one of several groups that are only identified by an abstract, seemingly uninformative symbol. In this way, similarity between members of a minimal group is not indicative of any shared behavioural characteristic of the individuals composing the group, but only of shared group identity. Five-year-old children prefer individuals allocated to the same minimal group as them over individuals allocated to a different minimal group. Furthermore, children not only prefer individuals belonging to the same minimal group but also have more positive expectations about in-group members' behaviour (Dunham et al. 2011).

This preference for similar others seems to occur across cultures (Kinzler et al. 2012; Cohen and Haun 2013). Children's relative reliance on particular cues, however, varies depending on the particular socio-cultural context. Recent studies comparing children in different townships along the Brazilian Amazon have demonstrated that children's preferences for certain cues are likely tuned according to locally relevant cue variation. For example, children from accent heterogeneous populations rely more strongly on accent as a similarity cue than children from accent homogeneous populations (Cohen and Haun 2013).

6.3.3 Children Prefer to Learn from Similar Others

Children's preference for similar others not only indirectly channels their own input by creating interaction bubbles of similar others but also has more immediate implications for children's social learning. Kinzler et al. (2011) demonstrated that 5-year-old children are more likely to learn the function of a novel object from an individual who speaks with the child's native accent than from an individual who speaks the same language with a foreign accent. A more recent study claimed that even infants preferentially learn from similar others (Buttelmann et al. 2012). In this study, 14-month-old infants listened to a story told either in their native language or in a foreign language. Children subsequently imitated the actions of the speaker of their native language more closely. Caution must be taken when interpreting this result, however, as the design confounds similarity with other factors such as the relative comprehensibility of the stories. Nevertheless, it raises the possibility that children select their models by similarity already in the second year of life.

Other evidence that children preferentially learn from similar others comes from research on the effects of being imitated. In a recent study, Over et al. (2013) found that 5- to 6-year-old children were more likely to adopt the preferences and novel object labels of an experimenter who had previously imitated their choices than those of an experimenter who had previously made independent decisions.

6.3.4 Summary

It appears from the above that the human preference for similar others likely far exceeds that of any other primate. This preference is present early in development and structures children's learning as well as their social interactions. In the next section, we outline the consequences of this homophilic social preference for the mechanisms underlying cultural transmission.

6.4 Like Me! The Consequences of Homophilic Preferences

If we prefer similar to dissimilar others, it follows that increasing the similarity between self and other can be a useful strategy for directing others' positive social activities towards the self. We contend that homophilic preferences in humans have interacted with the social learning mechanisms inherited from our common ancestor with the other great apes and transformed them into species-unique forms of copying behaviour which serve to maintain individuals' position within the group.

In contrast to previous accounts (e.g. Carpenter and Call 2009; Užgiris 1981), the homophilic account does not require children to have the goal of making themselves more similar to their social partners. Although children may, at times, actively seek to be like others (Carpenter 2006; Over and Carpenter 2013), the more typical pattern for children may be to learn through experience that imitation is successful in improving social relations, without any explicit awareness of this connection. In consequence, their only goal within the social situation may be to get along well with others.

In the remainder of this chapter, we discuss evidence that social learning mechanisms we share with other species—emulation and majority-biased transmission—have been transformed by homophilic preferences into a suite of human-unique social learning processes including social imitation, conformity, and a norm-psychology.

6.4.1 Emulation Becomes Imitation

Chimpanzees use a range of social learning strategies, including most prominently emulation (Call et al. 2005; Nagell et al. 1993). In emulation learning, an animal focuses on the outcome that is achieved in the physical world rather than the particular actions that were used to achieve it (Tomasello et al. 1993). If chimpanzees copy the particular actions of their conspecifics (i.e. imitate), they appear to do so infrequently and with relatively low fidelity (Tennie et al. 2009).

Children, in contrast, show a strong tendency to copy actions faithfully. In fact, children's imitation is sometimes so precise that they even copy actions that are superfluous or disadvantageous to solving the task at hand (Horner and Whiten 2005; Nagell et al. 1993; Nielsen 2006). For example, children from 3 to 5 years of age, who have been trained to identify the causally irrelevant parts of novel action sequences, still reproduce causally irrelevant actions, and they continue to do so even when specifically instructed by the experimenter to copy only necessary actions (Lyons 2009; Lyons et al. 2007). This phenomenon has come to be called overimitation (Lyons et al. 2007; Lyons et al. 2011). It emerges in the 2nd year of life (Nielsen 2006) and becomes increasingly pervasive through the preschool period (McGuigan and Whiten 2009; McGuigan et al. 2007).

The homophilic account presumes that these differences in social learning between chimpanzees and humans have been driven, at least in part, by human homophilic preferences. The increased importance of "how something is done" is owed to the significance of behavioural similarity amongst individuals of a group. Finding a different way to achieve the same ends is no longer functionally equivalent to exactly copying others' actions, since the former decreases similarity with others while the latter increases it. For humans, imitation could thus serve new social purposes. This added social dimension effectively turned emulation learning into faithful imitation.

In consequence, it is misleading to refer to high-fidelity imitation as "overimitation", since the term implies that children copy unnecessary parts of action sequences. Under the homophily account, these parts, while being causally irrelevant, still serve an important function for the learner: They produce a high level of similarity between the demonstrator and the learner.

Evidence in favour of the proposal that high-fidelity imitation is used to achieve social goals comes from data suggesting that children increase their tendency to imitate when affiliation is important to them. Over and Carpenter (2009) demonstrated that 5-year-old children who have been given a goal to affiliate (through priming with social exclusion) imitate the actions of a model more closely than children who have been given a neutral prime. Further evidence in favour of this hypothesis comes from work demonstrating that children are more likely to copy the specific actions of a model when that model is in the room and thus able to watch their imitation (Nielsen and Blank 2011). Imitation is also closely associated with social factors in younger children. Nielsen et al. (2008) demonstrated that 2-year-old children are more likely to copy the specific actions of a model who engages in a contingent social interaction with them than those of a model whose behaviour is not contingent on their own.

The tendency to make the self similar to others can also be used more strategically within social settings. That is to say, imitation can serve Machiavellian ends (Over and Carpenter 2012). Research with older children has shown that they are able to use imitation in order to increase their influence over others. For example, Thelen and colleagues demonstrated that 10-year-old children are more likely to copy the specific actions of a peer when they will later need to persuade that peer to do something (Thelen et al. 1980).

6.4.2 Majority-Biased Transmission Becomes Conformity

Homophilic preferences, we argue, have influenced not only how we interact with individual social partners but also how we respond to the group in general. One way in which humans interact with the group as a whole is through consideration of the majority.

If, due to any combination of underlying mechanisms, an individual is more likely to acquire the behaviour displayed by the majority, we refer to it as a majority-biased transmission (Haun et al. 2012). A recent study with chimpanzees showed that naïve individuals copy the behaviour of the majority instead of alternatives, even if those are equally frequent, equally familiar, and equally productive behaviours (Haun et al. 2012).

Thus, chimpanzees follow the majority when they have no prior information available. However, they do not follow a majority if they have to forgo their own behavioural tendencies in order to do so (Haun et al. submitted). We refer to the tendency to forgo personal preferences in favour of copying the majority as conformity (Haun et al. 2013; van Leeuwen and Haun 2013). The authors of a recent study argue that chimpanzees conformed against their own preference, based on the finding that individuals retained their socially acquired strategy even though the alternative yielded more preferred rewards (Hopper et al. 2011). However, because individuals only very rarely experienced the alternative strategy to yield more desirable foods, it remains highly questionable if individuals were, in fact, fully aware of the alternative.

Similar to chimpanzees, human children follow the majority if they have no relevant information available (majority-biased transmission, Haun et al. 2012). However, in contrast to other primates, human children also adjust their behaviour to the majority even when they have an equally effective but individually acquired strategy available already: In a situation in which one child, with high level of performance on a certain task, is confronted with a majority of peers who unanimously give a false response, children often choose to abandon their own judgement and to adjust their behaviour to the majority's response (Berenda 1950; Corriveau and Harris 2010; Haun and Tomasello 2011; Walker and Andrade 1996). Furthermore, children appear to consider the social consequences of conforming versus dissenting. Haun and Tomasello (2011) varied the privacy of the subjects while giving their response and found lower rates of conformity when preschool children were allowed to keep their response private from the majority. Most strikingly, children adjusted their level of conformity from trial to trial, depending on the privacy of their response, conforming more often when they gave their response in public. The authors concluded that the reduction in conformity in the private condition demonstrated a partial contribution of social motivations for children's conformity on the public trials. Hence, children, in contrast to other primates, are additionally guided by social motivations when conforming to a majority (Haun and Tomasello 2011).

In the absence of a social function, copying the majority when acquiring a new skill is adaptive on an individual level, but there is no reason to follow the majority in cases where the learner already has a different but equally productive strategy available to them. However, if conformity also serves a social function, then it pays for a learner to forgo their own strategy and adopt that of the majority, since sticking to the former will decrease similarity between the self and the group, while conforming to the latter will increase similarity between the self and the group. According to the homophily account, this added social dimension increased humans' tendency to conform to the majority, effectively turning majority-biased transmission into conformity.

6.5 The Emergence of Norm-Psychology

Non-human primates, such as chimpanzees, have "rules of conduct" that are reinforced. For example, subordinates tend to display certain gestures when meeting a dominant individual and violations of this behavioural pattern will result in aggression (Goodall 1986). Although the superficial structure of these patterns of behaviour might resemble that of human norms, they differ from norms in important respects (Tomasello 2008). For example, while human norms are often variable across groups, gestures negotiating the relationship between dominant and subordinate individuals in chimpanzees are highly similar across different, unrelated populations, thus suggesting that they are not culturally learned (Tomasello et al. 1997). Furthermore, chimpanzee "rules", unlike human norms, are not agent neutral. Subordinate chimpanzees failing to submit to the dominant might suffer aggression from the dominant (the affected party), but not from other (unaffected) group members. Chimpanzees do not appear to punish the violations of third parties (Riedl et al. 2012). Humans, on the other hand, punish the transgressions of others even if they do not concern them directly (Henrich et al. 2006). Hence, it does not seem to be the case that chimpanzees collectively intend to do things in a certain way and do not have any normative expectations about their conspecifics' behaviour, but that their social interactions are better characterized by behavioural regularities and individual and idiosyncratic preferences for certain behaviours.

Human norms are rich in their social interpretation: Norms describe the "right" way to do things, the way things "ought" to be done, the way "we" do things (Bruner 1993). Human children appear to spontaneously detect such norms in many behaviours even in the absence of normative language (Schmidt et al. 2011). After a single confident and intentional demonstration by an adult, children appear to assume that the way in which an action was demonstrated is normative. Following such a demonstration, children will not only follow that norm but also actively enforce it when later observing someone performing the action "incorrectly", often protesting using normative language about what people ought to do (Rakoczy et al. 2008). Thus, children readily enforce norms on others even if their violation does not impact upon them directly.

We have argued above that the social relevance of similarity amongst individuals gives previously socially neutral behaviours a new social relevance. The "way something can be done" is effectively elevated to the "way we do something", fitting actions with a social signalling function. This normative dimension to actions that have no intrinsic value (e.g. how to hold a fork) is, we argue, a direct consequence of the relevance of self–other similarity in cooperative groups of increasing size.

6.6 Conclusion

Many accounts exist for the species-unique structure of human social behaviour. All of them contain lists of human-specific social abilities and motivations for coordination (Tomasello et al. 2005), social learning (Tennie et al. 2009), teaching (Gergely and Csibra 2006), and norm-psychology (Chudek andHenrich 2011). We have provided an account that unites some of these previously unconnected sets of abilities and motivations. According to our homophily-based account, a preference for similar over dissimilar others underlies important aspects of human-unique social behaviour.

Evidence suggests that, from early development, children prefer to interact with, and learn from, individuals who are similar to themselves. This preference for similar others and the potential advantages reaped by being similar to others, ensures that children engage in high-fidelity copying of the group's behavioural repertoire. In consequence, seemingly irrelevant parts of actions gain social relevance by serving as a similarity marker. This tendency to interpret the physically irrelevant structure of actions as social signals, we argue, spawned a human-unique form of interpreting the actions of others: norm-psychology.

In summary, we argue that species-specific difference in homophilic preferences might be one key factor in explaining the origins of species-typical features of human cultural transmission. We predict that humans are unique amongst living primates in the extent of their preference for similar others. We also predict that this preference is universal across human cultures, albeit relying on different similarity cues in different populations (Logan and Schmittou 1998; Cohen and Haun 2013). Future studies should further test these predictions from cross-cultural and comparative angles.

Acknowledgments We would like to thank Malinda Carpenter, Emma Cohen, Emma Flynn, Katja Liebal, Nadja Richter, Carel van Schaik, Claudio Tennie, Marco Schmidt, Peter Richerson, and an anonymous reviewer for valuable comments on an earlier draft.

References

Agnetta, B., & Rochat, P. (2004). Imitative games by 9-, 14-, and 18-month-old infants. *Infancy, 6*(1), 1–36.

Apicella, C. L., Marlowe, F. W., Fowler, J. H., & Christakis, N. A. (2012). Social networks and cooperation in hunter-gatherers. *Nature, 481*(7382), 497–501.

Asendorpf, J. B., Warkentin, V., & Baudonnière, P.-M. (1996). Self-awareness and other-awareness II: Mirror self-recognition, social contingency awareness, and synchronic imitation. *Developmental Psychology, 32*(2), 313–321.

Berenda, R. W. (1950). *The influence of the group on the judgments of children: An experimental investigation*. New York: King's Crown Press.

Boyd, R., & Richerson, P. J. (1987). The Evolution of Ethnic Markers. *Cultural Anthropology, 2,* 65–79.

Boyd, R., & Richerson, P. J. (2005). *The origin and evolution of cultures*. Oxford: Oxford University Press.

Boyd, R., & Richerson, P. J. (2009). Culture and the evolution of human cooperation. *Philosophical Transactions of the Royal Society B: Biological Sciences, 364*(1533), 3281–3288.

Bruner, J. (1993). Commentary on Tomasello et al. 'cultural learning'. *Behavioural and Brain Sciences, 16,* 515–516.

Buttelmann, D., Zmyj, N., Daum, M. M., & Carpenter, M. (2012). Selective imitation of in-group over out-group members in 14-month-olds. *Child Development, 84*(2), 422–428. doi:10.1111/j.1467-8624.2012.01860.x.

Call, J., Carpenter, M., & Tomasello, M. (2005). Focusing on outcomes and focusing on actionsin the process of social learning: Chimpanzees and human children. *Animal Cognition, 8,* 151–163.

Carpenter, M. (2006). Instrumental, social, and shared goals and intentions in imitation. Imitation and the development of the social mind: Lessons from typical development and Autism. In S. J. Rogers & J. H. G. Williams (Eds.), *Imitation and the social mind* (pp. 48–70). New York: Guilford.

Carpenter, M., & Call, J. (2009). Comparing the imitative skills of children and nonhuman apes. Revue de Primatologie, 1 (Document 6), http://primatologie.revues.org/263.

Chapais, B. (2008). *Primeval kinship: How pair-bonding gave birth to human society* (*illustrated ed.*). Cambridge: Harvard University Press.

Chartrand, T. L., & Bargh, J. A. (1999). The chameleon effect: The perception-behaviour link and social interaction. *Journal of Personality and Social Psychology, 76*(6), 893–910.

Chudek, M., & Henrich, J. (2011). Culture-gene coevolution, norm-psychology and the emergence of human prosociality. *Trends in Cognitive Sciences, 15*(5), 218–226.

Cohen, E. (2012). The evolution of tag-based cooperation in humans: The case for accent. *Current Anthropology, 53*(5), 588–616.

Cohen, E., & Haun, D. B. M. (2013). The development of tag-based cooperation via a socially acquired trait. *Evolution and Human Behavior, 34,* 230–235.

Corriveau, K. H., & Harris, P. L. (2010). Preschoolers (sometimes) defer to the majority in making simple perceptual judgments. *Developmental Psychology, 46*(2), 437–445.

Dunham, Y., Baron, A. S., & Carey, S. (2011). Consequences of "minimal" group affiliations in children. *Child Development, 82*(3), 793–811.

Fawcett, C. A., & Markson, L. (2010). Similarity predicts liking in 3-year-old children. *Journal of Experimental Child Psychology, 105*(4), 345–358.

Gergely, G., & Csibra, G. (2006). Sylvia's recipe: The role of imitation and pedagogy in the transmission of cultural knowledge. In N. J. Enfield & S. C. Levinson (Eds.), *Roots of human sociality: Culture, cognition, and human interaction* (pp. 229–255). Oxford: Berg.

Goodall, J. (1986). *The chimpanzees of gombe: Patterns of behaviour*. Cambridge: Harvard University Press.

Gruenfeld, D. H., & Tiedens, L. Z. (2010). Organizational preferences and their consequences. In S. T. Fiske, D. T. Gilbert, & G. Lindsay (Eds.), *The Handbook of social psychology*. New York: Wiley.

Harris, P. L. (2012). *Trusting what you're told: How children learn from others*. Cambridge: Belknap.

Haun, D. B. M., & Tomasello, M. (2011). Conformity to peer pressure in preschool children. *Child Development, 82*(6), 1759–1767.

Haun, D. B., Rapold, C. J., Call, J., Janzen, G., & Levinson, S. C. (2006). Cognitive cladistics and cultural override in hominid spatial cognition. *Proceedings of the National Academy of Sciences of the United States of America, 103*(46), 17568–17573.

Haun, D. B. M., Rekers, Y., & Tomasello, M. (2012). Majority-biased transmission in chimpanzees and human children, but not orangutans. *Current Biology, 22*(8), 727–731.

Haun, D. B. M., Van Leeuwen, E. J. C., & Edelson, M. G. (2013). Majority influence in children and other animals. *Developmental Cognitive Neuroscience, 3,* 61–71.

Haun, D. B. M., Rekers, Y., & Tomasello, M. (submitted). Great apes stick with what they know; children conform to others. Psychological Science.

Henrich, J., & Boyd, R. (1998). The evolution of conformist transmission and the emergence of between-group differences. *Evolution and Human Behaviour, 19*(4), 215–241.

Henrich, J., McElreath, R., Barr, A., Ensminger, J., Barrett, C., Bolyanatz, A., & Ziker, J. (2006). Costly punishment across human societies. *Science, 312*(5781), 1767–1770.

Hill, K. R., Walker, R. S., Božičević, M., Eder, J., Headland, T., Hewlett, B., & Wood, B. (2011). Co-Residence patterns in hunter-gatherer societies show unique human social structure. *Science, 331*(6022), 1286–1289.

Hopper, L. M., Schapiro, S. J., Lambeth, S. P., & Brosnan, S. F. (2011). Chimpanzees' socially maintained food preferences indicate both conservatism and conformity. *Animal Behaviour, 81*(6), 1195–1202.

Horner, V., & Whiten, A. (2005). Causal knowledge and imitation/emulation switching in chimpanzees (pan troglodytes) and children (homo sapiens). *Animal Cognition, 8,* 164–181.

Jones, J. T., Pelham, B. W., Carvallo, M., & Mirenberg, M. C. (2004). How do I love thee? Let me count the js: Implicit egotism and interpersonal attraction. *Journal of Personality and Social Psychology, 87*(5), 665–683.

Kahlenberg, S. M., Thompson, M. E., Muller, M. N., & Wrangham, R. W. (2008). Immigration costs for female chimpanzees and male protection as an immigrant counterstrategy to intrasexual aggression. *Animal Behaviour, 76*(5), 1497–1509.

Kallgren, C. A., Reno, R. R., & Cialdini, R. B. (2000). A focus theory of normative conduct: When norms do and do not affect behaviour. *Personality and Social Psychology Bulletin, 26,* 1002–1012.

Kinzler, K. D., Dupoux, E., & Spelke, E. S. (2007). The native language of social cognition. *Proceedings of the National Academy of Sciences of the United States of America, 104*(30), 12577–12580.

Kinzler, K. D., Shutts, K., Dejesus, J., & Spelke, E. S. (2009). Accent trumps race in guiding children's social preferences. *Social Cognition, 27*(4), 623–634.

Kinzler, K. D., Corriveau, K. H., & Harris, P. L. (2011). Children's selective trust in native-accented speakers. *Developmental Science, 14*(1), 106–111.

Kinzler, K. D., Shutts, K., & Spelke, E. S. (2012). Language-based social preferences among children in South Africa. *Language Learning and Development, 8*(3), 215–232.

Laland, K. (2004). Social learning strategies. *Learning and Behaviour, 32*(1), 4–14.

Logan, M. H., & Schmittou, D. A. (1998). The uniqueness of crow art: A glimpse into the history of an embattled people. *Montana: The Magazine of Western History, 48*(2), 58–71.

Lyons, D. E. (2009). The rational continuum of human imitation. In J. A. Pineda (Ed.), *Handbook of environmental engineering: Mirror neuron systems* (pp. 77–103). New York: Humana.

Lyons, D. E., Young, A. G., & Keil, F. C. (2007). The hidden structure of overimitation. *Proceedings of the National Academy of Sciences, 104*(50), 19751–19756.

Lyons, D. E., Damrosch, D. H., Lin, J. K., Macris, D. M., & Keil, F. C. (2011). The scope and limits of overimitation in the transmission of artifact culture. *Philosophical Transactions of the Royal Society, B, 366,* 1158–1167.

McElreath, R., Boyd, R., & Richerson, P. J. (2003). Shared norms and the evolution of ethnic markers. *Current Anthropology, 44*(1), 122–130.

McGuigan, N., & Whiten, A. (2009). Emulation and "overemulation" in the social learning of causally opaque versus causally transparent tool use by 23- and 30-month-olds. *Journal of Experimental Child Psychology, 104*(4), 367–381.

McGuigan, N., Whiten, A., Flynn, E., & Horner, V. (2007). Imitation of causally opaque versus causally transparent tool use by 3- and 5-year-old children. *Cognitive Development, 22*(3), 353–364.

Meltzoff, A. N. (1990). Foundations for developing a concept of self: The role of imitation in relating self to other and the value of social mirroring, social modeling, and self practice in infancy. In D. Cicchetti & M. Beeghly (Eds.), *The self in transition: Infancy to childhood* (pp. 139–164). Chicago: University of Chicago Press.

Nadel, J. (2002). Imitation and imitation recognition: Functional use in preverbal infants and nonverbal children with autism. In A. Meltzoff & W. Prinz (Eds.), *The imitative mind: Development, evolution, and brain bases* (pp. 42–62). Cambridge: Cambridge University Press.

Nagell, K., Olguin, R. S., & Tomasello, M. (1993). Processes of social learning in the tool use of chimpanzees (pan troglodytes) and human children (homo sapiens). *Journal of Comparative Psychology, 107*(2), 174–186.

Nielsen, M. (2006). Copying actions and copying outcomes: Social learning through the second year. *Developmental Psychology, 42*(3), 555–565.

Nielsen, M. (2009). The imitative behaviour of children and chimpanzees: A window on the transmission of cultural traditions. *Revue De Primatologie*, *1*(document 5), http://primatologie.revues.org/254; doi:10.4000/primatologie.254.

Nielsen, M., & Blank, C. (2011). Imitation in young children: When who gets copied is more important than what gets copied. *Developmental Psychology, 47*(4), 1050–1053.

Nielsen, M., Simcock, G., & Jenkins, L. (2008). The effect of social engagement on 24-month-olds' imitation from live and televised models. *Developmental Science, 11*(5), 722–731.

Over, H., & Carpenter, M. (2009). Priming third-party ostracism increases affiliative imitation in children. *Developmental Science, 12,* F1–F8.

Over, H., & Carpenter, M. (2012). Putting the social into social learning: Explaining both selectivity and fidelity in children's copying behaviour. *Journal of Comparative Psychology, 126*(2), 182–192.

Over, H., & Carpenter, M. (2013). The social side of imitation. *Child Development Perspectives, 7*(1), 6–11.

Over, H., Carpenter, M., Spears, R., & Gattis, M. (2013). Children selectively trust individuals who have imitated them. *Social Development,* 22(2), 215–224.

Paukner, A., Suomi, S. J., Visalberghi, E., & Ferrari, P. F. (2009). Capuchin monkeys display affiliation toward humans who imitate them. *Science, 325*(5942), 880–883.

Rakoczy, H., Warneken, F., & Tomasello, M. (2008). The sources of normativity: Young children's awareness of the normative structure of games. *Developmental Psychology, 44,* 875–881.

Rekers, Y., Cronin, K. A., & Haun, D. B. M. (submitted). Behavioural co-variation increases prosocial behaviour in 2-year-old children, and other great apes.

Riedl, K., Jensen, K., Call, J., & Tomasello, M. (2012). No third-party punishment in chimpanzees. *Proceedings of the National Academy of Sciences, 109,* 14824–14829.

Schmidt, M. F., Rakoczy, H., & Tomasello, M. (2011). Young children attribute normativity to novel actions without pedagogy or normative language. *Developmental Science, 14*(3), 530–539.

Tajfel, H., Billig, M. G., Bundy, R. P., & Flament, C. (1971). Social categorization and intergroup behaviour. *European Journal of Social Psychology, 1*(2), 149–178.

Tennie, C., Call, J., & Tomasello, M. (2009). Ratcheting up the ratchet: On the evolution of cumulative culture. *Philosophical Transactions of the Royal Society B: Biological Sciences, 364*(1528), 2405–2415.

Thelen, M. H., Miller, D. J., Fehrenbach, P. A., Frautschi, N. M., & Fishbein, M. D. (1980). Imitation during play as a means of social influence. *Child Development, 51,* 918–920.

Tomasello, M. (1999). *The cultural origins of human cognition*. Cambridge: Harvard University Press.

Tomasello, M. (2008). *Origins of human communication*. Cambridge: MIT Press.

Tomasello, M., Kruger, A. C., & Ratner, H. H. (1993). Cultural learning. Behavioural and Brain. *Sciences, 16*(03), 495–511.

Tomasello, M., Call, J., Warren, J., Frost, G. T., Carpenter, M., & Nagell, K. (1997). The ontogeny of chimpanzee gestural signals: A comparison across groups and generations. *Evolution of Communication, 1*(2), 223–259.

Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences, 28,* 675–735.

Turner, J. C. (1991). *Social influence*. Buckingham: Open University Press.

Užgiris, I. C. (1981). Two functions of imitation during infancy. *International Journal of Behavioural Development, 4*(1), 1–12.

van Leeuwen, E. J. C., & Haun, D. B. M. (2013). Conformity in nonhuman primates: Fad or fact? *Evolution and Human Behaviour, 34,* 1–7.

Walker, M. B., & Andrade, M. G. (1996). Conformity in the Asch task as a function of age. *The Journal of Social Psychology, 136*(3), 367–372.

Whiten, A., McGuigan, N., Marshall-Pescini, S., & Hopper, L. M. (2009). Emulation, imitation, over-imitation and the scope of culture for child and chimpanzee. *Philosophical Transactions of the Royal Society B: Biological Sciences, 364*(1528), 2417–2428.

Wilson, M. L., Kahlenberg, S. M., Wells, M., & Wrangham, R. W. (2012). Ecological and social factors affect the occurrence and outcomes of intergroup encounters in chimpanzees. *Animal Behaviour, 83*(1), 277–291.

Wrangham, R. W. (1999). Evolution of coalitionary killing. *American Journal of Physical Anthropology, 110*(suppl. 29), 1–30.

Yeaman, S., Bshary, R., & Lehmann, L. (2011). The effect of innovation and sex-specific migration on neutral cultural differentiation. *Animal Behaviour, 82*(1), 101–112.

Chapter 7
Evolutionary Media Psychology and Its Epistemological Foundation

Christine Hennighausen and Frank Schwab

7.1 The Human's Insatiable Interest in the Media

Why do a myriad of people go to the movies and watch million-dollar blockbusters such as *Titanic* or *Lord of the Rings* over and over again? Why does the reader of a novel feel happiness and satisfaction when his favorite character finally ends up marrying his formerly secret love? And why do most teenagers spend countless hours visiting social network sites (SNS) such as Facebook? To cut a long story short, why do humans universally produce and enjoy media, and devote such a considerable amount of time for it? At all times, *Homo sapiens sapiens* enjoyed telling and listening to stories, watching art and entertainment, and, more recently, reading and being absorbed by exciting books or movies (Schwab and Schwender 2010). Nowadays, in our so-called information age, the media have become a central part of our daily life. Current research shows that individuals spend an average of 4 h a day watching TV, 3 h a day listening to the radio, and almost 1.5 h a day browsing the Internet (ARD-ZDF-Onlinestudie 2012). Media psychology aims at describing and explaining behavior and experiences concerning the usage of mass and individual media (Winterhoff-Spurk 2004). Mass media include the press, radio, TV, and cinema, whereas individual media encompass landline telephones, mobile devices, and social network services. In addition to learning and knowledge acquisition, the main focus of media psychology is to find an answer to the question why humans are willing to invest such a great amount of time and money into (media) entertainment and art. One step towards a satisfactory answer is to analyze the phenomena of media usage by applying an evolutionary approach.

C. Hennighausen (✉) · F. Schwab
Department of Media Psychology, University of Würzburg, Würzburg, Germany
e-mail: christine.hennighausen@uni-wuerzburg.de

F. Schwab
e-mail: frank.schwab@uni-wuerzburg.de

T. Breyer (ed.), *Epistemological Dimensions of Evolutionary Psychology*,
DOI 10.1007/978-1-4939-1387-9_7

7.2 "Old" Darwin and the "New" Media: The Missing Link

For a long time, evolutionary theories and their application to everyday phenomena of human life has been a neglected field of psychology. In the twentieth century, psychological predictions and hypotheses were predominantly derived from neo-behaviorism, social learning theory, cognitive approaches, modern psychoanalysis, and postmodern radical constructivist approaches (Schwab 2010a). Cosmides and Tooby (1994) even state that "cognitive psychology has been conducted as if Darwin never lived" (p. 43). Traditional anthropology, sociology, communication sciences, and cognitive psychology primarily focus on current and ontogenetic environmental factors as well as on social milieu factors when analyzing human experience and behavior (Schwab 2010b; Sherry 2004). The human body is implicitly considered as a product of nature, whereas the human mind is regarded as a sole product of culture or environment. Even though nobody would deny the fact that our brain and central aspects of our mental functioning are a result of natural selection, it rarely leads to any consequences in the development of theories and the deduction of hypotheses. But this dichotomous and exclusive view of nature and culture as opposites proves to be too narrow-minded. Human behavior is always a result of an interaction between *evolved psychological mechanisms (EPMs)* and environmental factors, as all evolved human mental programs need environmental input to be activated (Buss 1995; Confer et al. 2010). In cognitive psychology, the human mind is generally compared with computer hardware on which selected and specific cultural software is run (Hagen 2005). The evolved human nature is almost completely ignored when psychologists and media researchers as "software experts" mainly apply cognitive "learning-only approaches" to almost all phenomena they are interested in. This ignorance of human nature may result in some kind of one-eyedness or "nature blindness" (Sherry 2004). If, nevertheless, researchers apply a biological approach, foremost biological mechanisms such as genetics, neurophysiology, and peripheral physiology are considered. Researchers predominantly ask how these biological mechanisms work and influence human perception and behavior proximately. From an evolutionary perspective, however, the important ultimate question, namely why exactly these mechanisms have evolved and which purpose they initially served, remains disregarded.

Contrary to traditional cognitive psychology approaches, evolutionary psychologists view the human mind as an *adaptive toolbox* (Gigerenzer and Selten 2001; Schwab and Schwender 2010). They suppose that the human mind rather consists of several domain-specific EPMs than of only a few general mental mechanisms (Barkow et al. 1992; Buss 1999; Schwab 2010a). It is assumed that the specific EPMs have evolved in our early ancestors, who were faced with multigenerational adaptive problems occurring in a long-lasting and stable—mainly Pleistocene—environment (Barkow et al. 1992; Cosmides and Tooby

1992; Schwab 2006; Schwab and Schwender 2010). The successful solution of these multigenerational adaptive problems led to the development and the genetic transfer of specific EPMs, since individuals equipped with these had a reproduction advantage over individuals lacking these EPMs. The modern human, concerning both his nature and behavior, can thus be regarded as a result of an evolutionary process of adaptation (Buss 1995; Confer et al. 2010; Schwab and Schwender 2010). The fact that the modern human is still equipped with the ancestral EPMs, but lives in a world radically differing from ancient Pleistocene, is referred to as the *mismatch of evolutionary design* (Workman and Reader 2008). Does the mismatch of evolutionary design also appear in the human's usage of the media? Some media researchers suppose so. According to Schwender (2006), the effects of media events and figures are similar to those of a scarecrow: A scarecrow is designed to deter other birds by mimicking a predatory bird through its shape and design. The scarecrow deceives the other birds' perception, triggers specific EPMs (that are usually triggered by the sight of a predatory bird), and frightens them. Similarly, the media trigger the recipient's emotions by activating EPMs through visual and aural stimuli or cues. Fictional media characters and events evoke happy or sad feelings in the recipient; recipients even form long-term parasocial relationships with fictional characters (Horton and Wohl 1956; Schramm 2008; Schwab and Schwender 2010). The evolutionary psychologist Pinker (1997) also explains human enjoyment of art and entertainment as a by-product of evolution and interprets it in terms of the mismatch of evolutionary design. He describes the entertaining and aesthetic aspects of media reception as "the futile tickling of the human pleasure centers" (Pinker 1997, p. 526). For a better illustration, Pinker introduced the so-called *cheesecake metaphor*: Humans love and enjoy eating cheesecake, although they have not evolved a special taste for it. Pinker gives the following explanation: in the course of evolution, circuits have evolved in the human brain that react to the sweet taste of ripe fruit, fat, and the oil of nuts or meat. These stimuli, indicating highly nutritious food, elicit feelings of pleasure. As a cheesecake provides an overdose of highly nutritious food, combining fat, oil, and ripe fruit, it especially tickles these pleasure centers (Pinker 1997). Pinker assumes that the same mechanisms work when humans use and enjoy the media.

So far, it has not been definitely determined for which purpose the media and artful entertainment evolved and which specific EPMs they trigger in the recipient. Evolutionary media psychologists aim at answering this and other questions and transfer the framework of evolutionary psychology to the research on modern media. They hold the view that the human's use of the media is—if not exclusively, then at least essentially—determined by the domain-specific EPMs that have evolved during Pleistocene to solve adaptive problems. Evolutionary media psychology focuses on mental features involved in media usage that are designed by natural and sexual selection. In addition, the ultimate and proximate (biological) functions of media production and use, such as motivation, reception, and effects, are analyzed (Schwab 2010a).

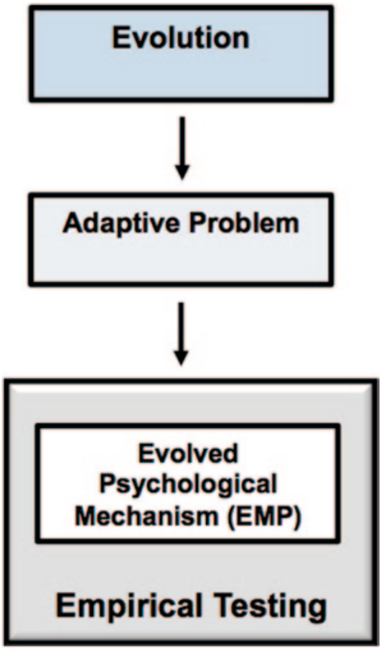


Fig. 7.1 Empirical testing in evolutionary psychology. (Based on Holcomb 1998)

7.3 Generating and Testing Hypotheses in Evolutionary Psychology

The principle aim of evolutionary media psychologists is to analyze the specific elements of the human mind's adaptive toolbox and their ultimate functions in modern human's use of the media. But how do evolutionary psychologists proceed when generating and testing theories and hypotheses?

In general, methods in evolutionary psychology are derived from evolutionary biology as well as from experimental and cognitive psychology. In his article "Testing Evolutionary Hypotheses," Holcomb (1998) describes the common scientific procedure in evolutionary psychology, which consists of three different steps (see Fig. 7.1):

First, a significant physical or intellectual adaptive problem our ancestors faced in Pleistocene is identified. Second, hypotheses about a specific EPM that developed through selection processes in order to solve this ancestral adaptive problem are formulated. When generating specific hypotheses, presumed features and conditions of the ancestral Pleistocene environment are taken into account. It is then considered how our ancestors could have solved the encountered adaptive problems appropriately. Finally, the derived hypotheses and theories are empirically tested. It is examined if the postulated EPMs still exist as, for example, species-wide psychological bias or apparent irrational behavior. Moreover, it is important to distinguish between current effects and past psychological adaptations (cf. Holcomb 1998).

Evolutionary psychology hypotheses are tested by applying psychological methods and by collecting empirical data. The empirical testing is complemented by archeological data, data of hunters and gatherers, observational data, self-report data, life course data, public statistics, and human cultural artifacts, such as cave paintings or media content (Buss 1995; Confer et al. 2010; Schwab 2010a). Evolutionary psychologists explicitly do not aim at testing the underlying phylogenetic processes and Darwin's theory of evolution directly (Ketelaar and Ellis 2000). They rather test the EPMs through specific hypotheses previously derived from Darwin's theory of evolution and modern evolutionary theories (Buss 1995). Derived assumptions and hypotheses can either compete or conflict. Thus, assumptions about human phylogeny are used to generate hypotheses, but they are by no means explanatory luxury. Moreover, evolutionary explanations are on no account intended as ex post facto speculations (see Schwab 2010a; Schwab and Schwender 2010).

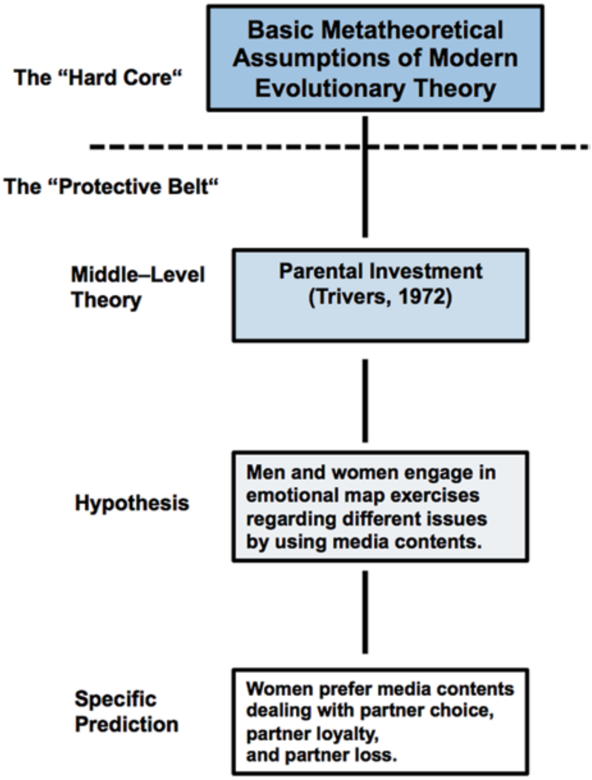
7.4 The Philosophy of Science in Evolutionary Psychology: Popper Versus Lakatos

Although evolutionary psychology is considered to be a promising new research program in psychology (e.g., Barkow et al. 1992; Buss 1995, 1999; Confer et al. 2010), it has likewise been criticized (e.g., Buller 2005; Fodor 2005; Lewontin 1990, 1998; Schlinger 2002; Silvers 2010; Trafimow and Gambacorta 2012). As evolutionary psychologists usually tested theories and derived hypotheses rather than the underlying basic framework, in this case Darwin's theory of evolution, they are recurrently faced with the reproach that specific evolutionary models and their hypotheses are hence unfalsifiable and just-so stories as well as post hoc explanations of observed phenomena (Confer et al. 2010; Holcomb 1998). It is further claimed that evolutionary psychologists neglect to apply standards of evidence referring to established principles of epistemology when evaluating their derivative models and hypotheses (Ketelaar and Ellis 2000). Ketelaar and Ellis (2000) respond to the critics in their target article "Are Evolutionary Explanations Unfalsifiable? Evolutionary Psychology and the Lakatosian Philosophy of Science." Starting by applying the Popperian philosophy of science, followed by the Lakatosian philosophy of science, they discuss in detail the scientific foundations of evolutionary psychology.

According to Popperian epistemology, science can be distinguished from pseudoscience by employing the method of falsification (Popper 1959). During the scientific progress, theories are developed and deduced hypotheses are empirically tested. The derivate hypotheses can be either preliminarily verified, i.e., supported by collected data, or falsified, i.e., rejected in case of inconsistencies with collected data. If hypotheses drawn from the theory have been proven false, this falsification also falsifies the underlying theory. Ketelaar and Ellis (2000) suppose that exclusively applying the method of falsification to the science of psychology would lead to "too many empirical findings being cast in terms of their support or refutation of binary oppositions, such as nature versus nurture, central versus peripheral, serial versus parallel" (Newell 1973, as cited in Ketelaar and Ellis (2000), p. 3). They further state that using the rather discriminative Popperian epistemology might primarily result in the accumulation of negative knowledge about the human mind and that positive knowledge about the mechanisms of the human mind would only be derived from hypotheses that were not falsified. Therefore, a Popperian scientific approach does not provide an efficient strategy to test psychological theories that generate hypotheses. Consequently, the authors call for a more approximating scientific perspective to complement Popper's method of falsification and propose the Lakatosian philosophy of science (cf. Ketelaar and Ellis 2000). Lakatos (1970, 1978) supports a more approximating scientific approach. First, a set of basic assumptions (e.g., Newton's four principles in physics) is agreed upon. This is also referred to as the "hard core." This set of basic assumptions is then used to generate a broader metatheoretical research program. A metatheory provides a research heuristic and can be compared with a "scientific map": it guides scientists in the development of theories and hypotheses, protects them from predictive failures, gives established criteria for the observation and recognition of significant natural phenomena, and prompts them to question generated explanations (cf. Ketelaar and Ellis 2000). Metatheories allow the locating and linking of new theories and hypotheses within an existing metatheoretical framework and, hence, provide a tool to systematically arrange and accumulate new knowledge. Furthermore, a metatheory facilitates the generation of new theories and hypotheses, as it is much easier to distinguish between plausible and implausible a priori psychological hypotheses. Competitive metatheories are evaluated as being either progressive or degenerative, depending on the falsification or verification of their derivative hypotheses. During the scientific progress, theoretical frameworks are constantly refined and reformulated and can thus be considered rather as approximations to observed phenomena. From a Lakatosian perspective, theories are always evaluated in relation to rivaling theories in a specific scientific area. Despite certain scientific failures, researchers may keep the theory that best explains the observed phenomena (cf. Ketelaar and Ellis 2000).

The core metatheory is surrounded by mid-level and derivative auxiliary hypotheses that constitute its "protective belt." This "protective belt" serves different functions. First, it allows scientists to link the core metatheory to empirically collected data. By verifying mid-level theories and auxiliary hypotheses within the "protective belt," the assumptions of the underlying metatheory are indirectly empirically verified, too. Second, the "protective belt" protects the core assumptions from being refuted, as it is the mid-level theories that are directly empirically tested. If data is gathered that is inconsistent with the mid-level theory or a derivative hypothesis, it does not mean that the entire underlying metatheory is equally proven to be wrong. In this case, the falsified mid-level theory can be replaced by an alternative theory based on the metatheoretical assumptions. The mid-level theories are consistent with the metatheory, but may not be consistent with each other. Consequently, when sufficient empirical data are collected during the scientific progress, the mid-level theory explaining the examined phenomena best rules competitive mid-level theories out and finally displaces them.

Fig. 7.2 The hierarchical structure of evolutionary psychological explanations based on the example of evolved sex preferences in media contents. (Schwab 2010a, b; based on Buss 1995 and Ketelaar and Ellis 2000)



In evolutionary psychology, the Lakatosian levels of analysis can be further differentiated into a total of four: the core metatheory, mid-level theories, hypotheses, and specific predictions (cf. Buss 1995; Ketelaar and Ellis 2000; see also Figs. 7.2, 7.4, and 7.6). At the top of the Lakatosian model, the metatheoretical "hard core" consists of Darwin's theory of adaption through natural and sexual selection and the general principles of genetic evolution derived from modern evolutionary theories (e.g., Dawkins 1976, 1982; Hamilton 1964; Tooby and Cosmides 1992). On the next lower level, the evolutionary mid-level theories are located, connecting the basic metatheoretical assumptions with derivative hypotheses and specific predictions. By testing the derivative hypotheses and predictions, mid-level theories are either confirmed or contradicted. On the next level of scientific analysis, the derived hypotheses are situated. Competitive hypotheses can be inferred either from one single mid-level theory or from different mid-level theories. Within evolutionary psychology, rivaling mid-level hypotheses can be derived either from an evolutionary framework (e.g., the evolution of human art can be described as an *evolutionary by-product* or as an *evolutionary adaptation*; for a more detailed discussion, see below; Miller 2000; Pinker 1997), or from non-evolutionary theoretical frameworks (e.g., preferences in human mate selection can be explained based on sexual strategies theory (SST) or on social role theory; Buss and Barnes 1986;

Eagly et al. 2000). Lastly, specific predictions can be found at the lowest level of scientific analysis. As hypotheses are often too general to be tested in a specific context, specific empirically testable predictions are generated. By testing these predictions, it becomes increasingly obvious whether a mid-level theory is supported or not, since its value is measured by the total of verified and falsified predictions and hypotheses.

7.5 Examples and Evidence from Evolutionary Media Psychology

7.5.1 The Culture by Nature Perspective

Applying an evolutionary approach to the analysis of the media might raise some questions at first sight, as the media is rather regarded as a product of human culture than one of human evolution. Likewise, the term *culture* is often used to describe ideas, representations, beliefs, behavior patterns, and artifacts, including art and the media (cf. Buss 2001). The media, however, can be also viewed as a result of human evolution, because it is the cognitive architecture of the human mind, shaped by evolutionary selection processes, that underlies any human behavior leading to cultural phenomena (Miller 1998; Tooby and Cosmides 1992).

Evolutionary processes are manifested in the individual's phenotype. The concept of the *classic phenotype* comprises, in general, the sum of all (externally) detectable features of an individual such as its morphology, ontogenetic development, biochemical, or physiological properties (Dawkins 1982, 1999). An individual's phenotype is a mixed result of the expression of an organism's genes and of the influence the environment exerts on the individual. It is further the product of the genotype–environment interaction that takes place during the individual's ontogenetic development. Typical examples of the classic phenotype provide physical characteristics, that is, the color of an organism's eyes, its hair, its height, or its strength. As addendum to the classic phenotype, the ethologist Dawkins (1982, 1999) introduced the concept of the *extended phenotype.* According to Dawkins, the concept of the phenotype is not restricted to biological processes (e.g., protein biosynthesis) only, but can also be extended to all effects a gene has on the organism it is located in, on its environment, or even on other organisms (e.g., a cuckoo chick manipulates its host to feed it; Dawkins 1982, 1999). He argues that the effects of the genes are mediated by the individual's behavior, which, in turn, results from the individual's genes, specific alleles, and the gene–environment interaction. Niche construction can be described as a special case of the extended phenotype: the ability to show a certain behavior, that is, constructing a beaver dam, is passed on genetically and aims at enhancing the organism's differential reproductive success by altering the environment to the organism's advantage (Dawkins 2004).

7.5.2 Media Entertainment: An Evolutionary Adaptation or a By-Product?

Two competing theories exist with regard to the evolution of entertainment: The first describes media entertainment as a by-product of evolution (e.g., Pinker 1997; Reeves and Nass 1996) and the second as an evolutionary adaptation (e.g., Dissanayake 1992; Miller 2000; Schwab 2008; Schwab and Schwender 2010; Tooby and Cosmides 2001).

Evolutionary by-products are not a product of direct selection mechanisms, but occur as a side effect of an adaptation. In humans, mechanisms to learn a language have evolved as an adaptation, while the ability to write is most likely to be seen as a by-product (Fitch et al. 2005). As mentioned above, Pinker (1997) introduced the cheesecake metaphor to explain media production and reception in humans. With this metaphor, he compares the effect media have on the human brain with the effect the consumption of a cheesecake has on the human brain. Pinker argues that specific adaptive EPMs that have once evolved within an ancestral stable Pleistocene environment to process and react to certain important stimuli are also triggered by (artificial) objects providing the same or related important stimuli. As certain circuits of the human brain strongly respond with pleasant feelings to the taste of a sweet and fatty cheesecake (because this reaction of the human brain to the consumption of nutritious and high-energy food proved to be adaptive in our ancestors' world), there are also specific EPMs of the human brain that respond automatically and with great pleasure to media contents providing specific stimuli.

Media Equation The *media equation approach* and the *ethopoeia assumption* (Reeves and Nass 1996), respectively, are both examples of theories that postulate an evolutionary by-product effect occurring in human–media interaction. The authors describe the observable phenomenon that, given certain circumstances, humans react to the media in a similar way as they would react to real social persons. Especially interactive media such as computers can provide low-level social cues and thus trigger social behavior. The evoked social behavior from humans can be attributed to the activation of specific EPMs that developed as adaptations to interact with other human individuals. Studies of Reeves and Nass (1996) demonstrated that participants apply the same rules of politeness when interacting via a computer as in human–human interactions. In an experimental setting, computers provided participants with several facts and, afterwards, asked the participants to rate the computers' performance. When the participants were asked to give the performance evaluation on the same computer that had provided the facts, they gave significantly more positive ratings than participants who completed the evaluation on a different computer. This is in line with the social behavior people show in interactions with other humans, as it is considered to be rude to give a mediocre or bad performance rating directly to the person concerned. In two additional experiments, Reeves and Nass (1996) showed that participants rated the computer's performance differently depending on whether the computer had previously provided them with positive feedback or not. Even gender stereotyping could be found in human–computer interaction: in a knowledge task, participants rated a computer talking in a female voice less competent regarding the communication of technical knowledge (a domain of expertise regarded as typically male) than a computer talking in a male voice. These findings can be interpreted in terms of an evolutionary by-product, as it proved to be adaptive for our ancestors to automatically comprehend systems displaying social cues as psychological entities, such that they could immediately show an appropriate social response and predict the other's individual reaction (Dennett 2008).

The Ornamental Mind On the other hand, there are also theories explaining media entertainment as an evolutionary adaptation. One explanation is provided by the theory of the *ornamental mind* proposed by Miller (2000), which is based on Darwin's theory of sexual selection. According to the theory of natural selection (Darwin 1859), males and females of one species are both equally exposed to natural and environmental selection pressures, as they share the same habitat. Thus, the sexes should evolve either similar, if not exactly the same, physical and psychological characteristics (e.g., body size and strength, fur and plumage, risk taking and aggressive behavior, etc.) as an adaptation to the natural and environmental selection pressures of this habitat. In most species, however, the sexes have evolved significantly different physical and psychological characteristics: in birds, males are usually equipped with a brighter plumage than females (e.g., the peacock's tail), and in deer, only the males develop powerful antlers and engage in aggressive struggles. Because natural selection theory cannot sufficiently explain the evolution of sex-typical differences, Darwin (1871) proposed a second selection mechanism: sexual selection. The theory of sexual selection distinguishes between *intrasexual* and *intersexual* selection. The former refers to the competition between males of a species to gain access to fertile females. The latter describes the phenomenon that, in most species, it is the female that selects the mating partner. Females are a scarce resource for males because they have much more restricted reproductive capacities due to a limited number of ova. In addition, females show more parental investment (Trivers 1972), as they usually carry, bear, and nurture the offspring. They are therefore at a greater risk of losses and wasted resources when choosing a mate of inferior genetic quality or one that cannot (or is not willing to) provide them with sufficient resources. Due to their choosiness, females exert a selection pressure on males. Driven by the force of sexual selection, males have evolved costly sexual ornamentation as an adaptation. By means of this often splendid sexual ornamentation, males signal the females' health, strength, resources, and, in general, "good genes," because the quality of the sexual ornamentation highly depends on the male's physical condition and genetics (e.g., the quality of the peacock's tails varies with the degree of parasitic infestation). Females select males by their sexual ornamentation and, in turn, males use their ornamentation to seduce and charm the females. The ornamental mind theory (Miller 2000) continues the idea of sexual selection and transfers it to the human mind. Accordingly, artistic and entertaining abilities have been favored through the mechanism of sexual selection during the human evolution. Just as Darwin (1871) said, namely that, "high cost, apparent uselessness, and manifest beauty usually indicated that a behavior had a hidden courtship function," the human mind is also more entertaining, intelligent, creative, and talented than necessary to survive and reproduce. In terms of entertaining and artistic abilities, Miller (2000) compares the human mind with the peacock's tail: Just like the peacock's tail indicates health and fitness to peahens, characteristics of the human mind, expressed through art, creativity, or humor, serve as fitness indicators of the individual. In summary, he describes the human mind as an "entertainment system that evolved to attract sexual partners" (Miller 2000, p. 29). Accordingly, products of the human mind, such as art, moral, language, and creativity, are considered adaptations in order to seduce and charm the opposite sex (e.g., a conversation with a potential mate provides a number of cues to his intelligence and personality). Consequently, some aspects of the modern-day production and selection of media entertainment can also be interpreted as an adaptation to attract and seduce reproductive mates.

Social Cohesion A second explanation of art as an evolutionary adaptation is provided by Dissanayake (1988, 1995, 2000), who primarily focuses on attention and its influence. She assumes that the regulation and management of attention are central to explore effects of spectacular artifacts, such as cave paintings or archaic sculptures, and to investigate creative and artistic processes. Dissanayake assumes that art has primarily arisen in order to establish and strengthen social cohesion. Art meets the criteria of an evolutionary adaptation, since it is found in every culture, consumes a lot of resources, energy, and time, triggers intensive and pleasant emotions (every evolutionary adaptive behavior "rewards" individuals by eliciting positive and pleasant feelings to reinforce this behavior), is associated with biologically important behavior (such as play), and is ontogenetically developed by the majority of individuals even without special training. Furthermore, she describes art as "making special" and as a kind of behavior that shows similarities to behavioral patterns also observable in other species. She particularly refers to play and ritual, as both encompass specific forms of movement and expression that do not serve any obvious purpose other than evoking pleasant feelings. Formalized and fixed rituals are central to animal behavior. Exaggerated, elaborated, and repeated forms of ritual behavior are foremost used to communicate with other individuals, as it is obvious in lekking or mating behavior (Schwab 2004). According to Dissanayake (2000), human art originates from the intimate pre-speech conversational exchange between parent and child. This pre-speech conversational exchange displays clear characteristics of rituals: it is finely drawn, rhythmical, coordinated between the individuals, exaggerated, elaborated, repetitive, and surprising (Dissanayake 2000; Schwab 2010a; Schwab and Schwender 2010; Stern 2002; Tomasello and Call 1997).

Functional and Organizational Mode Another approach examines entertainment in the context of training and learning. In their *evolutionary theory of aesthetics, fiction, and the arts,* Tooby and Cosmides (2001) assume that human imagination plays a crucial role for the development of human mental abilities regarding thoughts, feelings, and fantasy. In their theory, the authors predominantly focus on fictional narrations. They suppose that art, including fictional narration, serves an evolutionary adaptive function, as humans across all cultures enjoy spending time in fictional worlds and narratives. Furthermore, a number of findings indicate that specific psychological mechanisms have evolved to deal with fiction. First, fiction affects the experience of emotions, whereas there is no effect on behavior or motor activity. Second, humans are easily able to distinguish between factual and fictional information (Schreier and Appel 2002). Third, autistic individuals do not develop the ability to *pretend play,* which hints at the fact that pretend play, a pre-adaptation for dealing with fiction, is based on a specific EPM (cf. Schwab and Schwender 2010). Tooby and Cosmides (2001) propose two different modes in which cognitive adaptation can work: the *functional mode* that refers to the normal mental operation, and the *organizational mode* that describes a test or training operation which includes playing, learning, dreaming, narrative fictional entertainment, and other kinds of art. The organizational mode supports the development and the expansion of the mind and seems to be only activated in times of great security and satiation (Früh 2002; Schwab 2010a; Schwab and Schwender 2010). Similar to the theory developed by Tooby and Cosmides (2001), the so-called *Verhaltensdiversifikations-Protokognitionstheorie des Spiels* (behavior diversification protocognition theory of play) supposes that play has evolved as an evolutionary adaptation, because it contributes to the expansion and development of behavioral alternatives (Ohler 2001; Ohler and Nieding 2006). According to this theory, specific behavior patterns that go beyond established behavior accidentally evolved during evolution. Displaying these extended behavior patterns turned out to be adaptive and advantageous, as individuals having trained extra behavioral patterns in play (e.g., hunting or fighting) were able to reproduce them more effectively in crucial “real-life” situations (e.g., a struggle) than individuals lacking these behavior patterns trained in play before. As a result, play and training behavior was passed on to following generations and became established within the species’ genetic pool. Mechanisms of play especially evolved in species equipped with more complex cognitive systems (Ohler and Nieding 2006). These species are able to imagine fictional objects and entities and, thus, to go through and evaluate different alternatives in their imagination.

Entertainment as Emotional Map Exercise A related approach explains entertainment as an adaptive emotional map exercise aiming at emotional fine-tuning and training of emotional mechanisms. Higher primates are able to mentally simulate behavioral alternatives, which can be put on a level with fantasy and productive reasoning (Bischof 1985). The mental simulations can be interpreted as a protective mechanism, because they allow the individual to run though different behavioral alternatives within its imagination and to evaluate them as successful or unsuccessful before eventually displaying the behavior. Emotions play a crucial role in the evaluative process, as they “indicate” which behavioral alternative to select: positive emotions usually point at a promising and successful behavioral alternative, while negative emotions resulting from anticipated negative outcomes prevent individuals from choosing the respective alternative behavior (cf. Dennett 1997). In line with the theory of emotional map exercise, it is supposed that the themes and contents of entertaining media are mainly determined by evolutionarily relevant problems: multigenerational adaptive problems of our ancestors, such as mating, status, resources, and kin protection. To sum up, modern entertainment media provides a place to mentally simulate behavioral alternatives that might be carried out in crucial life situations, predominantly when evolutionarily relevant decisions and problems occur (Schwab 2008; Schwab 2010a).

7.6 Selected Evidence of Evolutionary Media Psychology

In the following section, selected studies analyzing and explaining human's use and production of media within an evolutionary context are presented. To demonstrate Lakatosian epistemology in evolutionary media psychology, the Lakatosian scientific process is pointed out in detail for each of the selected studies.

7.6.1 Study 1: Sex-Typical Media Preferences

As males and females of one species show different amounts of parental investment in their offspring (Trivers 1972), they are faced with specific adaptive problems and, thus, follow different reproductive strategies depending on a long-term or a short-term mating context (see Buss and Barnes 1986; Buss and Schmitt 1993). Females usually invest more resources in offspring (at least they bear the costs of gestation and lactation), but have much more limited reproductive capacities than males (a limited number of female ova vs. an almost unlimited number of male sperm). To avoid a waste of precious resources, females are choosy with whom to mate, and, as an adaptation to the problem of mate selection, they evolved a taste for sexual partners displaying specific desired characteristics serving as reliable indicators of the male's health and high genetic quality. Males, on the other hand, have developed costly sexual ornamentation (e.g., antlers in deer or bright plumage in birds) to demonstrate their genetic fitness as an adaptation to selection pressure by females (intersexual selection, Darwin 1871). Once females have made their choice and produced offspring, they are motivated to tie their mate to themselves to guarantee the provision of resources for them and their offspring. In contrast, males are less willing to commit to one female only and to invest all their resources in her and the offspring. They are rather motivated to mate with a high number of sexual partners to guarantee the most extensive gene distribution. Several studies have provided evidence for the different mating strategies of men and women and their adaptations: men more often seek extra-pair mating, express a higher desire for sexual variety, report a higher average number of sexual partners in their lifetime, are more willing to lower their standards regarding their sexual partners in a short-term mating context than females, and rate physical attractiveness and youth in their mate as one of the most important characteristics, as this indicates health and fertility (Buss 1989; Buss and Schmitt 1993).

Females seek resources, status, sexual faithfulness, specific character traits that suggest the male to be both a good father and a good mate (e.g., agreeableness, generosity, humor), and genetic fitness in a long-term mate (see SST, Buss 1989; Greiling and Buss 2000; Regan 1998). As pointed out earlier, within an evolutionary context, men and women are faced with a variety of different adaptive problems and pursue different interests and objectives. To solve these adaptive problems, individuals are supposed to engage in emotional map exercise and in mental testing of behavioral alternatives (Bischof 1985; Ohler 2001; Ohler and Nieding 2006; Schwab 2010a; Tooby and Cosmides 2001). As described above, the media and its contents provide a place for the mental simulation and emotional evaluation of behavioral alternatives concerning evolutionarily relevant adaptive problems. Given their different adaptive problems, males and females should engage in or prefer different emotional map exercises provided by different media contents. Schwab (2010a) assumed that the emotional map exercise men and women engage in explains part of the sex-typical differences found in media preferences and selection. Within Lakatosian epistemology, parental investment theory (Trivers 1972; SST; Buss and Schmitt 1993) and the theories postulating emotional map exercise (Ohler 2001; Ohler and Nieding 2006; Tooby and Cosmides 2001) constitute mid-level theories, since they are derived from the core evolutionary theories (see Fig. 7.2).

The hypothesis that the sexes use media contents to perform emotional map exercises concerning different sex-typical adaptive problems is situated at the hypothesis level of analysis. Based on the mid-level theories and the deduced hypotheses, it can be specifically predicted that woman prefer media contents dealing with issues of mate choice, partner loyalty, and the loss of a partner. It can be assumed, on the other hand, that men are more likely to select media contents dealing with kin protection, rivalry, status, power, and the acquisition and maintenance of resources. These specific predictions are directly empirically testable and located at the bottom of the Lakatosian model, i.e., at the level of the specific predictions. In his research, Schwab further used an integrative model to explain sex differences (Asendorpf 1996). According to this model, sex-typical differences in media preferences are also supposed to be the result of the interplay between sex and gender identity (see Fig. 7.3; Schwab 2010a).

Consequently, the participants' sex and gender role and identity were assessed. Gender role and identity were measured by the androgyny framework developed by Bem (1974), postulating two unipolar and independent femininity and masculinity dimensions. The participants' sex was collected on both the apparent genetic and the hormonal level. Genetic sex was assessed by physical appearance; the hormonal sex was measured by applying the method of digit ratio 2D:4D, that is, calculating the ratio of the index finger and the ring finger. This ratio serves as a reliable indica-

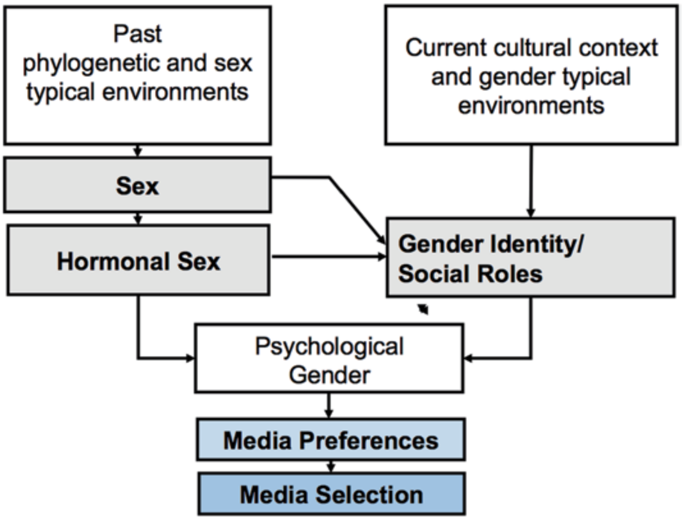


Fig. 7.3 Integrative model to explain sex differences in media preferences and selection. (Schwab 2010a)

tor of the hormonal sex, because it is affected by the prenatal concentration of sex hormones. Evidence shows that men develop a lower 2D:4D ratio resulting from a higher fetal concentration of testosterone and a lower concentration of estrogen (Manning 2002; Voracek and Loibl 2009). Studies have further revealed that prenatal and current concentration of testosterone, naturally or resulting from medication, may influence aspects of psychological gender that manifest themselves in certain behaviors, such as dominance, aggression, willingness for competition, and sexual activity (Bischof-Köhler 2002; Mealey 2000; Reinisch 1981). To investigate to which extent media preference can be attributed to genetic sex, hormonal sex, and gender identity, a sample of moviegoers was surveyed. As a dependent measure, the moviegoers' selected film genre was collected. According to the formulated predictions, the results demonstrated that women significantly prefer melodramatic and romantic movies, whereas men significantly favor action films. For other genres (thriller, comedy, horror), no significant difference between the sexes was found. Further analyses showed that genetic sex significantly accounted for variance in the preference for romance, action, and drama, and that the prenatal hormonal sex appears to explain a significant proportion in the preference for romance. Finally, gender identity also proved to explain a large part of the media content preferences, as it was particularly evident that feminine role identity increased the preference for melodrama, while masculine role identity contributed to the selection of action movies.

To sum up, Schwab's research program (2010a) shows that both culturally constructed or socialized gender identity as well as hormonal and genetic sex contribute to the sex-typical differences in the preference regarding media content.

7.6.2 Study 2: Musical Seduction: The Entertainer's Ornamental Mind

The ornamental mind theory postulates that human art and creativity have evolved as a result of sexual selection (Miller 2000): the human mind and its products (including art and the media) evolve to indicate and promote the individual's genetic fitness in order to attract and seduce potential mates. Miller (2000) compares the human mind with the sexual ornamentation found in the animal kingdom (e.g., the peacock's tail). Bringing together the concept of the extended phenotype (Dawkins 1982, 1999) and the ornamental mind theory (Miller 2000), media entertainment can be regarded as a cultural artifact designed by and for evolutionary purposes. Thus, the questions arises whether (media) entertainment also serves the purpose of charming a potential mate by indicating the individual's fitness or whether it still shows design features of this phylogeny. To answer this question, Schwab and Bordin (2009) conducted an experimental study in which female students were asked to give attractiveness ratings of potential mates either as a short-term (affair/sexual encounter) or as a long-term sexual partner (including a strong emotional relationship/marriage). The female students' attractiveness ratings were based on the mates' personal descriptions given in lonely hearts advertisements. These lonely hearts advertisement descriptions were manipulated concerning the degree to which the male indicated musical abilities (moderate vs. outstanding musician). As control variables, the socio-sexual orientation (Penke and Asendorpf 2008), the female's fertility status and the intake of contraceptives were collected. Within the Lakatosian epistemological framework, the ornamental mind theory (Miller 2000) can be classified as a mid-level theory within the "protective belt" of the general evolutionary theory (see Fig. 7.4).

Building on the ornamental mind theory, different assumptions can be generated, which are located at the next lower Lakatosian level of analysis, the level of hypotheses. Because the degree of musicality (as an expression of art) serves as an indicator of the men's genetic fitness, Schwab and Bordin (2009) hypothesized that women would rate men describing themselves as outstanding musicians as better sexual partners than men indicating only moderate musical abilities, regardless of the mating context. It was further hypothesized that women in their fertile phase would show a stronger preference for outstanding musicians than women being in a non-fertile phase of their ovulatory cycle, as they would engage in more profound information processing when they have a greater risk of getting pregnant in case of sexual intercourse. Finally, specific predictions, situated at the lowest level of scientific analysis in the Lakatosian model, were formulated and tested. They encompass, among other things, that women rate men describing themselves as being outstanding musicians as more attractive than men indicating only moderate musical abilities in both mating contexts. However, the results showed that outstanding musicians were only favored over moderate musicians in a short-term relationship context; no selection differences emerged in a long-term mating context. Moreover, during their fertile phase, women showed a significantly stronger preference for

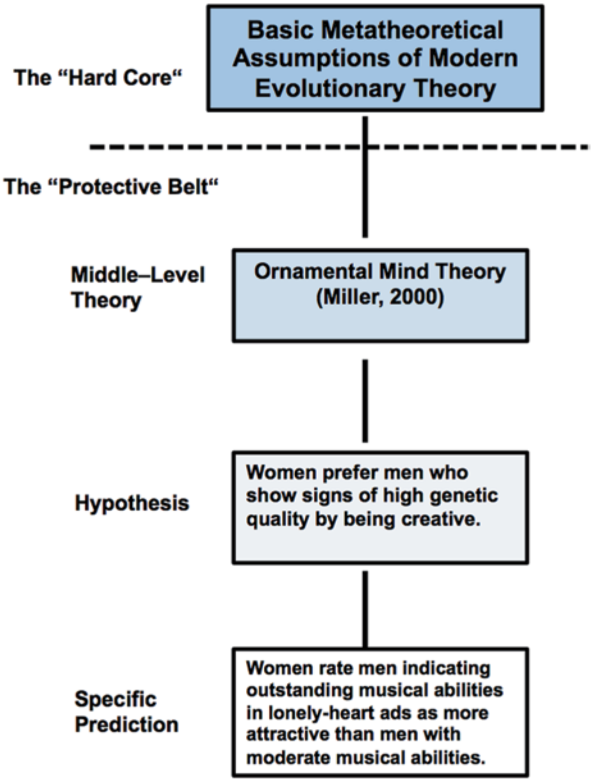


Fig. 7.4 The hierarchical structure of evolutionary psychological explanations based on the example of evolved women's preferences for men indicating high genetic quality by describing musical abilities. (Schwab and Bordin 2009; based on Buss 1995 and Ketelaar and Ellis 2000)

outstanding musicians than for moderate musicians in a short-term mating context, but not in a long-term mating context. During a non-fertile phase, no such effect could be found. To sum up, the results demonstrated that high musical abilities in men significantly attracted women only in the short-term mating condition, whereas in the long-term mating condition, no reproductive advantage for outstanding musicians could be found. Furthermore, women in the fertile phase of the ovulatory cycle expressed a stronger preference for more musical men, but only in a short-term mating context.

These findings can be interpreted in the framework of the *dual mating hypothesis* (Gangestad and Simpson 2000; Pillsworth and Haselton 2006a, b). This hypothesis states that women form long-term relationships with investing mates, whereas they seek "good genes" (i.e., genetic fitness) from short-term sexual partners (extra-pair mates). Of course, women would seek both resources and high genetic quality in one mate; however, mates displaying indicators of high genetic quality are highly attractive as sexual partner. Thus, they are difficult to commit to a long-term partnership, as they are offered a large number of reproductive opportunities by females. Hence, women might pursue the dual mating strategy when choosing the outstanding musician, which indicates high genetic fitness, as a short-term mate and thus seek his high-quality genes. On the other hand, females might favor the moderate musician as a long-term mate. Since he is less desirable (due to his lower genetic quality indicated by his musical abilities) and thus offered fewer mating chances by other females, females might assume that this mate will provide a higher and more stable amount of investment. Furthermore, females might select their sexual partners more thoroughly during their fertile phase, because they are at a higher stake of becoming pregnant resulting from sexual intercourse. During their fertile window, women are more attracted to men who display indicators of high genetic quality compared to men of inferior genetic quality (as measured by the degree of musicality).

The findings of Schwab and Bordin (2009) are a second example for a fruitful application of an evolutionary approach to media psychology. Their results support the validity of the ornamental mind theory (Miller 2000) in the context of musicality in lonely heart advertisements provided by the media. The ornamental mind of an entertainer (such as a musician) proved to serve as an indicator of high genetic quality and charmed and attracted potential sexual partners, just like the tail of the peacock evolved to seduce peahens.

7.6.3 Study 3: Gossip 2.0

Gossip refers to

> information about the deviant or surprising (which both depend on the context) traits and behaviors of a (or more) third person(s) (most often non-present, but potentially present in the conversation), where the goal of the conversation is non-professional, and the sender has true/false knowledge of the gossip content. (De Backer 2005, Chap. 1.5.3[1])

Research shows that gossip is ubiquitous, as up to two thirds of all human conversations deal with social issues (Dunbar 1998a). Engaging in gossip, however, takes a considerable amount of time and resources; so why do people universally gossip? This question can be answered by considering two different perspectives: the *proximate* perspective, explaining behavior in terms of causation (behavioral causes on a chemical, psychological, neuro-ethological, psychological, and social level) and *ontogeny* (the effect of the individual's experience on its behavior), and the *ultimate* perspective, describing behavior in terms of *survival value* (purpose and function of behavior on an individual level) and *evolution* (evolutionary function of individual behavior leading to an advantage in natural and sexual selection; Tinbergen 1963). In terms of proximate explanations, gossip serves as social glue (Turner et al. 2003), since it connects individuals and supports the development of positive relationships (Foster 2004). Moreover, individuals gossip to obtain information, influence, and entertainment (Rosnow 1977). According to the ultimate perspective, making allies within the social group against common enemies or other group members is also a very important function of gossip, because the forming of alliances significantly

---

[1] As the doctoral thesis of De Backer (2005) is solely available in an online HTML version, it is not possible to cite the exact page. Thus, the respective chapter is cited.

contributes to the individual's survival chance and reproduction success (Carolus 2014; Dunbar 1998a, b).

Dunbar (1998a, b, 2004) assumes that human gossip has evolved as an elaborated form of the *social grooming* behavior that can be observed in primates, the nearest still living relatives of modern man. The *social brainhypothesis* (Dunbar 1998b) explains the evolution of human language as follows: during human evolution, our ancestors living in social groups changed their habitat away from life in the trees to life on the ground of the savannah. Living on the ground, however, led to a higher predation risk. As an adaptation to this new risk, the group size increased to guarantee a better protection of its members. However, living in larger groups did not only provide advantages for the individuals, but also downsides, such as a higher stress level due to multiple conflicts resulting from the struggle for the access to resources. To solve this problem, our ancestors formed alliances that helped and supported them in conflict situations. This contributed to the social cohesion of the group. These alliances were enabled by social grooming. Living in a large social group demanded advanced social cognition from the group members: for a beneficial interaction, it was critical to understand the mental states of others. Thereby, it was possible to further manipulate their behavior for one's own benefits within the social group. An example of this advanced social cognition is the development of a *theory of mind*, which refers to the ability to attribute inner mental states (e.g., beliefs, desires, knowledge, and intents) to oneself and to other individuals and to understand that the inner mental states of others might differ from one's own mental states (e.g., Astington 1993). Consequently, primates and humans evolved a deep sociality. Sociality and mind reading, however, put heavy cognitive demands on our ancestors' brains, which resulted in an increased brain size. Living in larger social groups required higher "social investment," i.e., social grooming. But due to other mandatory activities (e.g., foraging), the available time for social grooming (a strictly one-to-one activity) was restricted. This led to an upper limit for group sizes. Dunbar (1992, 1993, 1998a) found a maximum of 80 individuals in a group of primates, whereas humans tend to live in groups twice as large. To allow this increase in group size, a more efficient mechanism for bonding individuals evolved: language. Compared to social grooming, language provides critical advantages: By talking, individuals can interact with more than one individual at the same time and form several relationships simultaneously. In addition, individuals can gain much more insight into the group's social network, as they receive information about other individuals during conversations with third persons. Monkeys lack this ability and are therefore limited to visual information about other individuals.

Based on Dunbar's social brain hypothesis (1998b), De Backer (2005) clustered and differentiated contents of gossip around evolutionarily relevant fitness problems (see Fig. 7.5). She roughly distinguishes between *strategy learning gossip* (learning behavior strategies cost-effectively from other's experiences) and *reputation gossip* (learning about the reputation of other social network members and manipulating reputation to one's own advantage). Strategy learning gossip deals with issues of *survival* (food supply, health, shelter), *mating* (how to tie a sexual partner to oneself), and

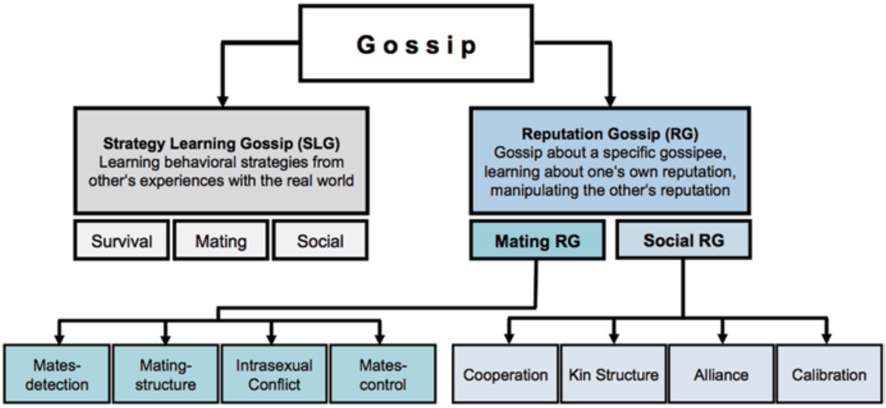


Fig. 7.5 Classification of gossip contents. (Carolus 2014; adapted from De Backer 2005)

social norms (how to interact with other individuals according to the social and cultural norms of the group). Reputation gossip is divided into *mating reputation gossip*, including *mates detection, mating structure* (successful mating requires the detection of potential mates with a high level of fitness), *intrasexual conflict* (detecting and eliminating potential same-sex rivals by spreading rumors to damage their reputation and to portray oneself in a good light), *mates control* (gaining information about the own mate's behavior to find out about cheating), *social reputation gossip,* encompassing *cooperation* (detecting altruistic individuals and free riders), *kinship* (information on family relationships), *alliance* (finding allies and maintaining them), and *calibration* (regularly updating the existing impression of interaction partners).

In summary, De Backer (2005) subdivides gossip into different fields, with every field representing a specific adaptive problem our ancestors had to face during Pleistocene (cf. Carolus 2014). As these adaptive problems are still relevant for modern humans today, the gossip contents, according to De Backer (2005), are supposed to be found in everyday communication settings. Since its invention, the Internet has been largely used for communication. Especially with the emergence of SNS (e.g., Facebook), the World Wide Web has provided the individual user with a “stage” for self-portrayal and self-presentation. By means of SNS, users can display self-generated information about themselves (e.g., personal information), about other individuals (e.g., photos and wall posts), and virtually connect with family members and friends they know from real life. The average individual user is depicted in his “real” social network with his “real-life identity,” because he usually connects online with individuals he also encounters off-line (Boyd and Ellison 2007).

Given the parallels between the virtual and the real social networks, it is likely that SNS are used to convey gossip contents. Carolus (2014) conducted a content analysis of the user profiles of SNS in order to examine the gossip contents

Table 7.1 Mating-relevant gossip information. (Carolus 2014; based on Buss 2004; Buss and Barnes 1986; Buss and Schmitt 1993; Schmitt and Buss 2000)

Mating-relevant information
Youth and physical attractiveness, indicating fertility and reproductive value
Willingness to provide a family
Sexual exclusiveness
Information on (prospective) status
Creativity/intelligence
Sexual accessibility

according to De Backer (2005) within the Internet.[2] First analyses revealed that user profiles provided manifold information that can be interpreted as mating or social reputation gossip (e.g., photos depicting the user with friends or mates, relationship status information). However, De Backer's classification of gossip contents could not be empirically supported, as a clear differentiation between all postulated categories proved to be difficult. As an initial starting point, Carolus (2014) therefore focused on the analysis of the higher-order categories of social and reputation gossip. Criteria for mating reputation gossip were derived from Buss and his colleagues, who identified relevant characteristics men and women seek in short-term and long-term mates (see Table 7.1; Buss 2004; Buss and Barnes 1986; Buss and Schmitt 1993; Schmitt and Buss 2000). They distinguished between men's preferences in a short-term mate (promiscuous and sexually experienced women; women who do not intend to form a long-term relationship, physically attractive women) and in a long-term mate (young and faithful women indicating fertility and reproductive value). Likewise, it was differentiated between women's preferences in a short-term mate (single men that are strong, athletic, generous, intelligent, and creative; indicating high genetic quality) and a long-term mate (men who occupy a high social status and lots of resources; educated, kind, emotionally stable, mature, and ambitious men; see also Schmitt 2005). According to SST (Buss and Schmitt 1993), both men and women were assumed to display mating-relevant gossip information in their SNS profiles, showing significant differences between the sexes.[3] Carolus (2014) analyzed different sections of SNS profiles, including profile photos, uploaded and linked photos, demographic and professional information, and explicit mating relevant information (relationship status) in terms of the deduced mating relevant aspects (Table 7.1).

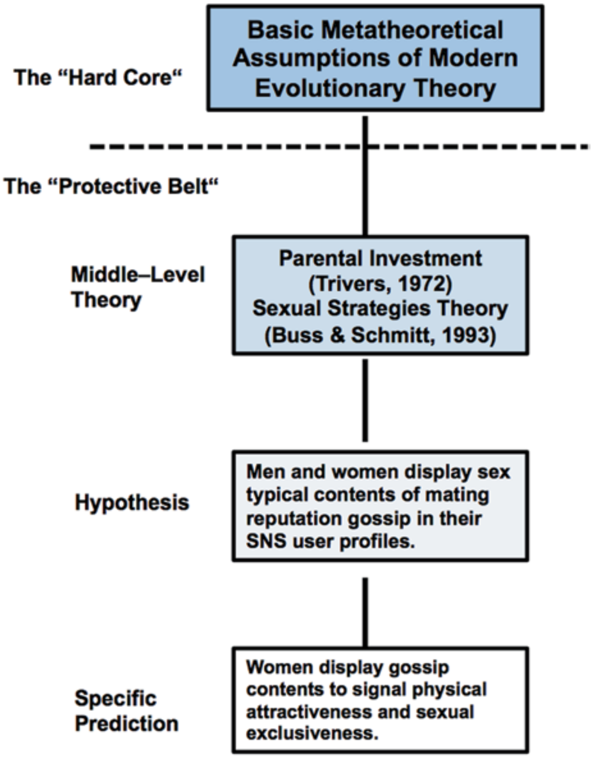
Within the Lakatosian structure of scientific analysis, SST serves as mid-level-theory (Buss 2004; Buss and Barnes 1986; Buss and Schmitt 1993; Schmitt and Buss 2000; see Fig. 7.6). The hypothesis that sex-specific gossip contents regarding mating reputation gossip are found in SNS is located on the next lower level. Finally, the specific prediction that women underline their physical attractiveness and sexual exclusiveness in their self-presentation on their SNS profile photo is situated

---

[2] The SNS http://www.xing.com (a typical example of professional SNS) and http://www.wer-kennt-wen.de (a typical example of private SNS) were analyzed.

[3] Due to research heuristics, it was only focused on strategies of heterosexual individuals.

Fig. 7.6 The hierarchical structure of evolutionary psychological explanations based on the examples of men's and women's display of sex-typical mate reputation gossip in SNS. (Carolus 2014; based on Buss 1995 and Ketelaar and Elist 2000)



on the lowest level of the Lakatosian epistemological model. Results showed that there were almost an equal number of men and women having a user profile in SNS (Carolus 2014). Detailed analyses of the owner's profile photo revealed a trend that women uploaded profile photos more frequently than men. This can be explained in terms of "self-advertising": men value the partner's physical qualities in mate choice more than women (Buss 1989; Buss and Barnes 1986), as the outward appearance serves as an honest signal of the woman's genetic fitness and reproductive value (Furnham et al. 2006; Grammer et al. 2003). However, no differences between the sexes were found concerning the amount of exposed naked skin. Naked skin is also viewed as an indicator of genetic fitness and reproductive value, such that women should be more inclined to upload profile photos showing a greater amount of naked skin. Moreover, about 5 % of men and women depicted themselves with an opposite-sex partner, which was interpreted as a sign of sexual exclusiveness. The analysis of the profile photos showed that women provided more hints to sexual exclusiveness than men, but due to the small number of cases, no statistically significant differences between the self-presentation of the sexes were found. In addition to profile pictures, Carolus (2014) examined the profile owners' relationship status. Roughly 36 % of men and 30 % of women indicated a relationship status. Approximately, 18 % of men and 23 % of women declared to be in a long-term relationship, whereas 18 % of men and only 8 % of women indicated to be single. In line with SST (Buss and Schmitt 1993), these results support the hypothesis that men aim at mating with a high number of sexual partners, as by indicating to be single they present themselves as sexually available. Women, on the other hand, elicit the impression to be sexually faithful and signal sexual exclusiveness by declaring to be in a long-term relationship. By conveying such information, women increase their value as a mate as they guarantee high paternity certainty (Buss 1989). This, in turn, increases the likelihood that their mate will provide for them and their children (Buss 1989).

Carolus (2014) showed that gossip contents derived from an evolutionary framework can also be found in computer-mediated communication. As predicted, the content analysis of SNS user profiles revealed the display of mating reputation gossip, such as relationship status, reproductive value, and sexual exclusiveness. Moreover, in the gossip contents, Carolus (2014) found the sex-typical differences deduced from SST (Buss and Schmitt 1993). The assumption, however, that women show more naked skin on their profile photos than men could not be empirically supported; hence, this specific prediction can be falsified. Since according to Lakatosian epistemology, the underlying theories are not tested directly, but derived predictions and hypotheses that constitute their "protective" belt, the rejection of this specific prediction does not imply that the underlying evolutionary theories of the "hard core" have to be refuted as well.

7.7 Conclusion

This chapter presented several mid-level theories from the "protective belt" of evolutionary psychology, explaining the phenomenon of human's universal interest in the media from the viewpoint of evolutionary media psychology. Based on selected studies dealing with sex-typical media preferences, musical seduction, and gossip in SNS, it was demonstrated that applying an evolutionary approach to media psychology results in a fruitful research heuristic.

All three exemplary studies start from the metatheoretical "hard core" of modern evolutionary theory. Study one uses the well-established assumptions of parental investment (Trivers 1972) as a mid-level theory to deduce hypotheses and specific predictions about sex-typical preference in movie genres. While focusing on evolutionary psychology explanations, the study also addresses Standard Social Science Model explanations (Tooby and Cosmides 1992), namely social role theory (Eagly et al. 2000), and identifies the explanatory contribution of each theory by applying a multiple regression analysis (Schwab 2010a). This study shows that not only can evolutionary psychology theories within the "protective belt" compete against each other, but that theories based on non-evolutionary explanations (see arbitrary culture; Alcock 2005) can also be taken into account. In the second study, based on the ornamental mind theory (Miller 2000), it was examined why humans produce (media) entertainment and how their abilities and performances are evaluated in terms of mate quality (Schwab and Bordin 2009). It was shown that the musical abilities of men influenced women's mate choices differently, depending on a long-term or short-term mating context. Not only is this a new scientific approach in communicator research, but it also clarifies the seemingly irrational, pretentious, extravagant, and dazzling aspects of the media. The last study deals with gossip in SNS. Evolutionary gossip theories (De Backer 2005; Dunbar 1998b) were used to design and conduct a media content analysis of users' profiles on the Internet (Carolus 2014). As this analysis of gossip content was carried out in a more exploratory manner, some assumptions were not supported by the collected data. This study demonstrated that predictions, hypotheses, and even theories within the "protective belt" can be rebutted. Refuting falsified predictions, hypotheses, or theories within the "protective belt," however, does not directly challenge the "hard core" assumptions of the metatheoretical setting; it rather calls for better-developed and more valid theories on the mid-level of explanation.

Hopefully, this chapter has been able to convey the exciting and promising state of research that the still relatively young science of evolutionary psychology currently finds itself in. Especially in media psychology, a field of science that constantly has to defend itself against the allegations of being a mere product of culture and socialization, an evolutionary approach is an enlightening and essential step for further research.

In a nutshell, this chapter provided clear examples showing that evolutionary psychology offers a research heuristic that is a useful and promising tool for scientific media research. It further connects the science of (media) psychology with the other life sciences. Although evolutionary psychology is still a young science, this chapter contributed to the scientific validity and quality of evolutionary psychology by demonstrating in detail how the scientific approach of evolutionary media psychology is founded on Lakatosian philosophy of science.

References

Alcock, J. (2005). *Animal behavior: An evolutionary approach* (8th ed.). Sunderland: Sinauer Associates.

ARD-ZDF-Onlinestudie. (2012). http://www.ard-zdf-onlinestudie.de/index.php?id=353. AccesseD 22 Jan 2013.

Asendorpf, J. (1996). *Personality psychology. Basics, Psychologie der Persönlichkeit. Grundlagen*. Berlin: Springer.

Astington, J. (1993). *The child's discovery of the mind*. Cambridge: Harvard University Press.

Barkow, J. H., Cosmides, L., & Tooby, J. (1992). *The adapted mind: Evolutionary psychology and the generation of culture*. New York: Oxford University Press.

Bem, S. L. (1974). The measurement of psychological androgyny. *Journal of Consulting and Clinical Psychology, 42*(2), 155–162. doi:10.1037/h0036215.

Bischof, N. (1985). *The puzzle of Oedipus, Das Rätsel Ödipus*. München: Piper.

Bischof-Köhler, D. (2002). *Different by nature, Von Natur aus anders*. Stuttgart: Kohlhammer.

Boyd, D. M., & Ellison, N. B. (2007). Social network sites: Definition, history, and scholarship. *Journal of Computer-Mediated Communication, 13*(1), 210–230. doi:10.1111/j.1083-6101.2007.00393.x.

Buller, D. J. (2005). Evolutionary psychology: The emperor's new paradigm. *Trends in Cognitive Sciences, 9*(6), 277–283. doi:10.1016/j.tics.2005.04.003.

Buss, D. M. (1989). Sex differences in human mate preferences: Evolutionary hypotheses tested in 37 cultures. *Behavioral and Brain Sciences, 12*(1), 1–49. doi:10.1017/S0140525X00023992.

Buss, D. M. (1995). Evolutionary psychology: A new paradigm for psychological science. *Psychological Inquiry,6*(1), 1–30. doi:10.1207/s15327965pli0601_1.

Buss, D. M. (1999). *Evolutionary psychology: The new science of the mind*. Needham Heights: Allyn & Bacon.

Buss, D. M. (2001). Human nature and culture: An evolutionary psychological perspective. *Journal of Personality, 69*(6), 955–978. doi:10.1111/1467–6494.696171.

Buss, D. M. (2004). *Evolutionary psychology: The new science of the mind* (2nd ed.). Boston: Pearson.

Buss, D. M., & Barnes, M. (1986). Preferences in human mate selection. *Journal of Personality and Social Psychology, 50*(3), 559–570. doi:10.1037/0022-3514.50.3.559.

Buss, D. M., & Schmitt, D. P. (1993). Sexual strategies theory: An evolutionary perspective on human mating. *Psychological Review, 100*(2), 204–232. doi:10.1037/0033-295X.100.2.204.

Carolus, A. (2014). *Gossip 2.0.—media communication in social network sites, Gossip 2.0.—Mediale Kommunikation in Sozialen Netzwerkseiten*. Stuttgart: Kohlhammer.

Confer, J. C., Easton, J. A., Fleischman, D. S., Goetz, C. D., Lewis, D. G., Perilloux, C., & Buss, D. M. (2010). Evolutionary psychology: Controversies, questions, prospects, and limitations. *American Psychologist, 65*(2), 110–126. doi:10.1037/a0018413.

Cosmides, L., & Tooby, J. (1992). Cognitive adaptations for social exchange. In J. H. Barkow, L. Cosmides, & J. Tooby (Eds.), *The adapted mind: Evolutionary psychology and the generation of culture* (pp. 163–228). New York: Oxford University Press.

Cosmides, L., & Tooby, J. (1994). Beyond intuition and instinct blindness: Toward an evolutionarily rigorous cognitive science. *Cognition*, *50*(1–3), 41–77. doi:10.1016/0010-0277(94)90020-5.

Darwin, C. (1859). *On the origin of species by means of natural selection: Or the preservation of the favoured races in the struggle for life*. London: John Murray. doi:10.1037/14088-000.

Darwin, C. (1871). *The descent of man, and selection in relation to sex* (Vol. 1). London: John Murray. doi:10.1037/12293-000.

Dawkins, R. (1976). *The selfish gene*. New York: Oxford University Press.

Dawkins, R. (1982). *The extended phenotype*. San Francisco: Freeman.

Dawkins, R. (1999). *The extended phenotype: The long reach of the gene*. Revised edition. New York: Oxford University Press.

Dawkins, R. (2004). Extended phenotype—but not too extended. A reply to Laland, Turner and Jablonka. *Biology and Philosophy, 19*(3), 377–396. doi:10.1023/B:BIPH.0000036180.14904. 96.

De Backer, C. (2005). *Like Belgian chocolate for the universal mind: Interpersonal and media gossip from an evolutionary perspective*. http://www.ethesis.net/gossip/gossip_contence.htm. Accessed 16 Jan 2005.

Dennett, D. C. (1997). *Darwin's dangerous inheritance, Darwins gefährliches Erbe*. Hamburg: Hoffmann und Campe.

Dennett, D. C. (2008). Intentional systems, *Intentionale Systeme*. In P. A. Bieri (Ed.), *Analytical philosophy of the mind, Analytische Philosophie des Geistes*. Weinheim: Beltz.

Dissanayake, E. (1992). Homo Aestheticus: Where art comes from and why. Seattle: University of Washington Press.

Dissanayake, E. (1988). *What is art for?* . Seattle: University of Washington Press.

Dissanayake, E. (1995). *Homo aestheticus: Where art comes from and why*. Seattle: University of Washington Press.

Dissanayake, E. (2000). *Art and intimacy: How the arts began*. Seattle: University of Washington Press.

Dunbar, R. (1998a). *Grooming, gossip, and the evolution of language*. Cambridge: Harvard University Press.

Dunbar, R. I. M. (1992). Neocortex size as a constraint on group size in primates. *Journal of Human Evolution, 22*(6), 469–493. doi:10.1016/0047-2484(92)90081-J.

Dunbar, R. I. M. (1998b). The social brain hypothesis. *Evolutionary Anthropology, 6*, 178–190. doi:10.1002/(SICI)1520–6505 (1998) 6:5 < 178::AID-EVAN5 > 3.0.CO;2–8.

Dunbar, R. M. (1993). Coevolution of neocortical size, group size and language in humans. *Behavioral and Brain Sciences, 16*(4), 681–735. doi:10.1017/S0140525X00032325.

Dunbar, R. M. (2004). Gossip in evolutionary perspective. *Review of General Psychology, 8*(2), 100–110. doi:10.1037/1089-2680.8.2.100.

Eagly, A. H., Wood, W., & Diekman, A. B. (2000). Social role theory of sex differences and similarities: A current appraisal. In T. Eckes & H. M. Trautner (Eds.), *The developmental social psychology of gender* (pp. 123–174). Mahwah: Lawrence Erlbaum Associates Publishers.

Fitch, W., Hauser, M. D., & Chomsky, N. (2005). The evolution of the language faculty: Clarifications and implications. *Cognition, 97*(2), 179–210. doi:10.1016/j.cognition.2005.02.005.

Fodor, J. (2005). Reply to Steven Pinker 'So how does the mind work?'. *Mind & Language, 20*(1), 25–32. doi:10.1111/j.0268-1064.2005.00275.x.

Foster, E. K. (2004). Research on gossip: Taxonomy, methods, and future directions. *Review of General Psychology, 8*(2), 78–99. doi:10.1037/1089-2680.8.2.78.

Früh, W. (2002). *Entertainment through television. A molar theory, Unterhaltung durch das Fernsehen. Eine molare Theorie*. Konstanz: UVK.

Furnham, A., Swami, V., & Shah, K. (2006). Body weight, waist-to-hip ratio and breast size correlates of ratings of attractiveness and health. *Personality and Individual Differences, 41*(3), 443–454. doi:10.1016/j.paid.2006.02.007.

Gangestad, S. W., & Simpson, J. A. (2000). The evolution of human mating: Trade-offs and strategic pluralism. *Behavioral and Brain Sciences*, *23*(4), 573–644. doi:10.1017/S0140525 X0000337X.

Gigerenzer, G., & Selten, R. (2001). *Bounded rationality: The adaptive toolbox*. Cambridge: MIT.

Grammer, K., Fink, B., Møller, A. P., & Thornhill, R. (2003). Darwinian aesthetics: Sexual selection and the biology of beauty. *Biological Reviews, 78*(3), 385–407. doi:10.1017/S146479 3102006085.

Greiling, H., & Buss, D. M. (2000). Women's sexual strategies: The hidden dimension of extra-pair mating. *Personality and Individual Differences, 28*(5), 929–963. doi:10.1016/S0191-8869(99)00151-8.

Hagen, E. H. (2005). Controversial issues in evolutionary psychology. In D. M. Buss (Ed.), *The handbook of evolutionary psychology* (pp. 145–173). Hoboken: Wiley.

Hamilton, W.D. (1964). The genetical evolution of social behavior. *Journal of Theoretical Biology, 7*(1), 17–52. doi:10.1016/0022-5193(64)90039-6.

Holcomb, H. (1998). Testing evolutionary hypotheses. In C. B. Crawford & D. L. Krebs (Eds.), *Handbook of evolutionary psychology: Ideas, issues, and applications* (pp. 303–334). Mahwah: Lawrence Erlbaum.

Horton, D., & Wohl, R. (1956). Mass communication and para-social interaction. *Psychiatry: Journal for the Study of Interpersonal Processes, 19,* 215–229.

Ketelaar, T., & Ellis, B. J. (2000). Are evolutionary explanations unfalsifiable? Evolutionary psychology and the Lakatosian philosophy of science. *Psychological Inquiry, 11*(1), 1–21. doi:10.1207/S15327965PLI1101_01.

Lakatos, I. (1970). Falsifications and the methodology of scientific research programmes. In I. Lakatos & A. Musgrave (Eds.), *Criticism and the growth of knowledge* (pp. 91–196). Cambridge: Cambridge University Press.

Lakatos, I. (1978). *The methodology of scientific research programs: Philosophical papers* (Vol. 1). Cambridge: Cambridge University Press.

Lewontin, R. C. (1990). The evolution of cognition. In D. N. Osherson & E. E. Smith (Eds.), *Thinking: An invitation to cognitive science* (Vol. 3, pp. 229–246). Cambridge: MIT

Lewontin, R. C. (1998). The evolution of cognition: Questions we will never answer. In D. Scarborough & S. Sternberg (Eds.), *Methods, models, and conceptual issues: An invitation to cognitive science* (Vol. 4, pp. 106–132). Cambridge: MIT.

Manning, J. T. (2002). *Digit ratio: A pointer to fertility, behavior and health*. New Brunswick: Rutgers University Press.

Mealey, L. (2000). Sex differences. Developmental and evolutionary strategies. San Diego: Academic Press.

Miller, G. (2000). *The mating mind: How sexual choice shaped the evolution of human nature*. New York: Doubleday.

Miller, G. F. (1998). Sexual selection for cultural displays. In R. Dunbar, C. Knight, & C. Power (Eds.), *The evolution of culture*. Edinburgh: University Press.

Newell A. (1973). You can't play 20 questions with nature and win: Projective comments on the papers in this symposion. In W. G. Chase (Ed.), Visual informactino processing (pp. 283–308). New Yord: Academic.

Ohler, P. (2001). *Play, evolution, cognition. From the origins of play to the computer games, Spiel, Evolution, Kognition. Von den Ursprüngen des Spiels bis zu den Computerspielen*. Bad Heilbrunn: Julius Klinkhardt.

Ohler, P., & Nieding, G. (2006). An evolutionary perspective on entertainment. In J. Bryant & P. Vorderer (Eds.), *Psychology of entertainment* (pp. 423–433). Mahwah: Lawrence Erlbaum.

Penke, L., & Asendorpf, J. B. (2008). Beyond global sociosexual orientations: A more differentiated look at sociosexuality and its effects on courtship and romantic relationships. *Journal of Personality and Social Psychology, 95*(5), 1113–1135. doi:10.1037/0022-3514.95.5.1113.

Pillsworth, E. G., & Haselton, M. G. (2006a). Male sexual attractiveness predicts differential ovulatory shifts in female extra-pair attraction and male mate retention. *Evolution and Human Behavior, 27*(4), 247–258. doi:10.1016/j.evolhumbehav.2005.10.002.

Pillsworth, E. G., & Haselton, M. G. (2006b). Women's sexual strategies: The evolution of long-term bonds and extrapair sex. *Annual Review of Sex Research, 17*(1), 59–100. doi:10.1080/10532528.2006.10559837.

Pinker, S. (1997). *How the mind works*. New York: W Norton.

Popper, K. R. (1959). *The logic of scientific discovery*. Oxford: Basic Books.

Reeves, B., & Nass, C. (1996). *The media equation: How people treat computers, television, and new media like real people and places*. Chicago: Center for the Study of Language and Information.

Regan, P. C. (1998). What if you can't get what you want? Willingness to compromise ideal mate selection standards as a function of sex, mate value, and relationship context. *Personality and Social Psychology Bulletin, 24*(12), 1294–1303. doi:10.1177/01461672982412004.

Reinisch, J. M. (1981). Prenatal exposure to synthetic progestins increases potential for aggression in humans. *Science, 211*(4487), 1171–1173. doi:10.1126/science.7466388.

Rosnow, R. L. (1977). Gossip and marketplace psychology. *Journal of Communication, 27*(1), 158–163. doi:10.1111/j.1460-2466.1977.tb01811.x.

Schlinger, H. D. (2002). Not so fast, Mr. Pinker: A behaviorist looks at the Blank Slate. A review of Steven Pinker's the Blank Slate: The modern denial of human nature. *Behavior and Social Issues, 12*(1), 75–79.

Schmitt, D. P. (2005). Fundamentals of human mating strategies. In D. M. Buss (Ed.), *The handbook of evolutionary psychology* (pp. 258–291). Hoboken: Wiley.

Schmitt, D. P., & Buss, D. M. (2000). Sexual dimensions of person description: Beyond or subsumed by the Big Five? *Journal of Research in Personality, 34*(2), 141–177. doi:10.1006/jrpe.1999.2267.

Schramm, H. (2008). Parasocial interaction, Parasoziale Interaktion. In N. C. Krämer, S. Schwan, D. Unz, & M. Suckfüll (Eds.), *Media psychology: Key terms and concepts, Medienpsychologie: Schlüsselbegriffe und Konzepte* (pp. 253–258). Stuttgart: Kohlhammer.

Schreier, M., & Appel, M. (2002) The distinction between reality and fiction as an aspect of a critical-constructive media-use-competence, 'Realitäts-Fiktions-Unterscheidungen als Aspekt einer kritisch-konstruktiven Mediennutzungskompetenz'. In N. Groeben & B. Hurrelmann (Eds.), *Media competency. Preconditions, dimensions, functions, Medienkompetenz. Voraussetzungen, Dimensionen, Funktionen* (pp. 33–56). Weinheim: Juventa.

Schwab, F. (2004) *Evolution and emotion: Evolutionary perspectives in emotion research and applied psychology, Evolution und Emotion: evolutionäre Perspektiven in der Emotionsforschung und der angewandten Psychologie*. Stuttgart: Kohlhammer.

Schwab, F. (2006). Are we amusing ourselves to death? Answers from evolutionary psychology. *SPIEL. Siegener Periodikum zur internationalen empirischen Literaturwissenschaft, 22*(2), 329–338.

Schwab, F. (2008). Evolutionary explanations, Evolutionäre Erklärungsansätze. In N. C. Krämer, S. Schwan, D. Unz, & M. Suckfüll (Eds.), *Media psychology: Key terms and concepts, Medienpsychologie: Schlüsselbegriffe und Konzepte* (pp. 41–46). Stuttgart: Kohlhammer.

Schwab, F. (2010a). *Moving pictures: An evolutionary media pychology of entertainment, Lichtspiele: Eine evolutionäre Medienpsychologie der Unterhaltung*. Stuttgart: Kohlhammer.

Schwab, F. (2010b). Darwin as vision aids for psychologists—evolutionary psychology, Darwin als Sehhilfe für die Psychologen—Evolutionspsychologie. In J. Oehler (Ed.), *The human being—evolution, nature, and culture: Contributions to our prevailing understanding of human beings, Der Mensch—Evolution, Natur und Kultur: Beiträge zu unserem heutigen* (pp. 75–98). Berlin: Springer.

Schwab, F., & Bordin, I. (2009). "Money for nothing and chicks for free"—A Darwinian perspective on entertainment behavior. Proceedings of the 6th Conference of the Media Psychology Division of the German Psychological Society, 9–11 September. Duisburg.

Schwab, F., & Schwender, C. (2010). The decent of emotions in media: Darwinian perspectives. In K. Doveling, C. V. Scheve, & E. A. Konijn (Eds.), *The Routledge handbook of emotion and mass media* (pp. 15–36). New York: Routledge.

Schwender, C. (2006) *Media and emotions. Evolutionary psychological components of a mediatheory, Medien und Emotione. Evolutionspsychologische Bausteine einer Medientheorie* (2nd ed.). Wiesbaden: DUV.

Sherry, J. L. (2004). Media effects theory and the nature/nurture debate: A historical overview and directions for future research. *Media Psychology, 6*(1), 83–109. doi:10.1207/s1532785xmep0601_4.

Silvers, S. (2010). Methodological and moral muddles in evolutionary psychology. *Journal of Mind and Behavior, 31*(1–2), 65–84.

Stern, D. N. (2002). *The first relationship: Infant and mother*. Cambridge: Harvard University Press.

Tinbergen, N. (1963). On the aims and methods of ethology. *Zeitschrift für Tierpsychologie, 20*(4), 410–433. doi:10.1111/j.1439-0310.1963.tb01161.x.

Tomasello, M., & Call, J. (1997). *Primate cognition*. New York: Oxford University Press.

Tooby, J., & Cosmides, L. (1992). The psychological foundations of culture. In J. H. Barkow, L. Cosmides, & J. Tooby (Eds.), *The adapted mind: Evolutionary psychology and the generation of culture* (pp. 19–136). New York: Oxford University Press.

Tooby, J., & Cosmides, L. (2001). Does beauty build adapted minds? Toward an evolutionary theory of aesthetics, fiction and the arts. *SubStance, 30*(94–95), 6–27. doi:10.2307/3685502.

Trafimow, D., & Gambacorta, D. (2012). How obvious are hypotheses in evolutionary psychology?. *The Journal of Social, Evolutionary, and Cultural Psychology, 6*(1), 1–12.

Trivers, R. L. (1972). Parental investment and sexual selection. In B. Campbell (Ed.), *Sexual selection and the descent of man: 1871–1971* (pp. 136–179). Chicago: Aldine.

Turner, M., Mazur, M. A., Wendel, N., & Winslow, R. (2003). Relational ruin or social glue? The joint effect of relationship type and gossip valence on liking, trust, and expertise. *Communication Monographs, 70*(2), 129–141. doi:10.1080/0363775032000133782.

Voracek, M., & Loibl, L. M. (2009). Scientometric analysis and bibliography of digit ratio (2D:4D) research. *Psychological Reports, 104*(3), 922–956. doi:10.2466/pr0.104.3.922-956.

Winterhoff-Spurk, P. (2004). *Media psychology. An introduction, Medienpsychologie. Eine Einführung* (2nd ed.) Stuttgart: Kohlhammer.

Workman, L., & Reader, W. (2008). *Evolutionary psychology: An introduction* (2nd ed.). New York: Cambridge University Press.

Chapter 8
Preparedness to Learn About the World: Evidence from Infant Research

Sabina Pauen and Stefanie Hoehl

A German version of this chapter can be found in Breyer et al. (2013).

8.1 Introduction

Developmental psychology aims at describing age-related changes and identifying their underlying mechanisms. Modern accounts acknowledge that developmental progress depends on biological and maturational processes as well as on learning experiences, personal interactions, and on broader socio-cultural influences. In this chapter, we will focus on one aspect of development that has recently received much attention: knowledge acquisition in social contexts.

Infants learn a great deal from other people. Through observation and imitation, infants acquire knowledge about the emotional valence of things (Mumme and Fernald 2003), the function of unfamiliar objects (Elsner 2007; Träuble and Pauen 2007), social norms (Schmidt et al. 2012), and many other aspects. In the following paragraphs, we will review recent evolutionary accounts on early cognitive development and knowledge acquisition in social contexts. Are the ways in which infants and young children learn from others shaped by evolved adaptive mechanisms? Has evolution provided us with mechanisms of cultural transmission that are universal in humans, but unique to our species? It is unequivocal that spoken language is unique to humans as a means of communication and transmission of knowledge. As will be demonstrated, we may also have evolved mechanisms that allow for

---

S. Pauen (✉)
Institute of Psychology, University of Heidelberg, Heidelberg, Germany
e-mail: sabina.pauen@psychologie.uni-heidelberg.de

S. Hoehl
Institute of Psychology, University of Heidelberg, Heidelberg, Germany
e-mail: stefanie.hoehl@psychologie.uni-heidelberg.de

T. Breyer (ed.), *Epistemological Dimensions of Evolutionary Psychology,*
DOI 10.1007/978-1-4939-1387-9_8

learning and teaching of cultural knowledge long before language comes into play (Csibra and Gergely 2009).

After providing a general overview of current theories and empirical findings on infants' preparedness to acquire knowledge, we will focus on one particular aspect of knowledge acquisition with high relevance to evolutionary psychology: observational fear learning. Illustrated by an example from our own research, we will show how evolutionary psychology can inform developmental research and vice versa. Studying infants provides a promising way to find out more about how evolution has prepared us to perceive, evaluate, and process information about objects that are potentially dangerous. Only if we examine cognitive processes in very young children with limited experience will we be able to fully evaluate the validity of preparedness accounts.

8.2 Evolutionary Accounts of Knowledge Acquisition

One fundamental question raised by philosophers as well as by developmental psychologists is whether infants come to life as *tabula rasa* (John Locke, 1632–1704), without any knowledge about the world, or whether infants come equipped with innate knowledge. Kant (1724–1804) introduced the term *categories of thinking,* referring to the argument that we analyze every object or situation that we encounter in terms of temporal and spatial relations. Modern developmental psychologists go even further and speak of *core knowledge* (e.g., Spelke and Kinzler 2007; Baillargeon and Carey 2012) or *innate knowledge acquisition devices* (Briscoe 2000), thereby suggesting that evolution might have prepared us to process information in ways that allow for quick adaptation and efficient learning. Just like the visual cortex in our brain has evolved to process visual information, more specific neural circuits may have developed to process information about objects, properties, or relations that have existed for thousands of years and that can be found in many places around the world.

Elaborating on this general idea, Elisabeth Spelke, as well as Renée Baillargeon and colleagues conducted numerous experiments exploring infants' innate knowledge about physics. The basic claim of the *core-knowledge account* is that infants are born with some core-concepts, allowing them to reason about physical relations form early on. For example, Spelke (1990) suggested that infants expect objects to take up a unique place in space and time. No material entity can move through another material entity. This core-principle is called "solidity." *Core-knowledge* never gets overwritten or invalid in later life, but may only become more differentiated with increased experience (Spelke and Kinzler 2007). How infants refine their innate knowledge structures may also be predetermined and follow specific rules of knowledge acquisition (Baillargeon and Carey 2012).

It soon became evident that the physical domain is not the only one in which innate knowledge might exist. Other concepts and knowledge acquisition devices or strategies have also been identified for the domains of language (e.g., Markman 1994), mathematics (e.g., Wynn 1995) or geometry (e.g., Spelke 2011). Hence, modern infant research refers to *domain-specific knowledge development*, assuming that the basis for discriminating between different domains has emerged in the course of human evolution and can be traced back to a preverbal age.

This idea became highly prominent when developmental researchers discovered that young infants make a fundamental categorical distinction between animate beings and inanimate objects, assuming that each kind of entity has typical visual features and shows a characteristic pattern of behavior (Gelman and Spelke 1981; Mandler 1992). It has been shown, for example, that infants less than 1 year of age expect only animals or humans to show self-propelled motion (Kosugi et al. 2003; Pauen and Träuble 2009; Spelke et al. 1995) that is goal-directed (e.g., Biro and Leslie 2007) and involves changes in speed, direction, and orientation (Rakison and Poulin-Dubois 2001). In contrast, they expect inanimate objects to start moving or to change their path and speed only after being contacted by an external force.

From birth on, infants seem highly sensitive to causes and effects of changes (Pauen 2000). Starting in their first year of life, they analyze any situation and event in causal and functional terms, always driven by the necessity to make valid predictions about the future actions of entities in their immediate environment. This early sensitivity to causal and functional information—combined with the fact that any natural environment contains entities capable of self-propelled biological motion as well as other entities that are not self-propelled—may lead them to develop a global animate–inanimate distinction early in life.

In line with this assumption, behavioral categorization research reveals that infants as young as 2 to 4 months of age are capable of discriminating animals from either furniture (Behl-Chadha 1996) or vehicles (Quinn and Johnson 2000) in a visual fixation task. When 3D toy-models are presented that can be explored visually and manually in an object examination task, infants again show signs of global category discrimination at an early age (Mandler and McDonough 1994; Pauen 2002a, b). We have recently been able to confirm this finding for 4- and 7-month-olds using brain waves (event-related potentials) as dependent measures. More specifically, we presented a very broad range of different exemplars from various basic-level categories of animals, such as fish, birds, reptiles, insects, or mammals and of furniture items, such as beds, chairs, tables, cupboards, lamps, and shelves. Each exemplar was presented on a computer monitor for only one second. Nonetheless, infants were well able to discriminate the categories in different task contexts (Jeschoneck et al. 2010; Pauen et al. 2011; Elsner et al. in press). This suggests that the brain of young infants is capable of deciding rather quickly whether a given entity is an animate being or an inanimate object.

Since we presented only static information about the appearance of stimuli and since participants were not yet familiar with every exemplar presented, this leaves room for speculations about the basis for this performance: (1) Could it be that our perceptual system is prepared to easily abstract visual correlational features, thus allowing the infant to learn category discrimination online while the test is still running? (2) Could it be that all animate beings have abstract visual features in common that infants are prepared to process automatically in a special way? (3)

Or might it be the case that infants have already learned to associate certain characteristic visual properties with behavioral properties (e.g., self-propelled motion) and that they respond to the activation of this memory? These different accounts do not rule each other out, thus making it rather difficult to determine the nature of infants' preparedness to identify animate beings. Based on the literature, we assume that there is some truth to each option mentioned above and that different abilities jointly contribute to infants' early categorization performance.

Most animals have the same general body structure, usually including a head with facial features needed for communicating and taking up nutrition, and some kind of appendages that allow for movements. They also have a clear axis of symmetry and characteristic surface features (e.g., skin, fur, feathers). This information may help infants to quickly recognize animate beings on pictures or in their surroundings. Early sensitivity to these features may have developed due to evolutionary pressure.

For example, it has been shown that even newborns seem prepared to process facial information in a special way. More specifically, they have been shown to respond with increased attention to human eyes that are unique in terms of showing a black pupil in front of a white sclera (Farroni et al. 2005). They also increase attention in response to "3-blob" configurations resembling a human face with two eyes above one mouth (both dark), located within an oval form of lighter color (Johnson and Morton 1991). According to the authors, this face-recognition mechanism is based on automatic perceptual processes that do not require sophisticated visual abilities or a higher-order cortical analysis.

Apart from static perceptual features, infants also respond with increased attention to behavioral aspects of animate beings. For example, newborns have been shown to be highly interested in modulated sounds similar to human voices (Vouloumanos et al. 2010). Furthermore, they start to analyze motion features in more detail and try to predict future movements rather early in life. This leads to an early discrimination between biological and non-biological motion (Simion et al. 2008). At 3–5 months of age, they are even able to detect the subtle differences between a walking and a running person when provided only with point-light displays that do not show the appearance of the person, but only the movement of important joints (Booth et al. 2002). Hence, young infants seem well prepared to identify animate beings based on static as well as dynamic information (Pauen and Träuble 2009). But this is not the entire story.

When another person establishes eye contact with the infant, this motivates him/her to closely follow what the communicative partner will do next (Csibra and Gergely 2009). Infants respond to contingency cues in social interactions (Deligianni et al. 2011). If the communicative partner ignores their signals or if he shows a considerable time delay in responding, this leads to loss of interest and/or irritation on part of the infant, as shown by studies with infants of depressed mothers who are not capable of establishing a normal interaction with their child (Tronick and Reck 2009). Children of such depressed mothers typically avoid eye contact with their mothers, experience physiological distress, and turn away if possible, thus showing exactly the opposite behavior as babies of healthy mothers who are used to sensitive and fine-tuned interactions (Reck et al. 2004).

Such findings suggest that human infants have specialized expectations regarding the communicative behavior of their interactive partner. If these expectations are not met, they later show abnormalities in development that not only have consequences for their social relations with others, but that may also affect cognitive development in general.

These findings illustrate an important point: Even if we identify innate patterns of responding to certain stimuli or events in the environment, this does not imply that experience is irrelevant for shaping development and knowledge acquisition. It always remains crucial! In early years, infants' experience is typically in control of the caregivers. Mother or father decides which stimuli their child is exposed to; it is usually adults who actively introduce the infant to different aspects of the environment. As suggested by various authors, evolution may have prepared us to modulate this process in specific ways (Herrmann et al. 2007; Csibra and Gergely 2009; Tomasello 2011).

8.3 Preparedness to Cultural Learning

Cultural learning implies the transmission of knowledge from one generation to the next. Cultural knowledge is not innate, nor does it have to be discovered by the individual. Rather, one person learns from another, thus allowing for rapid progress in cultural evolution. In adult life, these transmissions are predominantly bound to language in its spoken or written form. But this is not true for all ages. Even preverbal infants acquire cultural knowledge, thus raising the questions as to how they do it and whether there is some kind of "preparedness for cultural transmission" explaining the great success of humans in teaching their offspring from early on. Imitation and observational learning are important mechanisms in this context. Even newborns are capable of imitating the facial expressions of interactive partners (Meltzoff and Moore 1983). Later, they start to imitate actions that involve objects. Even if they are not yet capable of imitating a given action involving a tool, they already refer to corresponding observations when identifying functionally relevant attributes of a given artifact (Elsner and Pauen 2007; Träuble and Pauen 2007). In addition to infants' knowledge of space and causality (Träuble and Pauen 2010), their sensitivity to others' social signals has been shown to affect imitative actions (Brugger et al. 2007). Hence, imitation and observational learning both provide important means for explaining cultural transmission at a preverbal age.

Recently, Csibra and Gergely (2009) suggested that humans have evolved some kind of *natural pedagogy* to teach each other about the world. Whenever an adult wants to instruct an infant or child, she follows a specific series of actions, exploiting infants' interest in faces, movements, and the human voice: First, she looks at the child and greets him with a friendly modulated voice. Such cues serve to establish social contact and are called *ostensive cues*. Another ostensive cue is "contingent responding" (Delinguianni et al. 2011). Turn-taking in social communicative situations and responding promptly to social signals of the interactive partners indicate contact. Once social contact has been established, the adult starts to demonstrate the relevant action—often involving a toy or tool. She frequently checks whether the child is still following her explanation.

Infants pay increased attention to eye-information and movement (Farroni et al. 2003), and they have the innate ability to imitate (Meltzoff and Moore 1983). This allows them to follow the gaze of another person towards a target at a rather young age (e.g., Senju et al. 2008). Even though this capacity still undergoes substantial development during the first months of life, by 9–12 months of age, infants are already quite sophisticated eye-gaze followers and get involved in triadic interactions rather easily (Striano and Stahl 2005). As revealed by brain data from our own lab as well as from other labs, eye-gaze cues also affect the processing of object information. Objects that have previously been gaze-cued by another person appear more familiar to 4-month-olds than objects that have not been gaze-cued (Reid et al. 2004). Interestingly, this effect also depends on the emotional expression of the gaze-cueing person (Hoehl and Striano 2008; Hoehl et al. 2008) and the familiarity of the person guiding the attention (Hoehl et al. 2012).

Taken together, these findings show that infants are predisposed to process information about objects and persons in specialized ways. They start forming categories and identifying types of objects long before language comes into play, thereby considering those properties that seem most relevant for predicting the object's causal and functional relation with the environment as well as its relevance for the infants' survival. Core knowledge as well as knowledge acquisition mechanisms shape our development in close interaction with experience. Even though one might say that this does not reflect a fundamental difference between humans and other species, it seems noteworthy that only humans have evolved sophisticated mechanisms supporting cultural learning. Infants' early sensitivity to social signals such as eye-gaze information or vocal expressions helps them to identify instructional situations. Their ability to encode and interpret emotional expressions in the faces of other people and to relate this information to gaze-cued objects supports learning about the value of a given target. Infants' early emerging competencies to learn from observation and to imitate actions guarantee that cultural transmission takes place. As will be demonstrated in the following section, this preparedness may also be of high relevance for fear learning.

8.4 The Case of Preparedness for Fear Learning

It is pivotal to learn about threats in the environment. Learning which objects, situations, animals, or people might be dangerous is a task that is not restricted to early development. However, more than in the case of adults and older children, young infants' knowledge about the world is limited, whereas their curiosity and eagerness to explore novel things is often unlimited. Learning to avoid dangers like electric outlets, hot pans, poisonous snakes, etc., depending on the particular environment an infant or child is exposed to, is of great importance. One way to learn about these dangers is making one's own experiences. Through classic conditioning, it is possible to quickly and effectively associate a previously neutral stimulus (e.g., a hot pan) with the aversive experience of pain. This way of learning is effective, but potentially costly.

Another way of learning about threats is through observation of others. When observing others making bad experiences (e.g., an older sibling touching a hot pan), we can learn effectively about dangers and do so from a safe distance. Some authors have argued that very similar processes and neural structures may be involved in fear conditioning and observational fear learning, both relying heavily on the amygdala, a subcortical brain structure located within the medial temporal lobe (Olsson and Phelps 2007). Whereas the conditioned stimulus and the unconditioned stimulus are coupled in the lateral nucleus of the amygdala during classic conditioning, in observational fear learning, the emotional expression of another person (expressing fear, pain, or disgust) is thought to replace the unconditioned stimulus and is associated with the conditioned stimulus instead (Olsson and Phelps 2007).

By the end of the first year, human infants can learn to avoid unfamiliar objects by observing others' negative emotional reactions to them (Carver and Vaccaro 2007; Hertenstein and Campos 2004; Moses et al. 2001). In a study by Mumme and Fernald (2003), 12-month-old infants were shown televised reactions of an unfamiliar adult to a novel object. The adult directed either positive or disgusted facial and vocal expressions toward a target object while ignoring a distractor object. Infants then had the chance to interact with the objects themselves. Infants who saw an adult show disgust towards the target object approached and explored the object less and also showed more negative affect. Importantly, infants' reactions to the distractor object were not altered, suggesting that infants had learned about the specific value of the target object the adult had referred to instead of just adopting the adult's general affect through emotional contagion. In this study, as in other experiments, negative emotional expressions have proven more effective in eliciting avoidance than positive expressions in encouraging approach, suggesting an early emerging negativity bias in infants' emotional learning (Vaish et al. 2008).

However, infants do not only passively learn from others' experiences. By 9–12 months of age, infants start to actively search for emotional signals from adults when being exposed to an unfamiliar and ambiguous object or situation (Campos et al. 2003). They selectively turn towards adults who are responsive to them (Striano and Rochat 2000; Striano et al. 2006) and who are presumably knowledgeable in a given situation (Stenberg and Hagekull 2007). They also actively search for referential cues when it is not clear what exactly the emotional outburst of an adult refers to (Moses et al. 2001).

Observational fear learning is not unique to humans. Susan Mineka and colleagues observed in the 1980s that lab-reared rhesus monkeys are not afraid of snakes, although their wild-reared conspecifics, who were born in India, showed intensive fear-reactions to snakes (Cook and Mineka 1989; Mineka et al. 1984). Did all of the wild-reared monkeys have painful encounters with snakes in the wild? This assumption seems highly implausible given that all of them were still alive and well. Mineka and colleagues reasoned that, instead of classic conditioning, observational learning may have led to the wild monkeys' fear reactions to snakes.

In a first set of experiments, they showed lab-reared rhesus monkeys the fear reaction of wild-reared monkeys to snakes (Mineka et al. 1984). After only a few observations of their conspecifics' fearful behavior, the lab-reared monkeys also started to respond with fear to the sight of a snake. This suggests that, although fear of snakes is obviously not innate in rhesus monkeys, it can be learned relatively quickly. Can fear of harmless stimuli be learned just as easily? In another set of experiments, Mineka and colleagues tested this question using manipulated video recordings (Cook and Mineka 1989). First, they established that lab-reared rhesus monkeys could learn fear of snakes from video recordings as well as from life observations. Then they presented monkeys with manipulated recordings in which a wild conspecific reacted with fear to fear-irrelevant stimuli like flowers. In these cases, the monkeys did not adopt the fearful behavior and did not learn to fear harmless stimuli.

These observations in monkeys support the idea that fear learning is not unspecific, but that it is rather biased towards the learning of fear of phylogenetically "old" threats (Seligman 1971). In the 1970s, Seligman observed that phobias in humans seem to cluster around stimuli and situations that have posed a threat throughout primate evolution, e.g., fear of heights or narrow rooms. Specific phobias are characterized by exaggerated fear of certain stimuli, which can be very hindering and affect the daily lives of the concerned patients. Among the most common animal phobias are fear of snakes and fear of spiders (Mineka and Öhman 2002). Both types of animals have posed a threat to our ancestors in early primate evolution. It has therefore been suggested that humans (and other primates) have evolved a *preparedness* to quickly associate fear with these animals. In contrast, fear of modern threats to our well-being (e.g., electric outlets) are much less common, even though the immediate threat emanating from these stimuli is much greater than the threat of being bitten by a snake or a spider in many regions of the world today.

Is there evidence for selective fear learning in humans, apart from the occurrence of specific phobias? In fear conditioning experiments with healthy human adults, phylogenetically fear-relevant stimuli, such as snakes and spiders, are more readily associated with unconditioned aversive stimuli (e.g., mild electric shocks) than fear-irrelevant stimuli, such as flowers (see Öhman and Mineka 2001, for a review). Associations of fear-relevant stimuli with pain and fear seem to be more robust and less prone to extinction than associations with fear-irrelevant stimuli (Cook et al. 1986; Öhman and Mineka 2001). Furthermore, pictures of spiders and snakes are detected more quickly within large arrays of flower and mushroom distractor images than vice versa (Öhman et al. 2001). In the same study, it was found that fear of snakes or spiders specifically speeds up detection of the individually feared animals.

However, when working with adults, prior experiences with the stimulus categories can hardly be controlled for. It is possible that the participants have previously encountered spiders and snakes. More importantly, they have very likely perceived others' reactions or attitudes towards fear-relevant stimuli. Prior observational learning and cultural knowledge (spiders and snakes are considered disgusting animals in many cultures and societies) may affect adults' responses in conditioning experiments and their performance when detecting fear-relevant stimuli in complex displays. Testing individuals without corresponding experiences and knowledge is therefore crucial in order to test the hypothesis that selective fear learning indeed relies on evolved mechanisms. If there is an evolved preparedness for fear learning, there should be evidence for it in very young infants without prior experiences with or knowledge about fear-relevant animals.

As reviewed above, many studies have tested social referencing and social learning about the valence of novel objects in infants and toddlers. There is conflicting evidence regarding the selectivity of social referencing effects for objects that are ambiguous or potentially threatening. In some studies, un-threatening toys were used (Mumme and Fernald 2003), whereas in other studies, target objects were described as threatening (Carver and Vaccaro 2007; Klinnert et al. 1986). Sometimes larger effects were observed for ambiguous or threatening objects than for unthreatening objects (Gunnar and Stone 1984). In general, it seems that by the end of the first year, infants avoid novel objects associated with the negative emotional reaction of an adult even if these objects are not inherently threatening. However, is there evidence that fear-relevant stimuli are associated with fear *more easily* than fear-irrelevant and unambiguous stimuli in infants?

Only very few studies have investigated preparedness for fear-learning in preverbal infants. First, it was found that, just like adults, children and infants detect fear-relevant stimuli like snakes and spiders faster than other animals and objects (LoBue 2010; LoBue and DeLoache 2008, 2010). This attention bias may help humans to detect fear-relevant animals in complex visual environments from early on and may thus favor learning about these animals. But is there evidence that infants in fact associate spiders and snakes with fear more readily than other stimuli?

DeLoache and LoBue (2009) presented 7- to 18-month-old infants with films of snakes and films of fear-irrelevant exotic animals (e.g., giraffe, hippopotamus) that were shown simultaneously. Films were accompanied by tape recordings of a frightened or happy voice. Infants looked longer at videos of snakes when listening to a frightened voice than when listening to a happy voice. For fear-irrelevant exotic animals, there was no effect of the emotional tone of voice on infants' looking behavior. Thus, the presence of a fearful voice specifically increased infants' visual attention to videos of snakes. This finding is interesting, but it does not inform us about infants' learning about the animals, because animals and emotional voice were always presented simultaneously.

In another study, schematic fearful and happy faces were paired with fear-relevant and fear-irrelevant stimuli in a habituation paradigm (Rakison 2009). When 11-month-old females were habituated to a photo of a spider or a snake paired with a schematic fearful face they looked longer at test trials involving a novel spider or snake paired with a schematic happy face than at a novel fear-irrelevant stimulus (flower or mushroom) paired with a happy face. This suggests that female infants quickly and effectively associated the frightened expressions with the fear-relevant categories and were then surprised to see an exemplar of the fear-relevant category paired with a happy face. No such effect was found for boys and no effect was found when infants were first habituated to a fear-irrelevant stimulus paired with a schematic fearful face. The selectivity of the effect for females may relate to females' increased proclivity to develop specific phobias of snakes and spiders (Fredrikson et al. 1996), but it is unclear whether this gender difference is already stable in infancy, since it was not found in the study by DeLoache and LoBue (2009).

Recently, we conducted a study using event-related brain potentials (ERP) to test for selective fear-associations in pre-verbal infants (Hoehl and Pauen in preparation). We presented 9-month-old infants with photographs of real faces looking at a spider or a flower with either a fearful or a neutral facial expression. After each presentation of a face looking at a fear-relevant or fear-irrelevant target, the respective target was presented again alone and infants' brain responses were measured. Each infant received up to 160 trials presenting two different individuals (one male, one female) and several different flowers and spiders in alternation, and each infant received all of the four conditions in a within-subjects design.

We were particularly interested in the negative central (Nc) component of the infant ERP. The amplitude of this component has been related to the amount of attention infants direct toward a visual stimulus, with higher amplitude signaling increased attention (Reynolds and Richards 2005). Increased Nc amplitude has consistently been observed in response to fearful compared to happy faces in infants by seven months of age (De Haan et al. 2004; Grossmann et al. 2011; Nelson and De Haan 1996). We hypothesized that increased attention toward a target cued by a fearful face would be indicated by an increased Nc relative to targets cued by a neutral face.

We found significantly increased amplitude of the Nc component for spiders that were previously looked at by a fearful face relative to spiders cued by a neutral face. For flowers, similar to a previous study with unambiguous toys (Hoehl and Striano 2010), no effect of emotional expression on infants' responses to the stimuli was observed. Our findings suggest that 9-month-olds specifically direct increased attention toward spiders that were associated with a frightened facial expression. Although this temporary enhancement of attention may not suffice for building up long-lasting associations, it may facilitate the formation of stable fear-associations with fear-relevant targets.

It will be necessary to extend these findings to other fear-relevant categories such as snakes. Future studies should control for prior experiences with the stimulus categories, even when testing young infants, as previous experiences cannot completely be ruled out without at least asking parents about possible prior encounters (e.g., having spiders or snakes as pets or toys at home). It will be important to test whether infants' sensitivity to fear-relevant animals is restricted to phylogenetic threats or whether it extends to other stimuli that are equally unfamiliar and ambiguous to infants at this age. Furthermore, controlling for prior experiences allows for manipulating infants' experiences, e.g., by providing them with exemplars or realistic toy replicas representing fear-relevant categories. Possibly, infants are less prone to associate a fear-relevant target with fear when they have encountered the animal in the form of a harmless toy prior to testing.

Finally, it will be interesting to assess how preparedness for fear learning relates to the theory of natural pedagogy. This theory postulates that emotional signals directed at a target without communicative cues to an observer are interpreted as signaling individual attitudes of the model toward the target instead of generic knowledge about the general valence of this kind of stimulus (Gergely et al. 2007). Does this hold true when the target is a fear-relevant animal and the emotion directed at the target is fear or disgust? Or do infants "by default" associate fear or disgust with fear-relevant animals, even without the presence of communicative ostensive cues, and then expect other people to share this knowledge? These questions need to be clarified by future studies.

8.5 Conclusions and Future Directions

This chapter summarized existing work on infants' preparedness to learn about the world. More specifically, we looked at infants learning about fear-relevant objects in social contexts. As demonstrated by many studies, infants are born with quite sophisticated abilities to form object categories based on experiences with static and dynamic features. Even at a very young age, they can identify category membership of individual exemplars within milliseconds, provided that making a categorical contrast is relevant for infants' survival. Infants also come equipped with competencies to identify faces, to read emotional expressions, and to pay increased attention to gaze-information. These are necessary prerequisites for following the attentional focus of an interactive partner and to learn about objects and their emotional value via observing others. As has been shown by multiple behavioral as well as neurophysiological studies, object processing is influenced by (a) categorical identity of the target object, (b) gaze-information, and (c) the emotional facial expression of the interactive partner. Until today, the complex interplay between these different aspects is still not fully understood. When trying to explain the emergence of fear-responses towards stimuli such as snakes or spiders that provided a great threat to the survival of our ancestors, but have only little relevance today, it seems as if evolution has prepared us to learn about such stimuli from our conspecifics with increased sensitivity.

We conclude that infant research can inform us about which type of information has been relevant for the survival of our ancestors and that evolutionary theory can inform us about potentially relevant core knowledge and/or knowledge acquisition devices. By combining these different lines of research, much can be learned about our human nature and how we became what we are today.

References

Baillargeon, R., & Carey, S. (2012). Core knowledge and beyond: The acquisition of physical and numerical knowledge. In S. Pauen (Ed.), *Early childhood development and later outcome*. Cambridge: Cambridge University Press.

Behl-Chadha, G. (1996). Basic-level and superordinate-like categorical representations in early infancy. *Cognition, 60,* 105–141.

Biro, S., & Leslie, A. M. (2007). Infants' perception of goal-directed actions: Development through cue-based bootstrapping. *Developmental Science, 10*(3), 379–389.

Booth, A. E., Pinto, J., & Bertenthal, B. I. (2002). Perception of the symmetrical patterning of human gait by infants. Developmental Psychology, Vol 38(4), Jul 2002, 554-563. doi:http://psycnet.apa.org/doi/10.1037/0012-1649.38.4.554/10.1037/0012-1649.38.4.554.

Breyer, T., Etzelmüller, G., Fuchs, T., Schwarzkopf, G., Tanner, K., & Welker, M. (Eds.). (2013). *Interdisziplinäre Anthropologie: Leib – Geist – Kultur*. Heidelberg: Winter.

Briscoe, T. (2000). Grammatical acquisition: Inductive bias and coevolution of language and the language acquisition device. *Language, 76*(2), 245–296.

Brugger, A., Lariviere, L. A., Mumme, D. L., & Bushnell, E. W. (2007). Doing the right thing: Infants' selection of actions to imitate from observed event sequences. *Child Development, 78*(3), 806–824.

Campos, J. J., Thein, S., & Owen, D. (2003). A Darwinian legacy to understanding human infancy—Emotional expressions as behavior regulators. *Emotions Inside Out, 1000,* 110–134. doi:10.1196/annals.1280.040.

Carver, L. J., & Vaccaro, B. G. (2007). 12-month-old infants allocate increased neural resources to stimuli associated with negative adult emotion. *Developmental Psychology, 43*(1), 54–69. doi:2006-23020-005 [pii] 10.1037/0012-1649.43.1.54.

Cook, E. W., Hodes, R. L., & Lang, P. J. (1986). Preparedness and phobia: Effects of stimulus content on human visceral conditioning. *Journal of Abnormal Psychology, 95*(3), 195–207.

Cook, M., & Mineka, S. (1989). Observational conditioning of fear to fear-relevant versus fear-irrelevant stimuli in rhesus monkeys. *Journal of Abnormal Psychology, 98*(4), 448–459.

Csibra, G., & Gergely, G. (2009). Natural pedagogy. *Trends in Cognitive Sciences, 13*(4), 148–153. doi:S1364-6613(09)00047-3 [pii]10.1016/j.tics.2009.01.005.

De Haan, M., Belsky, J., Reid, V., Volein, A., & Johnson, M. H. (2004). Maternal personality and infants' neural and visual responsivity to facial expressions of emotion. *Journal of Child Psychology and Psychiatry, 45*(7), 1209–1218. doi:10.1111/j.1469-7610.2004.00320.x JCPP320 [pii].

Deligianni, F., Senju, A., Gergely, G., & Csibra, G. (2011). Automated gaze-contingent objects elicit orientation following in 8-month-old infants. *Developmental Psychology, 47*(6), 1499–1503.

Deloache, J. S., & LoBue, V. (2009). The narrow fellow in the grass: Human infants associate snakes and fear. *Developmental Science, 12*(1), 201–207. doi:DESC753 [pii]10.1111/j.1467-7687.2008.00753.x.

Elsner, B. (2007). Infants' imitation of goal-directed actions: The role of movements and action effects. *Acta Psychologica, 124*(1), 44–59. doi:S0001–6918(06)00122–3[pii]10.1016/j.actpsy.(2006.09.006).

Elsner, B., & Pauen, S. (2007). Social learning in artifact function in 12- and 15-month-olds. *European Journal of Developmental Psychology, 4*(1), 80–99.

Elsner, B., Jeschonek, S., & Pauen, S. (in press). Event-related potentials for 7-month-olds processing of animals and furniture items. *Developmental Cognitive Neuroscience, 3,* 53–60.

Farroni, T., Mansfield, E. M., Lai, C., & Johnson, M. H. (2003). Infants perceiving and acting on the eyes: Tests of an evolutionary hypothesis. *Journal of Experimental Child Psychology, 85*(3), 199–212.

Farroni, T., Johnson, M. H., Menon, E., Zulian, L., Faraguna, D., & Csibra, G. (2005). Newborns' preference for face-relevant stimuli: Effects of contrast polarity. *Proceedings of the National Academy of Sciences USA, 102*(47), 17245–17250.

Fredrikson, M., Annas, P., Fischer, H., & Wik, G. (1996). Gender and age differences in the prevalence of specific fears and phobias. *Behavioral Research and Therapy, 34*(1), 33–39. doi:0005796795000483 [pii].

Gelman, R., & Spelke, E. S. (1981). The development of thoughts about animate and inanimate objects: Implications for research in social cognition. In J. H. Flavell & L. Ross (Eds.), *The development of social cognition in children* (pp. 43–66). Cambridge: Cambridge University Press.

Gergely, G., Egyed, K., & Kiraly, I. (2007). On pedagogy. *Developmental Science, 10*(1), 139–146. doi:DESC576 [pii]10.1111/j.1467-7687.2007.00576.x.

Grossmann, T., Johnson, M. H., Vaish, A., Hughes, D. A., Quinque, D., Stoneking, M., & Friederici, A. D. (2011). Genetic and neural dissociation of individual responses to emotional expressions in human infants. *Developmental Cognitive Neuroscience, 1*(1), 57–66. doi:S1878–9293(10)00002–2 [pii] 10.1016/j.dcn.2010.07.001.

Gunnar, M. R., & Stone, C. (1984). The Effects of Positive Maternal Affect on Infant Responses to Pleasant, Ambiguous, and Fear-provoking Toys. Child Development, 55, 1231–1236.

Herrmann, E., Call, J., Lloreda, M., Hare, B., & Tomasello, M. (2007). Humans have evolved specialized skills of social cognition: The cultural intelligence hypothesis. *Science, 317,* 1360–1366.

Hertenstein, M. J., & Campos, J. J. (2004). The retention effects of an adult's emotional displays on infant behavior. *Child Development, 75*(2), 595–613. doi:10.1111/j.1467-8624.2004.00695.xCDEV695 [pii].

Hoehl, S. & Pauen, S. (in preparation). Do infants associate evolutionary fear-relevant stimuli with fearful facial expressions?

Hoehl, S., & Striano, T. (2008). Neural processing of eye gaze and threat-related emotional facial expressions in infancy. *Child Development, 79*(6), 1752–1760.

Hoehl, S., & Striano, T. (2010). The development of emotional face and eye gaze processing. *Developmental Science, 13*(6), 813–825.

Hoehl, S., Wiese, L., & Striano, T. (2008). Young infants' neural processing of objects is affected by eye gaze direction and emotional expression. *PLoS One, 3*(6), e2389. doi:10.1371/journal.pone.0002389.

Hoehl, S., Wahl, S., Michel, C., & Striano, T. (2012). Effects of eye gaze cues provided by the caregiver compared to a stranger on infants' object processing. *Developmental Cognitive Neuroscience, 2*(1), 81–89.

Jeschoneck, S., Marinovic, V., Höhl, S., Elsner, B., & Pauen, S. (2010). Do animals and furniture items elicit different brain responses in human infants? *Brain Development, 32*(10), 863–871. doi:10.1016/j.braindev. 2009.11.010.

Johnson, M., & Morton, M. H. (1991). Conspec and conlearn: A two-process theory of infant face recognition. *Psychological Review, 98,* 164–181.

Klinnert, M. D., Emde, R. N., Butterfield, P., & Campos, J. J. (1986). Social Referencing: The infant's use of emotional signals from a friendly adult with mother present. *Developmental Psychology, 22,* 427–432.

Kosugi, D., Ishida, H., & Fujita, K. (2003). 10-month-old infants' inference of invisible agent: Distinction in causality between object motion and human action. *Japanese Psychological Research, 45,* 15–24.

LoBue, V. (2010). And along came a spider: An attentional bias for the detection of spiders in young children and adults. *Journal of Experimental Child Psychology, 107*(1), 59–66.

LoBue, V., & DeLoache, J. S. (2008). Detecting the snake in the grass: Attention to fear-relevant stimuli by adults and young children. *Psychological Science, 19*(3), 284–289. doi:PSCI2081 [pii] 10.1111/j.1467-9280.2008.02081.x.

LoBue, V., & DeLoache, J. S. (2010). Superior detection of threat-relevant stimuli in infancy. *Developmental Science, 13*(1), 221–228. doi:DESC872[pii]10.1111/j.1467-7687.2009. 00872.x.

Mandler, J. M. (1992). How to build a baby II: Conceptual primitives. *Psychological Review, 99*(4), 587–604.

Mandler, J. M., & McDonough, L. (1994). Concept formation in infancy. *Cognitive Development, 8,* 291–318.

Markman, E. M. (1994). Constraints on early word learning in language acquisition. *Lingua, 92,* 199–227.

Meltzoff, A. N., & Moore, M. K. (1983). Newborn infants imitate adult facial gestures. *Child Development, 54,* 702–709.

Mineka, S., & Öhman, A. (2002). Phobias and preparedness: The selective, automatic, and encapsulated nature of fear. *Biological Psychiatry, 52*(10), 927–937. doi:S000632230 2016694 [pii].

Mineka, S., Davidson, M., Cook, M., & Keir, R. (1984). Observational conditioning of snake fear in rhesus monkeys. *Journal of Abnormal Psychology, 93*(4), 355–372.

Moses, L. J., Baldwin, D. A., Rosicky, J. G., & Tidball, G. (2001). Evidence for referential understanding in the emotions domain at twelve and eighteen months. *Child Development, 72*(3), 718–735.

Mumme, D. L., & Fernald, A. (2003). The infant as onlooker: Learning from emotional reactions observed in a television scenario. *Child Development, 74*(1), 221–237.

Öhman, A., & Mineka, S. (2001). Fears, phobias, and preparedness: Toward an evolved module of fear and fear learning. *Psychological Review, 108*(3), 483–522.

Nelson, C. A., & de Haan, M. (1996). Neural correlates of infants' visual responsiveness to facial expressions of emotion. *Developmental Psychobiology, 29*(7), 577–595. doi:10.1002/(SICI)1098–2302(199611)29:7<577::AID-DEV3>3.0.CO;2-R[pii]10.1002/(SICI)1098-2302(199611)29:7<577::AID-DEV3 >3.0.CO;2-R.

Öhman, A., Flykt, A., & Esteves, F. (2001). Emotion drives attention: Detecting the snake in the grass. [Research Support, Non-U.S. Gov't]. *Journal of Experimental Psychology: General, 130*(3), 466–478.

Olsson, A., & Phelps, E. A. (2007). Social learning of fear. *Nature Neuroscience, 10*(9), 1095–1102. doi:nn1968 [pii]10.1038/nn1968.

Pauen, S. (2000). Developing ontological categories of matter: Stable dimensions and changing concepts. In W. Schnotz, S. Vosniadou, & M. Carretero (Eds.), *New perspectives on conceptual change* (pp. 15–31). Amsterdam: Elsevier.

Pauen, S. (2002a). The global-to-basic level shift in infants' categorical thinking. evidence from a longitudinal study. *International Journal of Behavioral Development, 26*(6), 492–499.

Pauen, S. (2002b). Evidence for knowledge-based categorization in infancy. *Child Development, 73,* 116–1033.

Pauen, S., & Träuble, B. (2009). How 7-month-olds interpret ambiguous motion events: Category-specific reasoning in infancy. *Cognitive Psychology, 59,* 275–295.

Quinn, P. C., & Johnson, M. H. (2000). Global before basic object categorization in connectionist networks and 2-month-old infants. *Infancy*, *1*(1), 41–46.

Rakison, D. H. (2009). Does women's greater fear of snakes and spiders originate in infancy? *Evolution and Human Behavior, 30*(6), 439–444. doi:10.1016/j.evolhumbehav.2009.06.002.

Pauen, S., Wahl, S., & Höhl, S. (2011). The beginnings or infant categorization: An ERP study with 4- and 7-month-olds. Poster presented at the biennial meeting of the Society for Research in Child Development, Montreal, Canada.

Rakison, D., & Poulin-Dubois, D. (2001). Developmental origin of the animate-inanimate distinction. *Psychological Bulletin*, *127,* 209–228.

Reck, C., Hunt, A., Fuchs, T., Weiss, R., Noon, A., Moehler, E., Downing, G., Tronick, E. Z., & Mundt, C. (2004). Interactive regulation of affect in postpartum depressed mothers and their infants: An overview. *Psychopathology, 37*(6), 272–280.

Reid, V., Striano, T., Kaufmann, J., & Johnson, M. H. (2004). Eye-gaze cueing facilitates neural processing of objects in 4-month-olds. *NeuroReport, 15*(6), 2553–2555.

Reynolds, G. D., & Richards, J. E. (2005). Familiarization, attention, and recognition memory in infancy: An event-related potential and cortical source localization study. *Developmental Psychology, 41*(4), 598–615. doi:2005-08221-002 [pii] 10.1037/0012-1649.41.4.598.

Schmidt, M. F., Rakoczy, H., & Tomasello, M. (2012). Young children enforce social norms selectively depending on the violator's group affiliation. [Research Support, Non-U.S. Gov't]. *Cognition, 124*(3), 325–333. doi:10.1016/j.cognition.2012.06.004.

Seligman, M. E. (1971). Phobias and preparedness. *Behavior Therapy, 2,* 307–320.

Senju, A., Csibra, G., & Johnson, M. H. (2008). Understanding the referential nature of looking: Infants' preference for object-directed gaze. *Cognition, 108,* 303–319.

Simion, F., Regolin, L., & Bulf, H. (2008). A predisposition for biological motion in the newborn baby. *Proceedings of the National Academy of Sciences of the United States of America, 105*(2), 809–813.

Spelke, E. S. (1990). Principles of object perception. Cognitive Science, 14, 29–36.

Spelke, E. S. (2011). Core systems and the growth of human knowledge: Natural geometry. In A. M. Battro, S. Dehaene, & W. J. Singer (Eds.), The proceedings of the working group on human neuroplasticity and education, vol. 117. *Human neuroplasticity and education* (pp. 73–99). Vatican City: The Pontifical Academy of Sciences.

Spelke, E. S., & Kinzler, K. D. (2007). Core knowledge. *Developmental Science, 10*(1), 89–96.

Spelke, E. S., Phillips, A. T., & Woodward, A. L. (1995). Infants' knowledge of object motion and human action. In D. Sperber, D. Premack, & A. Premack (Eds.), *Causal cognition: A multidisciplinary debate* (pp. 44–78). Oxford: University Press.

Stenberg, G., & Hagekull, B. (2007). Infant looking behavior in ambiguous situations: Social referencing or attachment behavior? *Infancy, 11*(2), 111–129.

Striano, T., & Rochat, P. (2000). Emergence of selective social referencing in infancy. *Infancy, 1*(2), 253–264.

Striano, T., & Stahl, D. (2005). Sensitivity to early triadic interaction in infancy. *Developmental Science, 8*(4), 333–343.

Striano, T., Vaish, A., & Henning, A. (2006). Selective looking by 12-month-olds to a temporally contingent partner. *Interaction Studies, 7,* 233–250.

Tomasello, M. (2011). Human culture in evolutionary perspective. In M. Gelfand, C. Chiu, & Y. Hong (Eds.), *Advances in culture and psychology* (vol. 1, pp. 5–51). Oxford: University Press.

Träuble, B., & Pauen, S. (2007). The role of functional information for infant categorization. *Cognition, 105*(2), 362–379. doi:S0010-0277(06)00211-3[pii]10.1016/j.cognition.2006. 10.003.

Träuble, B., & Pauen, S. (2010). Cause or effect—What matters? How 12-month-old infants learn to categorize artefacts. *British Journal of Developmental Psychology, 29*(3), 357–374. doi:10.1348/02615 1009X479547.

Tronick, E., & Reck, C. (2009). Infants of depressed mothers. *Harvard Review Psychiatry, 17*(2), 147–156.

Vaish, A., Grossmann, T., & Woodward, A. (2008). Not all emotions are created equal: The negativity bias in social-emotional development. *Psychological Bulletin, 134*(3), 383–403. doi:2008–04614-002 [pii]10.1037/0033-2909.134.3.383.

Vouloumanos, A., Hauer, M. D., Werker, J. F., & Martin, A. (2010). The tuning of neonates preference for speech. *Child Development, 81*(2), 517–527.

Wynn, K. (1995). Origins of numerical knowledge. *Mathematical Cognition, 1,* 35–60.

Chapter 9
Depression as an Adaptation: The Infection–Defense Hypothesis and Cytokine Mechanisms

Alexander Neumann and Sven Walter

9.1 Introduction

Major depression is a surprisingly common, unipolar mood disorder, characterized by episodes of extraordinary sadness and dejection. Although its frequency varies across cultures, it is rather ubiquitous, with lifetime prevalence rates usually in the range of 8–12 % (Andrade et al. 2003). According to the WHO, depression affects more than 350 million people of all ages worldwide, making it one of the leading causes of disability, and the WHO predicts that by the year 2020, depression will be the second most common chronic health issue in the world. According to the Diagnostic and Statistical Manual of Mental Disorders (DSM)-IV-TR, a diagnosis requires a major depressive episode, initially or recurring, during which subjects have to exhibit at least five of nine symptoms within a period of 2 weeks. The symptoms must include a depressed mood or a loss of interest or pleasure for nearly the whole day on most days as well as other possible indicators such as physiological (e.g., weight changes, sleep disturbances, psychomotor agitation, or retardation), behavioral (e.g., fatigue), or psychological (e.g., excessive guilt, suicidal tendencies, feelings of worthlessness, concentration disorders, indecisiveness) ones. Major depressive episodes can last from a few weeks to years and often recur (e.g., Spijker et al. 2002).

A variety of proximal biological, genetic and neurochemical, environmental, and psychological causal influences on depression have been found. Family, twin, and adoption studies suggest a moderate genetic influence (e.g., Nes et al. 2013): Compared to nonbiological siblings, for instance, biological relatives are eight times more likely to develop unipolar depression (Wender et al. 1986). Several genes,

---

A. Neumann (✉)
Maastricht University, Maastricht, Netherlands
e-mail: alex.neumann@student.maastrichtuniversity.nl

S. Walter
University of Osnabrück, Osnabrück, Germany
e-mail: mail@svenwalter.eu

T. Breyer (ed.), *Epistemological Dimensions of Evolutionary Psychology,*
DOI 10.1007/978-1-4939-1387-9_9

gene–gene, and gene–environment interactions have been said to increase the risk of experiencing depression (e.g., Kim-Cohen and Gold 2009). One gene polymorphism that possibly contributes to this risk increase lies in the promoter region of a gene coding for the serotonin transporter (5-HTT) responsible for removing serotonin in the synaptic cleft. That promoter region comes in short (s) and long (l) variants, where long alleles result in the production of more transporters than short ones. The higher the number of stressful live events an individual experiences, the likelier are ss carriers to experience depression compared to sl carriers, while the ll carriers are least likely to experience depression (Caspi et al. 2003); in all cases, however, the likelihood increases with the number of stressful life events, i.e., the s allele only predisposes individuals to be more vulnerable to stress. Other etiologically relevant factors include an elevated response of the hypothalamic–pituitary–adrenal (HPA) axis to stress, resulting in elevated cortisol levels (e.g., Gotlib et al. 2008),[1] and a lowered activity of the hypothalamic–pituitary–thyroid axis that influences metabolism by secreting thyroid hormones (e.g., Garlow and Nemeroff 2003). Among the suggested neurochemical factors, reduced serotonin functionality plays an important role, as well as abnormal norepinephrine and dopamine levels (e.g., Southwick et al. 2005), and some brain regions—including the prefrontal and anterior cingulate cortex, the hippocampus, and the amygdala—respond differently in depressed subjects (e.g., Davidson et al. 2002). Other biological and environmental factors associated with depression include disturbances in sleep and circadian rhythms (e.g., Germain and Kupfer 2008) and changes in sunlight and seasons (e.g., Lambert et al. 2002).

Psychologically, most depressive episodes are preceded by major stressful life events. In particular, relationship and interpersonal loss stressors have been said to play a crucial role, with early adversity possibly moderating the response. Stressful *dependent* life events that at least partly hinge on the subject's own choices better predict the onset of depression than independent ones that would have occurred regardless of the subject's behavior or personality (Hammen 2005). A tendency to attribute negative events to internal, stable, and global causes (as opposed to external, unstable, and specific causes) is also characteristic of depressive subjects (e.g., Sweeney et al. 1986), as are psychological traits like perfectionism (e.g., Shafran and Mansell 2001) and introversion (e.g., Janowsky 2001).

Interesting and important as these insights into the proximal mechanisms of depression and more distal explanations of the ontogenesis of depression are, they provide answers only to two of Tinbergen's (1963) famous four questions, viz., *What are the mechanisms of a trait?* and *How does it develop?,* but leave unaddressed the *ultimate causes* of depression, i.e., Tinbergen's other two questions: *What is the function of a trait?* and *What is its phylogenetic origin?* They fail to explain why, on the one hand, only a minority of people develops depression when facing the same challenges and why, on the other hand, the minority is still so large, given the apparently maladaptive effects. *Evolutionary accounts of depression* ar-

---

[1] A higher HPA axis reactivity is associated with the ss polymorphism, which suggests a possible mechanism for its risk effects.

gue that depression is an adaptation that serves some evolutionarily important function or other. In this chapter, we will be primarily concerned with the hypothesis that depression is an adaptation for fighting and preventing infections, evaluating the suggestion that this function could be achieved by cytokines, messengers of the immune system.

Section 9.1 argues that the mere fact that depression has mostly harmful effects and is found only in a minority of people does not show that it cannot be an adaptation. Section 9.2 surveys some evolutionary accounts of depression. Section 9.3 introduces the infection–defense hypothesis and argues that the evidence for it is at best incomplete. Section 9.4 shows that supplementing the infection–defense hypothesis with the cytokine hypothesis—i.e., with an account that bridges the gap between ultimate and proximal causes by detailing exactly how the alleged evolutionary function of depression may actually be achieved at the level of neurobiological mechanisms—results in a plausible overall account. Section 9.5 sums up the results and makes some concluding remarks.

9.2 Can Depression Be an Adaptation?

Although there is room for cultural variation and environmental influences, depression is found ubiquitously in all cultures and in nonhuman animals as well.[2] Individuals having to cope with the symptoms of depression have thus been subjected to selection forces for quite some time. Why, then, did natural selection not eliminate depression, given its detrimental effects that impair functioning and so apparently reduce one's fitness, in the worst case leading to suicide? And if for some reason depression was indeed selected for, then why is it not even more prevalent?

Notoriously, however, evolution and natural selection are much more complex than these simple questions suggest. The mere fact, for instance, that depression sometimes leads to suicide does not entail that it is not an adaptation, i.e., a trait we today have because it was adaptive in our ancestors' evolutionary environment. Although 30% of all suicides occur during a depressive episode, suicide is still comparatively rare (and apparently influenced by many other factors: In Iran, for example, the prevalence of depression is high, but that of suicide is low; Keedwell 2008, p. 10), and while most suicide *attempts* occur between age 18 and 24, most *successful* suicides are committed by those aged 65 and older (Butcher et al. 2013, Chap. 7). That some depressed subjects commit suicide is thus no reason to think that depression cannot be an adaptation: Since those few most likely to commit suicide are most likely to already have offspring, suicide is shielded from natural

---

[2] For instance, primates show reactions comparable to depressive behavior in humans when separated from socially important individuals (see, e.g., Comer 2013, Chap. 8), and they also exhibit "learned helplessness" (i.e., upon learning that they cannot change the outcome of an unpleasant event, they stop trying to do so), a feature often associated with depression (e.g., Abrahamson et al. 1978).

selection in a way similar to Alzheimer and other old age diseases which are not selected against because they do not affect reproductive success.

Moreover, the mere fact that a state is distressing and (thus) potentially disadvantageous in the short run does not mean it is necessarily detrimental in the long run, all things considered (see, e.g., Keedwell 2008): For instance, feeling sick and vomiting after food poisoning or during the first months of pregnancy is certainly distressing and temporarily reduces functioning, but it is still adaptive and arguably an adaptation for protecting the individual or the fetus (e.g., Profet 1992). Similarly, the sadness and other adversary effects of depression are unpleasant and maybe even temporarily maladaptive, but precisely this unpleasantness may be an important motivation to avoid or better cope with comparable situations in the future (in particular, given that these are most likely dependent life events over which the subject has some control).

Another uncompelling argument is that if depression was selected for, it should eventually be reliably triggered in appropriate circumstances in virtually everyone, and that since this is not the case, it is not an adaptation (Nettle 2004). Depression is obviously not universally and reliably triggered: While a liability to depression is to some extent heritable and many people indeed suffer from depression at some point in their life, most do not, and while the majority of depressive episodes are preceded by stressful major life events, most people do not respond to such events with depression. However, since there is no reason why adaptations or adaptive traits should be driven to universal fixation, this does not show that depression is not an adaptation or fails to be adaptive. First of all, the adaptive value of a trait can be frequency dependent, i.e., a trait may be adaptive as long as it is a “minority strategy,” but become evolutionarily neutral or even maladaptive once present with a certain frequency.[3] According to the *social navigation hypothesis* (Sect. 9.2.3), for instance, depression serves to elicit supportive behavior. If this is true, then depression may be selected for as long as there are enough nondepressed helpers, but become maladaptive once its adverse consequences are no longer counterbalanced because too many members of the population are depressed and thus unable to provide help. Second, given that depression vulnerability is influenced by pleiotropic genes, it may be adaptive and yet fail to be driven to universal fixation because other traits associated with these genes are selected against. The ss polymorphism of 5-HTTLPR, for example, is not only associated with depression, but also with aggression (e.g., Gonda et al. 2009). Thus, if aggressiveness is both selected for and against in different contexts, resulting in a balanced selection in which it is advantageous to have different polymorphisms, this will also affect depression: If, say, aggression is favored in war times, but not in times of peace, depression vulnerability will vary, too, regardless of the potential adaptive value of depression.

And finally, adaptations need not necessarily be currently adaptive (e.g., Tooby and Cosmides 1990). That we do many things detrimental to survival and

---

[3] As long as the chances of encountering other psychopaths are low, for instance, psychopathy can be an adaptive social strategy, but as more and more come to adopt this strategy, the negative effects start to outweigh its benefits due to the increased risk of costly interactions with other psychopaths (see, e.g., Mealey 1995).

reproduction (we use contraceptives, consume unhealthy doses of fatty food, and blow ourselves up in the middle of crowded market places), while at the same time refraining from activities that would be conducive to survival (eat healthy, exercise) or boost our reproductive success (donate sperm or eggs to cyrobanks) therefore does not show that our behavior is not the result of cognitive adaptations. Adaptations are present today because of their *past* selective advantage and our past environment differed notably from the modern metropolis in which we have been living for little more than 100 years in unprecedented large groups, consume fast food, and use contraceptives. Compared to this, our ancestors spent an unimaginably long time (roughly a period spanning 1.8 mya to 10,000 years ago) in small nomadic hunter–gatherer bands. The cognitive mechanisms produced by natural selection are adaptations for producing adaptive behavior in these circumstances, not for playing chess, navigating through lower Manhattan during rush hour, keeping an ideal weight in an environment full of fast food restaurants, or coping with all the other challenges of a globalized, fast-moving, technology-based, and increasingly anonymous world (Cosmides and Tooby 1994), which is why current maladaptivity does not indicate the absence of an adaptation.

9.3 What May Depression Be an Adaptation for?

Assuming that there is no good reason to deny that depression can be an adaptation, the next question is: What may it be an adaptation for, i.e., what may be the evolutionary function of depression?

9.3.1 Recovery After a Depressive Episode

Buist-Bouwman et al. (2004) compared a cohort of 164 subjects who had experienced a major depressive episode and a control group of 4178 subjects who had not during a 2-year period (between 1996 and 1998) with regard to various levels of functioning *before,during* and *after* the episode (according to the *Short-Form-36 Health Survey* which is used to measure a subject’s level of physical functioning, physical role functioning, vitality, pain, psychological health, psychological role functioning, social functioning, and general health). The depressed cohort fared worse than the controls at all times and with regard to all factors, except for psychological role functioning (problems related to job, family, and other daily activities as a consequence of emotional problems). While psychological role functioning in the cohort became worse during the depressive period, it improved again afterwards (compared to pre-episode scores) and was eventually no longer significantly different from the respective scores of controls. And while the psychological health scores (feelings of depression or nervousness) of the cohort never quite reached those of the controls and initially declined during the depressive episode, they afterwards on average also exceeded pre-episode scores. Depressive episodes thus worsen functioning in many areas, but on average seem to enhance psychological health and role functioning. However, in order for this to be taken as evidence that depression serves to improve psychological health and role functioning, one would need to know exactly why this warranted the significant worsening of general health, vitality, and physical functioning in our ancestors' social and natural environment (e.g., Richardson 2007).

9.3.2 Analytical Rumination

Some indication as to why improved psychological competences might have outweighed deterioration in other areas comes from the observation that depression is often caused by analytically difficult problems whose solution requires attention and analytical skills, skills which benefit from the rumination and the intense focus on a problem characteristic of depressive episodes (Andrews and Thomson 2009). The dependent stressful social life events that frequently trigger depression are typically multifaceted and require complex solutions that involve intricate tradeoffs, taking into account multifarious implications for one's relationship to the other members of one's social environment. If their tendency for rumination helps depressive subjects to analytically solve these problems, it becomes clear why depression is more typically associated with dependent problems: Ruminating over a problem that arose for reasons independent of one's own behavior or personality is a waste of resources.

There is some introspective evidence from reports of depressed subjects for the analytical rumination hypothesis (e.g., Lyubomirsky et al. 1999; see Andrews and Thomson 2009, p. 626). Andrews et al. (2007) sought empirical evidence by testing whether having to complete the *Raven's Advanced Progressive Matrices* (RAPM) task—an analytically difficult task invoking working memory and many other cognitive abilities—induces a depressive affect that in turn facilitates solving the RAPM. After their initial depressive affect was measured, participants were given some practice questions, then their depressive affect was measured again, and finally they had to complete the RAPM. In subjects with an initially low depressive affect, depressive affect increased and a stronger increase corresponded to a better RAPM performance; in contrast, for those with an initially high depressive affect, depressed affect decreased. According to Andrews and his colleagues, this is in line with the analytical rumination hypothesis: The analytically complex task triggered a depression-like state that focused the participant's resources on the RAPM, thus enhancing performance, while those who already were in a depressed state beforehand may have been focused on the problem causing their depression, so that the forced focus on another problem decreased depressive affect.[4]

---

[4] Note that as evidence for the analytical rumination hypothesis, the demonstration that depressive affect enhances performance on artificial tasks like the RAPM is less convincing than possible evidence that the rumination caused by depression helps to solve the kind of ecologically valid

One of the biological mechanisms through which depressive affect may boost performance on analytically difficult tasks is via the left ventrolateral prefrontal cortex (VLPFC), a brain area known to be involved in attentional control. There is evidence that the VLPFC is more active in depressed subjects and that increased activity in the VLPFC is indeed correlated with better RAPM performance (Gray et al. 2003; see Andrews and Thomson 2009, pp. 630–632). Another way in which depressive affect may enhance problem solving performance is via anhedonia: By making other activities less pleasurable, anhedonia makes it easier to focus on the depressogenic problem, and by making subjects weigh immediate rewards less, it supports a more unbiased rumination style (Watson and Andrews 2002). Similarly, the preference to evade social interaction during depressive episodes can help avoid distracting stimuli. Even anorexia and psychomotor retardation could have a similar function, given that oral or buccal activity potentially disrupts stimuli processing (e.g., Jacobs and Fornal 1999) and that reduced motor activity frees cognitive resources for rumination. Finally, Andrews and Thomson argue that a disturbed sleep pattern, specifically insomnia, is an adaptive feature of depression because "sleeping is not conducive to analytical processing" (2009, p. 632). This, however, can be contested. By means of its role in associating and integrating memories, for instance, sleep can improve performance in analytically complex problems like a number reduction task (e.g., Wagner et al. 2004). From this perspective, a lack of sleep would actually seem to be disadvantageous. On the other hand, however, negative memories appear to be more resilient to negative effects of sleep deprivation than positive memories (Walker 2009) and the resulting bias towards negative memories may serve to elicit rumination and a focus on the problem. Sleep deprivation may thus indirectly promote rumination while at the same time worsening analytical skills and making valuable insights less likely. In any case, depression is associated with both insomnia *and hypersomnia*, the latter being one of the defining features of atypical depression (see, e.g., Angst et al. 2002).

9.3.3 Social Navigation

The social navigation hypothesis (Watson and Andrews 2002) suggests that the function of depression is to elicit supportive behavior from others, both by signaling a need for help and by extortion: Depression signals that one is in serious trouble and in need of support, and its grave and adverse effects ensure that the signal is "honest" (Zahavi and Zahavi 1997)—the most drastic measures of "honest signaling" being suicide attempts. In addition to informing those who are already prepared to help, depression may also elicit help from those potentially unwilling to provide it: Social networks rely on "interacting reciprocal exchange contracts" (Watson and Andrews 2002, p. 8), and the fact that depressed subjects are unable to

---

problems that are known to trigger depressive episodes (for a review of allegedly relevant empirical evidence, see Andrews and Thomson 2009).

fulfill their social obligations extorts others to help in an attempt to alleviate pressures on the group, including themselves. Depressive behavior may also function to lessen the risks of social rank or power loss or social exclusion. According to the *social risk hypothesis* (Allen and Badcock 2006), depression serves to avoid risks and, thus, potential further damage, by triggering an increased sensitivity in perceiving social risks, signaling the need for support and making risky behavior unlikelier. The compliant behavior and the longing for solitude characteristic of depression, for example, function as de-escalation strategies: In fights or other social confrontations, the loser can avoid potential further damage, both physical and social, by showing subordinate behavior, withdrawing from the group, etc. Several biological mechanisms for how risk sensitivity and de-escalation behavior are achieved have been suggested (Allen and Badcock 2006). The amygdala, for instance, has been shown to be hyperactive during depression and is associated with evaluating threats and complex social judgments so that it may be involved in the detection of potential risks (e.g., Vorhold et al. 2007). Once risks are detected, they have to be avoided. A hypoactivation of the prefrontal cortex leads to less accurate predictions of future rewards and, thus, to a stronger focus on immediate rewards, which renders risky future goals less attractive (see Allen and Badcock 2006). While this seems to be at odds with Andrews and Thomson's claim that depression improves attentional control by an increased activity in the VLPFC, both claims may be compatible if risk aversion effects do not arise ventrolaterally, but from other parts of the prefrontal cortex.[5]

According to the evolutionary accounts introduced in this section, the cognitive and behavioral symptoms of depression serve to improve psychological functioning, the detection and solution of complex, typically social problems, to avoid further, typically also social, perils and damage or to elicit support from one's social community. Another interesting hypothesis is that the function of depression is to fight or prevent infections.

9.3.4 The Infection Hypothesis

It has been suggested that depression is an adaptation for fighting infections, avoiding the contamination of others, and protecting the subject from further pathogens when already infected or when in circumstances in which there is a heightened risk of infection, including psychologically stressful situations that potentially impair immune functioning (Anders et al. 2013; Kinney and Tanaka 2009). What renders this infection–defense hypothesis plausible is the apparent similarity between the symptoms of depression and general sickness (e.g., Schiepers et al. 2005). Sick people show signs of anorexia, weight loss, fatigue, somnolence, retardation, decreased locomotor activity, indifference, avolition, a diminished libido, impaired

---

[5] For example, damage to the ventromedial cortex seems to lead to abnormal performance on tasks requiring the evaluation of future consequences (Bechara et al. 1994).

cognitive abilities, dysphoria, or anhedonia, and the function of those symptoms is generally thought to be the conservation of energy expenditure for an adaptive febrile response and immunological activity, i.e., the allocation of the resources still available to the fighting of pathogens (see, e.g., Dantzer et al. 2007; Rabin 1999, Chap. 3). Since depression is characterized by the same symptoms, it is only natural to assume that they also serve a similar function. A reduced interest in social and sexual activity, psychomotor retardation, fatigue and hypersomnia, for instance, allow depressed subjects to allocate more energy to their immunological activity, a lack of appetite reduces the intake of food, and so the risk of infection (food being a major source of parasites), indecisiveness and a lack of confidence render injuries less likely and help the immune system by reducing psychological stress, etc. Moreover, while a negative body language, anhedonia, irritability or angry outbursts may signal a need for help (Sect. 9.2.3), they also create an aura of unapproachability that makes others less likely to want to get in contact, thereby making exposure to (further) pathogens less likely and at the same time protecting kin, friends, and other members of the social group from a potential contamination.

While the infection–defense hypothesis is an interesting alternative (or supplement; see Sect. 9.5) to the evolutionary accounts discussed in Sect. 9.2, it is not without problems. For instance, although the hypersomnia associated with atypical depression conserves energy and reduces social contacts, the insomnia that is extremely prevalent in major depression obviously does not. In an effort to explain why insomnia is so ubiquitous in depressed subjects, given its apparent disadvantages (compared to hypersomnia) with regard to disease avoidance, Kinney and Tanaka argue that insomnia is not a symptom of depression, but rather a cause or protective hypervigilant behavior (2009, p. 564). But as we will see, the causal relation, i.e., what exactly causes what, is not easy to establish.

Kinney and Tanaka also argue that the kind of agitation characteristic of depression and the excessive restlessness associated with it leads to an increased motor activity, which is important for immune functioning and counterbalances the negative effects of lethargy, while at the same time the subject does not actually have to go out and risk physical injury and social contact and, thus, infections and contaminations (Kinney and Tanaka 2009, p. 563; Anders et al. 2013). However, restlessness is also characteristic of normal, nonpathological anxiety (e.g., Gross and Hen 2004) and there it is quite unlikely to serve the prevention or fighting of infections.

The hypothesis is that depression makes exposure to pathogens less likely and restricts the spread of an infection by inducing social avoidance and isolation. But is depression really disadvantageous for social relationships? According to the social navigation hypothesis, for instance, it actually fosters helping behavior among members of a social group, and someone who is infected needs social support regarding protection, food, and the provision of other material resources.[6] And according to the social risk hypothesis, depression helps de-escalate potentially harmful social encounters, thereby again securing social harmony rather than having a negative

---

[6] This may be one of the reasons why people in a bad mood behave less generously and less altruistically (Kirchsteiger et al. 2006).

impact on social relations. Moreover, it is far from clear that isolation is adaptive for an infected individual, all things considered. The psychological stress induced by isolation may do more harm to the functioning of the immune system than good, for example by increasing the susceptibility to the common cold (e.g., Cohen et al. 1991). Moreover, there is evidence that within systems of close social relationships, depression actually lessens the level of aggression of those interacting with a depressed subject and thus fosters improved social relationships (e.g., Sheeber et al. 2001).

Finally, while food avoidance decreases the amount of contaminated food and the intake of nutrients necessary for the proliferation of pathogens (e.g., Yirmiya et al. 1999), it clearly also has negative effects on immune functioning. And even if it is true that one can minimize the risk of contamination by restricting food intake to certain kinds of food, say a diet high in carbohydrates (Kinney and Tanaka 2009, p. 564), then why did natural selection not favor this more selective strategy? And how are we to make evolutionary sense of the fact that while most subtypes of depression are characterized by a loss of appetite, atypical and seasonal versions are actually characterized by an increased appetite (see, e.g., Parker et al. 2002)?

Empirical evidence for the infection–defense hypothesis comes primarily from correlation studies. For example, infectious diseases like herpes simplex, West Nile encephalitis, HIV, and hepatitis C as well as other medical conditions that impair the functioning of the immune system, like various types of cancer, cardiovascular diseases, autoimmune disorders, omega-3 fatty acid deficiency, chronic pain, psychological stress, and chronic sleep deprivation, are associated with higher rates of depression, and depression indeed occurs more often in populations with low economic status and higher social stress, which are more prone to immune defects (Kinney and Tanaka 2009). The infection–defense hypothesis can also explain why premenstrual syndrome, premenstrual dysphoric disorder, postpartum depression, and seasonal affective disorder often go along with depressive episodes, given that under these circumstances, the immune system is clearly more vulnerable (Doyle et al. 2007; see Kinney and Tanaka 2009, pp. 564–565). However, such correlations between diseases and immunological vulnerability on the one hand and depression on the other do not show that the infection–defense hypothesis is not just another of those *prima vista* plausible, but eventually unjustifiable Panglossian "just-so stories," of which evolutionary psychology has so often been accused (e.g., Gould 1997). The problem is that these correlations alone do not tell us whether depression is indeed a consequence of infections, or even a cause, or maybe just a causally unconnected byproduct. Psychological stress, for example, could easily be a confounding variable: If psychological stress both induces depression and simultaneously, but independently, makes the body more susceptible to infections, this would lead to the observed correlations, although the function of depression could then of course not be what the infection–defense hypothesis says it is. Another possible confounding variable is childhood maltreatment. Rather than depression being a response to inflammation, both inflammation and depression may be consequences of childhood maltreatment. For instance, Danese et al. (2008) found a significant correlation between elevated inflammation markers (high-sensitivity C-reactive protein, fibrinogen, and leukocytes) and depression, controlling for medical records of recurring depression, socioeconomic status in childhood and adulthood, cardiovascular risk factors, smoking, and medications. However, once childhood maltreatment was taken into account, the correlation between elevated inflammation markers and depression became insignificant: Those who were depressed and maltreated were twice as likely to have clinically relevant elevated high-sensitivity C-reactive protein levels as the nondepressed controls.

What is therefore needed is evidence for an appropriate causal connection between infections and depressive episodes. As we will see, such evidence comes from the cytokine hypothesis, which can therefore supplement the ultimate account of depression offered by the infection–defense hypothesis with a proximal account at the level of neurobiological mechanisms that shows how infections manage to elicit depressive behavior.

9.4 The Cytokine Hypothesis

According to the cytokine hypothesis, pro-inflammatory cytokines released by the peripheral and perhaps the central immune system causally influence the development of depressive symptoms (Schiepers et al. 2005). As indicated at the end of Sect. 9.3, the cytokine hypothesis can supplement and support the infection–defense hypothesis by providing valuable insights into the neurobiological mechanisms, by means of which depression manages to fulfill its alleged evolutionary function. Unfortunately, some neurobiological details are unavoidable to fully understand the interplay between these two approaches.[7]

9.4.1 Cytokines and their Biological Influence on Brain and Behavior

Cytokines are signaling molecules secreted by leukocytes and used to regulate immune activity. They either increase inflammation, as in the case of interleukin (IL)-1↑, IL-6↑, interferon (IFN)-γ↑, and tumor necrosis factor (TNF)-α↑, or decrease inflammation, as in the case of IL-4↓, IL-10↓, and IL-13↓.[8] The neuroendocrine system can influence cytokine levels. The hypothalamus is able to increase the amount of corticosteroids like cortisol through the HPA axis as part of the stress response. This response has both immune enhancing and suppressing effects, which can be adaptive or pathological (e.g., Dhabhar 2009); psychological stress specifically can also suppress inflammatory activity. For instance, in one study those participants

---

[7] Unless otherwise noted, the following details about the workings of cytokines are taken from Schiepers et al. (2005).

[8] "↑" indicates that the cytokine is pro-inflammatory, "↓" indicates that it is anti-inflammatory.

who showed a greater cortisol response to mental stress tasks also displayed a lower IL-6↑ and IL-1↑ reactivity (Kunz-Ebrecht et al. 2003). The autonomic nervous system can also regulate immune responses.

The blood–brain barrier generally does not permit the crossing of relatively large hydrophilic molecules like cytokines. However, more permeable sites exist and cytokines can enter the brain via circumventricular sites as well. Trauma and diseases like multiple sclerosis can make the blood–brain barrier more permeable to cytokines, too, and there are also special proteins that carry cytokines across the blood–brain barrier. Moreover, the brain itself produces cytokines, for example when encountering an infection in case of brain damage. Finally, cytokines can also influence the brain "from a distance": Cerebral vascular endothelial cells can bind with cytokines, release nitric oxide, and prostaglandins, which in turn affect the brain, and the vagus nerve relays information of cytokine releases in immune organs to the nucleus tractus solitarius.

While cytokines have been linked to the development of the neural system, IL-1↑ can also damage neurons. Cytokines also influence neurotransmitter activity: IL-1↑, IL-2↑ IL-6↑ are able to increase serotonin, IL-1↑ additionally noradrenaline and dopamine, whereas IL-6↑ can lower dopamine levels. Moreover, cytokines can activate the HPA axis, albeit only to a certain degree, since the activation is countered by an inhibitory feedback mechanism (Sect. 9.4.4). When cytokines lead to an excessive production of corticosteroids, receptors in the hypothalamus and pituitary register the elevated levels and lower the production of corticosteroids.

High cytokine levels, as a result of a peripheral infection or a deliberate administration, can induce symptoms of general sickness. The impairment of cognitive abilities, for example, may be due to the uptake of Il-1↑ in the brain and its activation of the HPA axis (Banks et al. 2002).

9.4.2 Correlations Between Cytokine Action and Depression

The fact that the symptoms of general sickness and depression overlap significantly (Sect. 9.3) raises the question whether the immune mechanisms leading to sickness behavior are also involved in depression. Indeed, higher lymphocyte and phagocytic levels as well as increased amounts of the cytokines IL-1↑, IL-2↑, IL-6↑, and IFN-γ are found in patients with major depression. Furthermore, there seems to be a correlation between IL-1↑ and IL-6↑ concentration, the severity of depression, and the degree of activation of the HPA axis, and genetic polymorphisms encoding TNF-α↑ and IL-1↑ may predispose one to depression (Maes et al. 1995; Jun et al. 2003; see Schiepers et al. 2005, pp. 207–208).

Immune activity differs for different types of depression. Whereas nonmelancholic depression is characterized by increased leukocytes levels, melancholic depression is associated with normal concentrations and decreased cytokine activity (Rothermundt et al. 2001). Most studies, however, do not differentiate between different subtypes of depression, which may distort the findings. Another caveat is that while numerous studies show a correlation between the levels of a specific cytokine and depression, there are also studies that do not find significant correlations, and standard deviations are often large. In their recent meta-analysis, for example, Dowlati et al. (2010) did not find a significant correlation between most cytokines (IL-1β↑, IL-2↑, IL-4↓, IL-8↑↓,[9] IL-10↓, IFN-γ↑) and depression. Only TNF-α↑ and IL-6↑ were significantly correlated with depression. However, the highly variable results of the study suggest that TNF-α↑ and IL-6↑ activation cannot be responsible for depressive symptoms in all cases.

Like Kinney and Tanaka (2009), Schiepers et al. (2005) also argue that the fact that certain medical conditions tend to be accompanied by depression supports the claim that immune activation leads to depression. Once again, however, mere correlations between cytokine levels and depression say nothing about whether cytokines are causally involved in the etiology of depression in a relevant way (see Sect. 9.3).

9.4.3 *Causal Evidence Through Systematic Manipulation*

Important support for the cytokine hypothesis comes from experimental studies in which cytokine levels are artificially and systematically manipulated and a causal relationship between them and depression is established on the grounds that an artificial decrease or increase of cytokines results in a corresponding decrease or increase in depressive symptoms: Administration of cytokine synthesis inhibitors, cytokine antagonists, or cytokine gene manipulation can alleviate depressive symptoms, whereas activation of the immune system by endotoxins or cytokine administration intensifies them (Reichenberg et al. 2001; Schiepers et al. 2005, p. 208; for an overview of psychological side effects associated with cytokine therapy, see Kronfol and Remick (2000)).[10]

Interestingly, artificial stimulation of cytokine activity, while inducing depressive symptoms, need not also induce physical sickness symptoms (see also Maes et al. 2012). Reichenberg et al. (2001) used an endotoxin (lipopolysaccharide) that is not infectious but nevertheless invokes a response from the immune system, including flu-like symptoms with fever, uneasiness, and cytokine activity. They lowered the dose to a point where there was no sign of physical sickness symptoms, but still a rapid induction of cytokine activity, and administered this dose or a placebo to male subjects. The endotoxin increased heart rate and body temperature, but the subjects did not rate their physical symptoms worse. TNF-α↑ and IL-6↑ reached their peak quickly within 2 h, IL-1↑, cortisol, and temperature somewhat later. Measured anxiety and depressive mood strongly correlated with cytokine secretion. However, the fact that there was also a correlation with the cortisol level once again raises the question what caused the psychological responses, the cytokines or the cortisol. Since anxiety was already present before the increase in cortisol, anxiety indeed seems to have been caused by cytokine activity, which increased faster than cortisol

---

[9] IL-8 can act both pro- and anti-inflammatorily.

[10] For the philosophical details of this kind of manipulationist or interventionist account of explanation and causation that underlies such studies, see, e.g., Woodward (2003).

and can therefore better account for the response. With regard to depressed mood, however, Reichenberg and his colleagues conclude that the cytokine activation and the HPA axis were independently involved in causing the feelings (Reichenberg et al. 2001, p. 449).

Such studies suggest that cytokine activation is directly responsible for at least some depressive symptoms, thereby backing the interpretation that the correlations between cytokine levels and depression are indeed indicative of the fact that depression is a causal response to increased cytokine activity. Since psychological stress does not seem to play a crucial role in such studies, they may also be used to show that cytokine activation alone is causally sufficient to develop depression. However, almost everybody experiences at least some psychological stress most of the time, and it is impossible to rule out that the participants in these studies thought about potential troubles in their life or that the experimental setting itself was a stressor. One can of course exclude participants with depressed affect, but that still leaves the possibility that elevated cytokine levels make humans more sensitive to stress, stress which without elevated cytokine levels would not cause depressive behavior. Hence, increased cytokine levels alone may not be sufficient for triggering depression all by themselves, although they clearly exert an empirically verifiable causal influence.

9.4.4 Mechanisms of Cytokine Induced Behavior Change

Exactly how do cytokines causally affect depressive symptoms? Depression is usually associated with a decreased activity of serotonin, and cytokines can influence the activity of neurotransmitters by increasing their turnover, synthesis, reuptake, or receptor changes (Schiepers et al. 2005, pp. 209–210). Serotonin is synthesized from the precursor tryptophan. Cytokines like IL-6↑, for example, can activate the enzyme indoleamine-2,3-dioxygenase (IDO), which metabolizes tryptophan or uses tryptophan for synthesis of acute-phase proteins whose concentrations change during an immune response. This results in a decrease of tryptophan, which usually passes the blood–brain barrier and eventually becomes serotonin and thus leads to a shortage of newly synthesized serotonin. Moreover, cytokines increase turnover of serotonin by modifying its transporter and the amount and sensitivity of serotonin receptors. The increased turnover, combined with deficient renewal, leads to low serotonin levels. IDO may not only be responsible for deficient neurotransmitter synthesis, but also lead to the production of the metabolites 3-hydroxy-kynurenine (3OH-KYN) and quinolinic acid, neurotoxic substances known to be associated with anxiety and depression (e.g., Wichers and Maes 2004). 3OH-KYN has two effects: It increases reactive oxygen species, which changes the viscosity of serotonergic and catecholaminergic receptor membranes and thus potentially disturbs and lowers the activity of serotonergic pathways. And it increases monoamine oxidase activity, which in turn lowers the concentration of serotonin.

Cytokines do not only affect neurotransmitters, they also modulate the activity of the HPA axis, leading to the excessive production of cortisol that is typically linked to depression (see, however, Cowen 2002). As mentioned above (Sect. 9.4.1), a negative feedback loop is supposed to prevent the overproduction of cortisol, but immune activation can reduce its efficiency by manipulating corticosteroids receptors (e.g., via quinolinic acid).

Finally, cytokines can contribute to depression by decreasing neurogenesis in the hippocampus (Dowlati et al. 2010) and may thus be responsible for the loss of hippocampal volume observed in depression (e.g., Carroll 2004). For instance, neural progenitor cells have IL-6↑ and TNF-α↑ receptors, and Dowlati et al. (2010) suggest that IL-6↑ and TNF-α↑ negatively affect the proliferation of those progenitor cells, such that heightened levels of IL-6↑ and TNF-α↑ during depression could have an effect on a reduction of hippocampal volume.

9.4.5 The Effect of Antidepressants on Immune Functioning

If cytokines play an important causal role in the development of depression, then the question is whether antidepressants at least in part also lead to a modulation of immune activity instead of only affecting neurotransmitters. The evidence available so far is ambiguous (Schiepers et al. 2005, pp. 210–211). On the one hand, sickness behavior and neuroendocrinic changes in rats induced by lipopolysaccharides decreased or vanished completely as a result of the administration of antidepressants. In another study, the selective serotonin reuptake inhibitor paroxetine mitigated the risk of developing depression during IFN-α↑ therapy in humans (Musselman et al. 2001). On the other hand, however, IL-1↑ and IL-3↑ production increased after a treatment with clomipramine in a small study with depressed subjects (Weizman et al. 1994), and antidepressants in general do not seem to have an effect on IL-6↑. While these latter results are prima facie at odds with the cytokine hypothesis, it may be that instead of lowering the number of cytokines, antidepressant medication is just inhibiting their influence.

9.4.6 Cytokine Activity in Bipolar Disorder

According to the cytokine hypothesis, elevated cytokine levels are causally responsible for depressive symptoms. While mere correlations between elevated cytokine levels and depressive episodes (Sect. 9.4.2) and the effect of antidepressants on the immune system (Sect. 9.4.5) provide hardly any credible evidence, the systematic dependence of depressive symptoms on the manipulation of cytokine activity (Sect. 9.4.3) is a compelling reason to accept the cytokine hypothesis. At this point, then, it is tempting to ask exactly how specific the association between cytokines and depression is and to have a look, for instance, at the cytokine profiles associated with manic episodes in bipolar disorder, which in a sense are the exact opposites of depressive episodes. According to DSM-IV, bipolar disorder I is a disorder that features manic episodes or mixed episodes and, often, depressive episodes as well.

Manic episodes are characterized by an elevated, expansive, or irritable mood lasting for at least a week; symptoms include inflated self-esteem, decreased need for sleep, flight of ideas, distractibility, psychomotor agitation, and an excessive indulgence in pleasurable activities. In light of the radically different symptoms (elevated mood vs. dejection, high vs. low self-esteem, seeking pleasurable activities, both social and sexual, vs. anhedonia, and avoidance of social and sexual activity etc.), there should be no elevated cytokine levels during manic episodes if cytokines are indeed causally responsible for developing depression. Strikingly, however, elevated levels of pro-inflammatory cytokines and decreased levels of anti-inflammatory cytokines have been found during manic episodes in clients with bipolar disorder (e.g., Barbosa et al. 2012; Brietzke et al. 2009). Ortiz-Domínguez et al. (2007) have investigated phasic differences between manic and depressive episodes. They found that TNF-α↑ was similarly elevated during depression and mania, IL1-β↑ levels were similar during depression and healthy controls, but lower during mania when compared to both other conditions, IL-6↑ was higher in depression, but lower than in healthy controls and during mania (in fact, it was not detectable anymore during the manic episode), healthy controls had higher IL-2↑ levels than both bipolar phases, and the anti-inflammatory cytokine Il-4↓ was elevated during mania when compared to depression and healthy controls, which had similar levels. The fact that the level of IL1-β↑, IL-6↑, IL-2↑ was lower in mania is entirely in line with the claim that pro-inflammatory cytokines cause depression. What is perplexing is that TNF-α↑ was elevated. As mentioned above, elevated levels of TNF-α↑ are one of the most consistent findings in depression. Is it possible to explain the elevated cytokine levels without having to abandon the cytokine hypothesis?

Assuming that the cytokine levels found by Ortiz-Domínguez et al. (2007) are indeed characteristic of subjects with bipolar disorder,[11] one possible explanation is that TNF-α↑ accompanies depression, but is not causally involved in causing depressive symptoms. In this case, TNF-α↑ could be associated with manic episodes because of high cortisol levels due to a dysregulated HPA without causing depressive symptoms (Ortiz-Domínguez et al. 2007, p. 600). However, this seems implausible when considering evidence from TNF-α↑ administration in rats. After injection, these animals clearly showed depressive symptoms (e.g., Dantzer et al. 2008), suggesting that TNF-α↑ is indeed causally involved in inducing depressive symptoms.

Another possible explanation is that subjects with bipolar disorder react differently to TNF-α↑. Compared to major depression, bipolar disorder is much less frequent and has a high genetic component: The prevalence rate is around 1 % and heritability estimates are around 70–90 % for bipolar disorder and mania (e.g., Cardno et al. 1999; Weissman et al. 1996; interestingly, some of the genes associated with the risk for bipolar disorder I have also been associated with the immune system; e.g., Brietzke et al. 2011). One speculation is thus that genetic factors underlying bipolar disorder are responsible for a different reaction to TNF-α↑ in subjects with

---

[11] Their study took into account only 20 subjects with bipolar disorder. Further research is needed in order to see whether their findings can be replicated in studies with larger groups.

bipolar disorder, compared to humans and other animals without the disorder. It seems unlikely that some would react so drastically different compared to the rest of the population, with the opposite behavior to the same cytokine. Instead, subjects with bipolar disorder may simply not react (as strongly) to TNF-α↑ with depressive symptoms, which would mean that TNF-α↑ does not counteract the mechanisms producing manic symptoms.

Finally, the available evidence suggests that instead of a single cytokine level that predicts and explains depressive symptoms, specific profiles of different cytokines as a whole are responsible: For example, while TNF-α↑ may trigger depressive symptoms when other cytokines are normal, it may not be able to induce depressive symptoms when IL1-β↑, IL-6↑, IL-2↑ are lower, etc. Therefore, ratios to other cytokines, possibly also between pro-inflammatory and anti-inflammatory cytokines, could be more relevant for inducing depressive symptoms than absolute levels of single cytokines, making it necessary to take into account whole cytokine patterns and ratios in future research, instead of just comparing single cytokine levels in depressed subjects and healthy controls.

9.5 Concluding Remarks

While the infection–defense hypothesis, especially in combination with the cytokine hypothesis, offers a plausible evolutionary account of depression, some aspects of it are clearly problematic. It is obvious why the avoidance of normally pleasurable activity, the loss of self-esteem, a greater risk adversity, etc. are advantageous when it comes to avoiding infections, fighting them and preventing their spread, but the social isolation characteristic of depression may potentially be detrimental with regard to the subject's social relations and with regard to immune functioning, both directly and indirectly (Sect. 9.3). It is not at all clear whether the adaptive value of avoiding (further) infections and possible contaminations of others counterbalances these negative effects, and in particular, whether it did so in the evolutionary environment of our ancestors (Sect. 9.1). One interesting possibility, of course, is that there is just no such thing as *the* evolutionary function of depression, and that the infection–defense hypothesis has to be combined with other evolutionary accounts. For example, both the infection–defense hypothesis and the social navigation/risk hypothesis emphasize that depressive behavior minimizes risky behavior, where the former focuses on physical dangers and the latter additionally takes into account social risks. To the extent that it is important during an infection to gain social support and lower both physical and psychological stress (Sect. 9.3), these hypotheses are thus perfectly compatible: Depression may be an efficient means of fighting infections, not only by leading to anhedonia, somnolence, dysphoria, and all the other symptoms that conserve energy expenditure for an adaptive febrile response and an appropriate immunological activity, but also by minimizing social problems through social risk avoidance, through signaling the need for support, and maybe even through better problem solving abilities due to an increased tendency for rumination, given that these latter factors help reduce physical and psychological stress and secure social support, both of which are potentially crucial for a speedy recovery. Moreover, while the infection–defense hypothesis can explain why depression is preceded by major stressful life events, it fails to account for the fact that dependent stressful life events are much more important, given that the dependent/independent distinction should make no difference to immune functioning. The analytical rumination hypothesis, however, has a plausible explanation to offer: If the function of depression is indeed to facilitate problem-solving processes, then it does of course make more sense to concentrate on those problems upon which one has at least some influence (Sect. 9.2.2).

There are, however, also reasons why integrating the different evolutionary accounts may not be easy. To consider just one example, there seems to be a potential conflict between the infection–defense hypothesis and the analytical rumination hypothesis with regard to the fact that depression seems to improve attentional control, for instance in connection with the RAPM. In the study of Andrews et al. (2007), subjects with preexisting depressive affect experienced a decrease in depressive affect, whereas those without a significant preexisting depressive affect experienced an increase, which in turn was correlated with improved RAPM performance. *Prima facie* at least, the former seems to be at odds with the infection–defense hypothesis: Since the RAPM, due to the stress it induces, is potentially weakening the immune system, a preexisting depression should not led to a decrease in depressive affect. The neurobiological mechanisms of cytokines, however, appear to be in line with the biological explanations Andrews et al. (2007) offer for the observation that an increase in depressive affect is correlated with improved RAPM performance. Andrews and Thomson (2009, pp. 631–632; see also Morris et al. 1999) suggest that tryptophan depletion leads to greater activity of the dorsal raphe nucleus, which in turn increases attentional control. According to the cytokine hypothesis, cytokines may deplete tryptophan by activating IDO (Sect. 9.4.4) and may thus also foster attentional control. Observations of cytokine level changes after taking an exam (e.g., Dobbin et al. 1991) and improvements in working memory after endotoxin administration (Cohen et al. 2003) suggest that the immune system may indeed be relevant for the cognitive capacities involved in the execution of complex tasks.

Be that as it may, since it seems clear at the moment that none of the evolutionary hypotheses can account for all aspects of depression, it is a worthwhile future research endeavor to try to find out to what extent and in which way they might be rendered compatible. While, as we saw, such reconciliation is by no means straightforward, it is certainly not out of bounds either. In light of the above discussion, it will thereby be important to not focus solely on the evolutionary accounts, but to combine this research with the search for proximal accounts that are able to explain by means of which mechanisms depression can possibly fulfill all the various evolutionary functions attributed to it.[12]

---

[12] For example, in order to see to what extent the combined infection–detection/cytokine hypothesis is compatible with other evolutionary explanations, one may investigate whether cytokine induced depressive affect has features that are also said to be advantageous according to other

References

Abrahamson, L., Seligman, Y., & Teasdale, M. (1978). Learned helplessness in humans. *Abnormal Psychology, 87,* 49–74.

Allen, N., & Badcock, P. (2006). Darwinian models of depression. *Progress in Neuro-Psychopharmacology and Biological Psychiatry, 30,* 815–826.

Anders, S., Tanaka, M., & Kinney, D. (2013). Depression as an evolutionary strategy for defense against infection. *Brain, Behavior, and Immunity, 31,* 9–22.

Andrade, L., Caraveo-Anduaga, J., Berglund, P., Bijl, R., de Graaf, R., Vollebergh, W., Dragomirecka, E., Kohn, R., Keller, M., Kessler, R. C., Kawakami, N., Kiliç, C., Offord, D., Ustun, T. B., & Wittchen, H.-U. (2003). The epidemiology of major depressive episodes. *International Journal of Methods in Psychiatric Research, 12*(1), 3–21.

Andrews, P., & Thomson, J. Jr. (2009). The bright side of being blue. *Psychological Review, 116,* 620–654.

Andrews, P., Aggen, S., Miller, G., Radi, C., Dencoff, J., & Neale, C. (2007). The functional design of depression's influence on attention. *Evolutionary Psychology, 5,* 584–604.

Angst, J., Gamma, A., Sellaro, R., Zhang, H., & Merikangas, K. (2002). Toward validation of atypical depression in the community. *Journal of Affective Disorders, 72,* 125–138.

Banks, W., Farr, S., & Morley, J. (2002). Entry of blood-borne cytokines into the central nervous system. *Neuroimmunomodulation, 10,* 319–327.

Barbosa, I., Rocha, N., Bauer, M., de Miranda, A., Huguet, R., Reis, H., Zunszain, P., Horowitz, M., Pariante, C., & Teixeira, A. (2012). Chemokines in bipolar disorder. *European Archives of Psychiatry and Clinical Neuroscience, 263,* 159–165.

Bechara, A., Damasio, A., Damasio, H., & Anderson, S. (1994). Insensitivity to future consequences following damage to human prefrontal cortex. *Cognition, 50,* 7–15.

Brietzke, E., Stertz, L., Fernandes, B., Kauer-Sant'Anna, M., Mascarenhas, M., Escosteguy Vargas, A., Chies, J., & Kapczinski, F. (2009). Comparison of cytokine levels in depressed, manic and euthymic patients with bipolar disorder. *Journal of Affective Disorders, 116,* 214–217.

Brietzke, E., Stabellini, R., Grassis-Oliveira, R., & Lafer, B. (2011). Cytokines in bipolar disorder. *CNS Spectrums, 16*(7), 157–168.

Buist-Bouwman, M., Ormel, J., de Graaf, R., & Vollebergh, W. (2004). Functioning after a major depressive episode. *Journal of Affective Disorders, 82,* 363–371.

Butcher, J., Mineka, S., & Hooley, J. (2013). *Abnormal psychology*. New York: Pearson.

Cardno, A., Marshall, E., Coid, B., Macdonald, A., Ribchester, T., Davies, N., Venturi, P., Jones, L., Lewis, S., Sham, P., Gottesman, I., Farmer, A., McGuffin, P., Reveley, A., & Murray, R. (1999). Heritability estimates for psychotic disorders. *Archives of General Psychiatry, 56,* 162–168.

Carroll, B. (2004). Untreated depression and hippocampal volume loss. *American Journal of Psychiatry, 161,* 1309–1310.

Caspi, A., Sugden, K., Moffitt, T., Taylor, A., Craig, I., Harrington, H., McClay, J., Mill, J., Martin, J., Braithwaite, A., & Poulton, R. (2003). Influence of life stress on depression. *Science, 301,* 386–389.

Cohen, S., Tyrrell, D., & Smith, A. (1991). Psychological stress and susceptibility to the common cold. *New England Journal of Medicine, 325,* 606–612.

Cohen, O., Reichenberg, A., Perry, C., Ginzberg, D., Pollmächer, T., Soreq, H., & Yirmiya, R. (2003). Endotoxin-induced changes in human working and declarative memory associate with cleavage of plasma "readthrough" acetylcholinesterase. *Journal of Molecular Neuroscience, 21,* 199–212.

Comer, R. (2013). *Abnormal psychology*. New York: Worth.

Cosmides, L., & Tooby, J. (1994). Origins of domain specificity. In L. Hirschfeld & S. Gelman (Eds.), *Mapping the mind* (pp. 85–116). Cambridge: Cambridge University Press.

---

hypotheses. Regarding a compatibility with the analytical rumination hypothesis, for instance, one might try to find out whether an artificial increase of cytokine activity results in better RAPM scores.

Cowen, P. (2002). Cortisol, serotonin and depression. *British Journal of Psychiatry, 180,* 99–100.

Danese, A., Moffitt, T., Pariante, C., Ambler, A., Poulton, R., & Caspi, A. (2008). Elevated inflammation levels in depressed adults with a history of childhood maltreatment. *Archives of General Psychiatry, 65,* 409–417.

Dantzer, R., Bluthé, R.-M., Castanon, N., Kelley, K., Konsman, J.-P., Laye, S., Lestage, J., & Parnet, P. (2007). Cytokines, sickness behavior, and depression. In R. Ader (Ed.), *Psychoneuroimmunology* (Vol. 1, pp. 281–318). San Diego: Academic.

Dantzer, R., O'Connor, J., Freund, G., Johnson, R., & Kelley, K. (2008). From inflammation to sickness and depression. *Nature Reviews Neuroscience, 9,* 46–56.

Davidson, R., Pizzagalli, D., Nitschke, J., & Putnam, K. (2002). Depression. *Annual Reviews of Psychology, 53,* 545–574.

Dhabhar, F. (2009). A hassle a day may keep the pathogens away. *Integrative and Comparative Biology, 49,* 215–236.

Dobbin, J., Harth, M., McCain, G., Matin, R., & Cousin, K. (1991). Cytokine production and lymphocyte transformation during stress. *Brain, Behavior, and Immunity, 5,* 339–348.

Dowlati, Y., Herrmann, N., Swardfager, W., Liu, H., Sham, L., Reim, E., & Lanctôt, K. (2010). A meta-analysis of cytokines in major depression. *Biological Psychiatry, 67,* 446–457.

Doyle, C., Ewald, S., & Ewald, P. (2007). Premenstrual syndrome. *Perspectives in Biology and Medicine, 50,* 181–202.

Garlow, S., & Nemeroff, C. (2003). Neurobiology of depressive disorders. In R. Davidson, K. Scherer, & H. Goldsmith (Eds.), *Handbook of affective sciences* (pp. 1021–1043). Oxford: Oxford University Press.

Germain, A., & Kupfer, D. (2008). Circadian rhythm disturbances in depression. *Human Psychopharmacology, 23,* 571–585.

Gonda, X., Fountoulakis, K., Juhasz, G., Rihmer, Z., Lazary, J., Laszik, A., Akiskal, H., & Bagdy, G. (2009). Association of the s allele of the 5-HTTLPR with neuroticism-related traits and temperaments in a psychiatrically healthy population. *European Archives of Psychiatry and Clinical Neuroscience, 259,* 106–113.

Gotlib, I., Joormann, J., Minor, K., & Hallmayer, J. (2008). HPA axis reactivity. *Biological Psychiatry, 63,* 847–851.

Gould, S. (1997). Evolution: The pleasures of pluralism. *New York Review of Books*. http://www.nybooks.com/articles/archives/1997/jun/26/evolution-the-pleasures-of-pluralism/. Accessed 1 Sept 2014.

Gray, J., Chabris, C., & Braver, T. (2003). Neural mechanisms of general fluid intelligence. *Nature Neuroscience, 6,* 316–322.

Gross, C., & Hen, R. (2004). The developmental origins of anxiety. *Nature Reviews Neuroscience, 5,* 545–552.

Hammen, C. (2005). Stress and depression. *Annual Review of Clinical Psychology, 1,* 293–319.

Jacobs, B., & Fornal, C. (1999). Activity of serotonergic neurons in behaving animals. *Neuropsychopharmacology: official publication of the American College of Neuropsychopharmacology, 21,* S9–S15.

Janowsky, D. (2001). Introversion and extroversion. *Current Psychiatry Reports, 3,* 444–450.

Jun, T.-Y., Pae, C.-U., Hoon-Han, Chae, J.-H., Bahk, W.-M., Kim, K.-S., & Serretti, A. (2003). Possible association between -G308 A tumour necrosis factor-a gene polymorphism and major depressive disorder in the Korean population. *Psychiatric Genetics, 13,* 179–181.

Keedwell, P. (2008). *How sadness survived*. Abingdon: Radcliffe.

Kim-Cohen, J., & Gold, A. (2009). Measured gene-environment interactions and mechanisms promoting resilient development. *Current Directions in Psychological Science, 18,* 138–142.

Kinney, D., & Tanaka, M. (2009). An evolutionary hypothesis of depression and its symptoms, adaptive value, and risk factors. *Journal of Nervous and Mental Disease, 197,* 561–567.

Kirchsteiger, G., Rigotti, L., & Rustichini, A. (2006). Your morals might be your moods. *Journal of Economic Behavior and Organization, 59,* 155–172.

Kronfol, Z., & Remick, D. (2000). Cytokines and the brain. *American Journal of Psychiatry, 157,* 683–694.

Kunz-Ebrecht, S., Mohamed-Ali, V., Feldman, P., Kirschbaum, C., & Steptoe, A. (2003). Cortisol responses to mild psychological stress are inversely associated with proinflammatory cytokines. *Brain, Behavior, and Immunity, 17,* 373–383.

Lambert, G., Reid, C., Kaye, D., Jennings, G., & Esler, M. (2002). Effect of sunlight and season on serotonin turnover in the brain. *Lancet, 360,* 1840–1842.

Lyubomirsky, S., Tucker, K., Caldwell, N., & Berg, K. (1999). Why ruminators are poor problem solvers. *Journal of Personality and Social Psychology, 77,* 1041–1060.

Maes, M., Meltzer, H., Bosmans, E., Bergmans, R., Vandoolaeghe, E., Ranjan, R., & Desnyder, R. (1995). Increased plasma concentrations of interleukin −6, soluble interleukin −6, soluble interleukin −2 and transferrin receptor in major depression. *Journal of Affective Disorders, 34,* 301–309.

Maes, M., Berk, M., Goehler, L., Song, C., Anderson, G., Gałecki, P., & Leonard, B. (2012). Depression and sickness behavior are Janus-faced responses to shared inflammatory pathways. *BMC Medicine, 10,* 66.

Mealey, L. (1995). The sociobiology of psychopathy. *Behavioral and Brain Sciences, 18,* 523–599.

Morris, J., Smith, K., Cowen, P., Friston, K., & Dolan, R. (1999). Covariation of activity in habenula and dorsal raphé nuclei following tryptophan depletion. *NeuroImage, 10,* 163–172.

Musselman, D., Lawson, D., Gumnick, J., Manatunga, A., Penna, S., Goodkin, R., Greiner, K., Nemeroff, C., & Miller, A. (2001). Paroxetine for the prevention of depression induced by high-dose interferon alfa. *New England Journal of Medicine, 344,* 961–966.

Nes, R., Czajkowski, N., Røysamb, E., Ørstavik, R., Tambs, K., & Reichborn-Kjennerud, T. (2013). Major depression and life satisfaction. *Journal of Affective Disorders, 144,* 51–58.

Nettle, D. (2004). Evolutionary origins of depression. *Journal of Affective Disorders, 81,* 91–102.

Ortiz-Domínguez, A., Hernández, M., Berlanga, C., Gutiérrez-Mora, D., Moreno, J., Heinze, G., & Pavón, L. (2007). Immune variations in bipolar disorder. *Bipolar Disorders, 9,* 596–602.

Parker, G., Roy, K., Mitchell, P., Wilhelm, K., Malhi, G., & Hadzi-Pavlovic, D. (2002). Atypical depression. *American Journal of Psychiatry, 159,* 1470–1479.

Profet, M. (1992). Pregnancy sickness as adaptation. In J. Barkow, L. Cosmides, & J. Tooby (Eds.), *The adapted mind* (pp. 327–365). Oxford: Oxford University Press.

Rabin, B. (1999). *Stress, immune function and health*. New York: Wiley.

Reichenberg, A., Yirmiya, R., Schuld, A., Kraus, T., Haack, M., Morag, A., & Pollmächer, T. (2001). Cytokine-associated emotional and cognitive disturbances in humans. *Archives of General Psychiatry, 58,* 445–452.

Richardson, R. (2007). *Evolutionary psychology as maladapted psychology*. Cambridge: MIT.

Rothermundt, M., Arolt, V., Fenker, J., Gutbrodt, H., Peters, M., & Kirchner, H. (2001). Different immune patterns in melancholic and non-melancholic major depression. *European Archives of Psychiatry and Clinical Neuroscience, 251,* 90–97.

Schiepers, O., Wichers, M., & Maes, M. (2005). Cytokines and major depression. *Progress in Neuro-Psychopharmacology and Biological Psychiatry, 29,* 201–217.

Shafran, R., & Mansell, W. (2001). Perfectionism and psychopathology. *Clinical Psychology, 21,* 879–906.

Sheeber, L., Hops, H., & Davis, B. (2001). Family processes in adolescent depression. *Clinical Child and Family Psychology Review, 4,* 19–35.

Southwick, S., Vythilingam, M., & Charney, D. (2005). The psychobiology of depression and resilience to stress. *Annual Review of Clinical Psychology, 1,* 255–291.

Spijker, J., Graaf, R., Bijl, R., Beekman, A., Ormel, J., & Nolen, W. (2002). Duration of major depressive episodes in the general population. *British Journal of Psychiatry, 181,* 208–213.

Sweeney, P., Anderson, K., & Bailey, S. (1986). Attributional style in depression. *Journal of Personality and Social Psychology, 50,* 974–991.

Tinbergen, N. (1963). On aims and methods of ethology. *Zeitschrift für Tierpsychologie, 20,* 410–433.

Tooby, J., & Cosmides, L. (1990). On the universality of human nature and the uniqueness of the individual. *Journal of Personality, 58,* 17–67.

Vorhold, V., Giessing, C., Wiedemann, P., Schütz, H., Gauggel, S., & Fink, G. (2007). The neural basis of risk ratings. *Neuropsychologia, 45,* 3242–3250.

Wagner, U., Gais, S., Haider, H., Verleger, R., & Born, J. (2004). Sleep inspires insight. *Nature, 427,* 352–355.

Walker, M. (2009). The role of sleep in cognition and emotion. *Annals of the New York Academy of Sciences, 1156,* 168–197.

Watson, P., & Andrews, P. (2002). Toward a revised evolutionary adaptationist analysis of depression. *Journal of Affective Disorders, 72,* 1–14.

Weissman, M.M., Bland, R.C., Canino, G.J., Faravelli, C., Greenwald, S., Hwu, H.G., ... & Yeh, E.K. (1996). Cross-national epidemiology of major depression and bipolar disorder. *Jama, 276*(4), 293–299.

Weizman, R., Laor, N., Podliszewski, E., Notti, I., Djaldetti, M., & Bessler, H. (1994). Cytokine production in major depressed patients before and after clomipramine treatment. *Biological Psychiatry, 35,* 42–47.

Wender, P., Kety, S., Rosenthal, D., Schulsinger, F., Ortmann, J., & Lunde, I. (1986). Psychiatric disorders in the biological and adoptive families of adopted individuals with affective disorders. *Archives of General Psychiatry, 43,* 923–929.

Wichers, M., & Maes, M. (2004). The role of indoleamine 2,3-dioxygenase (IDO) in the pathophysiology of interferon-α-induced depression. *Journal of Psychiatry and Neuroscience, 29,* 11–17.

Woodward, J. (2003). *Making things happen*. Oxford: Oxford University Press.

Yirmiya, R., Weidenfeld, J., Pollak, Y., Morag, M., Morag, A., Avitsur, R., Barak, O., Reichenberg, A., Cohen, E., Shavit, Y., & Ovadia, H. (1999). Cytokines, "depression due to a general medical condition", and antidepressant drugs. In R. Dantzer, E. Wollman, & R. Yirmiya (Eds.), *Cytokines, stress and depression* (pp. 283–316). New York: Kluwer.

Zahavi, A., & Zahavi, A. (1997). *The handicap principle*. Oxford: Oxford University Press.

Chapter 10
When Misrepresentation is Successful

Michael Zehetleitner and Felix D. Schönbrodt

How can a representation lead to a successful action? It is widely taken for granted that a representation has to be true in order to be successful (e.g., Shea 2007; Millikan 1989; Whyte 1990; Ramsey and Moore 1927; Blackburn 2005). Consider for instance, how could a person successfully sit down on or avoid a chair, unless she has a true visual representation of the chair's shape and position? Or, to make use of an example by Ramsey (1927): How can the belief of a chicken, that a certain caterpillar is toxic be useful, unless the caterpillars are actually toxic? It even has been proposed that '[t]ruth just *is* the property of a belief that suffices for your getting what you want when you act upon it' (Whyte 1990, p. 149). Although this notion of truth being a prerequisite for success has a high face validity, we shall argue that it is wrong to suppose that all successful actions require true representations. Success does not require truth in all cases. Under certain circumstances, false representations can systematically cause successful actions. The goal of the chapter is to demonstrate what these defined circumstances are.

10.1 Success Semantics

Throughout this chapter, we assume a naturalistic theory of semantics, where representations, their content, their truth or falsehood are defined without recurrence to terms which themselves are already intentional. Our argumentation is invariant to what exact version of naturalized semantics is assumed (e.g., Fodor 1990; Millikan 1989; Papineau 1984, 2003; Dretske 1981). Now, what does it mean for a representation to be true versus false or successful versus unsuccessful?

M. Zehetleitner (✉) · F. D. Schönbrodt
Department of Psychology, Ludwig-Maximilians-University, Munich, Germany
e-mail: mzehetleitner@psy.lmu.de

T. Breyer (ed.), *Epistemological Dimensions of Evolutionary Psychology,*
DOI 10.1007/978-1-4939-1387-9_10

10.1.1 Representation

We understand that observer *O* has representation *r;* if and only if (iff) *r* is a (physical or biological) state or signal within *O,* which has the property of being about something else, that is to have content. To have content can be considered as a mapping from a set of representations *R*[1] (of which *r* is an element) to a specific set of states in the world external to the observer, the target domain *T*.[2]

$$cont : R \to T \tag{10.1}$$

Together with the content mapping, when an observer *O* has representation $r \in R$ and $cont(r) = t^*$ it is possible to state: Observer *O* has a representation *r* the content of which is t^*, or *O* believes that *T* is t^*.[4] Using this notation for content, it is possible to dissociate two aspects of 'aboutness' (or 'intentionality'): First, a set of representations *R* is about a specific set of world states *T* (the domains of the content mapping *cont*) and not about a different set of states, say *T*'. For instance, a specific representation *R* is about the velocity of an object and not about its size. This first aspect of 'aboutness' specifies which measurement unit the representation's content has (e.g. [m/s] or [m]). Second, a specific representation $r \in R$ has the specific content $cont(r) = t^* \in T$. To use a neurophysiological example, neuronal activity in orientation columns in primary visual cortex (of, e.g. cats or primates, including humans) would be a set of representations *R,* which is about the angular orientation (e.g. vertical or horizontal) of line segments and not about their colour (e.g. Hubel and Wiesel 1974). A specific pattern of neuronal activity then represents a specific angular orientation (e.g. horizontal) and not any other orientation (e.g. vertical).

10.1.2 Truth

In order to talk about the truth or falsehood of a representation, it is required to determine one further component: the actual state of the world, $\underline{t} \in T$. Remember, when the content of a certain cell assembly in a cat V1 is 'horizontal' (i.e. 0°) it is possible to state that the cat believes that the bar is horizontal. This belief is true, in case the actual state of the world (that is the line segment's orientation) indeed is 'horizontal' ($\underline{t} = t^*$), and false, in case it is not ($\underline{t} \neq t^*$):

Representation $r \in R$ is true:

---

[1] Capital letters here denote variables. A variable is a set of possible values together with a measurement operation which allows determination of what value currently is the case. For instance, the size of an apple can be considered as a variable S. The possible values of S are within the interval between zero and infinity. The measurement is a ruler, which also determines the quality of the variable, namely [cm] or [m].

[2] Here, we consider only first-order representations, the content of which is situated in the observers' external environment. In principle, representations can also be meta-representations, i.e. about other representations.

$$cont(r) = \underline{t}. \tag{10.2}$$

10.1.3 Success

Now, when is a representation successful? A representation is considered successful if it is useful for actions (Ramsey 1927), if it allows desires to be fulfilled (Blackburn 2005; Whyte 1990), or if it is causally effective in increasing the organism's fitness, i.e. the expected value of the number of reproducible offspring (e.g. Millikan 1989). A frequent assumption is that representations are successful if and only if they are true: for any representation $r \in R$, and the actual state of affairs $\underline{t} \in T$, an operation to determine the success of action *a, successful,* and the content mapping *cont* between *R* and *T*:

$$(successful(a) \wedge r \rightarrow a) \Leftrightarrow cont(r) = \underline{t} \tag{10.3}$$

that is the content of the representation is equal to the actual state of the world (10.2).[3]

Until here, truth and success have been introduced as binary. However, the terminology until now can also include gradual cases: For instance, a deer wants to jump over a 1.2-m-wide ditch when trying to evade a predator. If its internal representation of the ditch's width is true, it can plan its jump and successfully reach the other bank. If its internal representation however is false, it can be so in many ways. The width-representation's content can for instance be 1.199 m, which is false. Still, the evading jump could be successful. If the representation's content is for instance 1.1 m, the deer could still evade the predator, but twist its ankle. If its content is 0.5 m, its desire to evade the predator will most probably turn out not to be fulfilled. This example illustrates that for states of the world *T* for which an ordinal scale of measurement exists, the deviation from truth can be rank ordered, i.e. for two false representation one can be 'more false' than the other. That is, even though logical truth is still binary, the falsehood of representations can be gradually qualified and differentiated. Also, success can be used as a binary term. Therefore, it is necessary, to introduce a further term, which quantifies the outcome of each action combined with each state of the world *T*. Applied to the deer example, a jump of any length *a,* under the condition of any width of ditch, *t,* has a certain consequence, hence termed utility.[4] Thus, for a given length of jump *a,* the result differs in utility depending on the actual width of the ditch *t*. Utility should be maximal for that width

---

[3] Although it might be debated whether the truth of a representation 'guarantees' success (see Nanay 2013; Blackburn 2005), there is general consensus that a false representation is incapable of generating a successful action other than accidentally.

[4] Utility is used in a similar sense in evolutionary biology (e.g. Reeve and Sherman 1993; Fox and Westneat 2010). In decision science it is frequently termed pay-off, (cost) value, or reward (e.g. Green and Swets 1966; Gold and Shadlen 2007). It can be quantified in cardinal or ordinal scale.

of ditch, which corresponds to the length of jump, i.e. for $a=t$. An action a is called successful, if its result has the highest utility, given the current state of affairs, t, compared to all other possible actions.

A requirement is that utility can be assigned an at least ordinal scale, where the outcome of an action can differ in its results. For instance, biological fitness meets this requirement. The relationship between truth and success (10.3) in this case would state that a representation is true *iff* it causes the most utile action.

Consequently, independent of how utility is defined in detail, it at least comprises a measure which can be assigned to each action/world combination, because the utility of an action is always relative to what state the world is currently in. Applied to the deer example, for each width of ditch (states of the target domain T) and each jump length (set of actions A), a (for instance real) measure can be assigned quantifying the action's utility given the current state of the target domain:[5]

$$\Omega : A \times T \to \mathbb{R};\ (a,t) \mapsto \omega(a,t). \tag{10.4}$$

For the deer example, the utility surface is depicted in Fig. 10.1. For each state of the world, T, the width of ditch in the deer example, utility is maximal for a jump of the corresponding length (i.e. where the width representation is true), which is in line with success semantics (10.3) the successfulness of an action, *successful,* in Eq. (10.3) is defined using the construct of utility. An action $a' \in A$ is successful for a given $t \in T$ if a' has the highest utility value compared to all other $a \in A$ given t:

$$successful(a') : a' = argmax_a \omega(t,a). \tag{10.5}$$

Now, if there were a one-to-one mapping between the target domain and representations, all representations would always lead to maximally utile, i.e. successful actions, because every ditch width t would invariably lead to the true representation r. That is, there would be no chance of misrepresentation. In such a situation, in information theoretic terminology, the mutual information between the target domain T and the set of representations R would be equal to T's entropy, that is, there would be no loss of information. Generally, information between T and R is lost, if R is not solely causally affected by T, but also by any other source, Z (see Fig. 10.1).

10.2 Optimal Bias: No Challenge to Success Semantics

Let's apply this abstract conception to an example provided by Godfrey-Smith (1991), which has been first discussed by Millikan (1989; see also Usher 2001) and which has been considered as a potential counterexample for success semantics. A hunting tiger sees a movement of the grass. Either, the grass is moved by prey or by the wind, and the carnivore has to make the decision whether to jump and attack or

---

[5] Assuming that a representation invariably causes the same action, in the present case, a representation of x meters always leads to a jump of x meters.

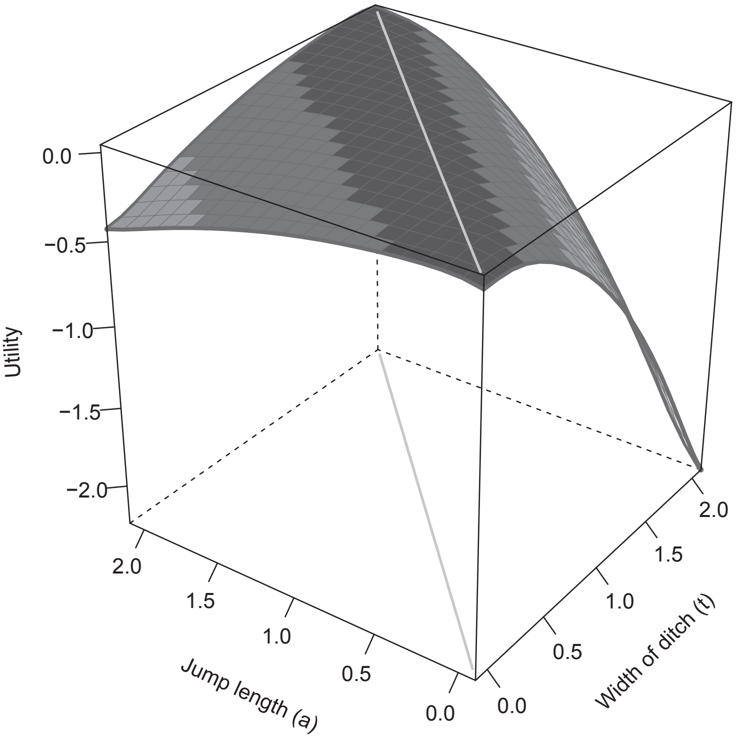


Fig. 10.1 This (hyper)-surface denotes the *utility* relationship between *width of ditch*, *t* (*x*-axis) and *length of jump* a (*y*-axis) on *utility* (vertical *z*-axis). One can observe that for each *width of ditch*, there is one *length of jump*, for which *utility* is maximal, represented by the *white line* on top of the surface and its projection onto the $T \times A$ plane. This maximal *utility* (i.e. success) is achieved for correspondence between *width of ditch* and *jump*, respectively, i.e. $a = t$. Jumping too short ($a < t$) leads to less *utility* than jumping too far ($a > t$); the latter only causing waste of energy, the former potential injuries or capture. The surface is not restricted to continuous variables of T and action A. For, e.g., dichotomic variables, the surface would consist of four points where, again, for each action the largest utility value is present at the true representation (see, for instance, Table 10.1)

Table 10.1 Utility values ω for each combination of *t* and *a* for the tiger example

Target	Action	
	Jump	Ignore
Prey	+20 (hit)	−0.1 (miss)
Non-prey	−5 (false alarm)	−0.1 (CR)

not. That is, the target domain set T consists of the elements {prey, non-prey}, the set of actions A of the elements {jump, ignore}.

Let X denote a signal in the tiger's visual system, the strength of which is proportional to the vehemence with which the grass is disturbed. The outcome of the tiger's decision about presence of prey is a representation R, the content of which is T, that is either prey or non-prey. As a simplification, following Godfrey-Smith,

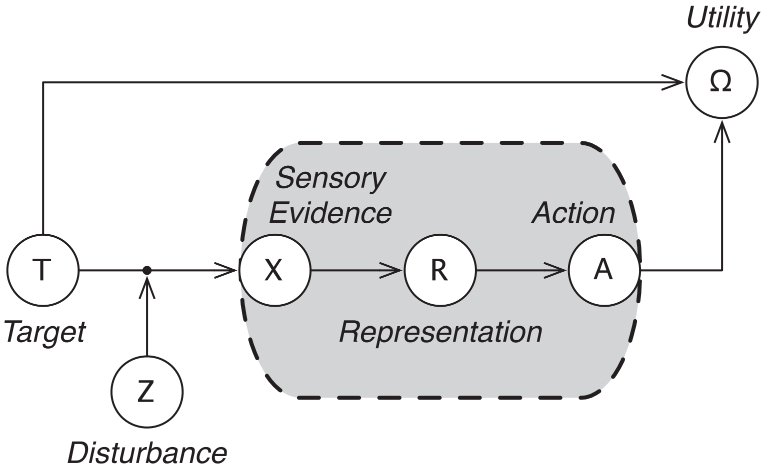


Fig. 10.2 Variables are denoted by *circles, arrows* denote causal connections. The *dashed line* denotes the boundaries of the organism *O,* which has representation *R*. *R* is about the state of the world *T,* external to *O*. *T* affects *O*'s sense organs producing sensory evidence *X*. *R* causes one of a possible repertoire of actions *A*. Ω denotes the utility of each action *A* combined with each state of the world *T*

let's assume that there is an invariable link between *R* and the tiger's action *A:* It always jumps when it believes that prey is present and ignores non-prey. How does the utility surface look like? First, the surface can be represented as a matrix, because both *T* and *A* are dichotome. Jumping when prey is present ultimately leads to feeding which here is assigned $+20$ in arbitrary units of utility. Jumping when no prey is present leads to a waste of energy and is assigned a utility of -5. The utility of ignoring is independent of whether prey is indeed present or absent and set to -0.1.[6] This reflects the fact that due to metabolic loss of energy continuous ignoring will lead to starvation. In general, the utility values for combinations of the target domain *T* and actions *A* are shown in Table 10.1.

Now, not only prey can disturb the grass, but also the wind. That means the information flow between the presence of prey *T* and the sensory state *X* is reduced by the disturbance (denoted *Z* in Fig. 10.2). The disturbance can have a source external to the organism (e.g. wind moving the grass) or internal to it (i.e. neuronal noise). In case no information is lost, there are no errors. In case of information loss, the animal has to balance which type of mistake to make: Should it be unbiased and equally often jump at non-prey as it ignores prey? Such a decision scheme would intuitively seem disadvantageous, given that jumping at non-prey is more costly[7] than

---

[6] The exact numerical values are arbitrary.

[7] Costs can be defined in two ways. First, negative entries in the utility matrix can be considered costs. The definition we are using in the subsequent chapter is the following: The incorrect action has less utility than the correct action. We understand as cost the amount of how much less utility an incorrect action has, compared to the correct action. That is, the 'costly' action can have a positive entry in the utility matrix, but we would still speak of cost, as utility would be positive but smaller than for the correct action.

ignoring prey (see Table 10.1). Indeed, decision-theoretic calculations confirm that the tiger acts optimal if it is biased towards jumping rather than ignoring, because the cost of a false decision is greater in case prey indeed is present compared to when prey is absent (e.g. Godfrey-Smith 1991). Thus, an optimistic tiger throughout life gains more calories than its sceptic fellow tiger.

The opposite is true for beavers (using an example by Millikan 1989): Optimistic beavers which, when the grass is disturbed, assume it to be the wind rather than a predator have a shorter life expectancy than skittish beavers which raise alarm and hide at the slightest disturbance of grass. Here, the cost of a false decision in case a predator is present is greater than the cost in case no predator is present.

Decisions can be (optimally) shifted not only depending on the cost functions but also on the base probability of events *T*. Consider a bear that is foraging at a river full of salmon, optimally has a very liberal criterion and claws the water at the merest indication of a flicker in the water. When fish are scarce, it is optimal to require more visual evidence for making a fishing attempt (for optimal foraging behaviour, see, e.g. Stephens and Krebs 1987).

The conclusion to be drawn from these examples is twofold. First, there can be calculated decision criteria, which are optimal in the sense that they maximize the expected value of utility, taking into account the cost of different types of errors and the a priori frequencies within the target domain.[8] Second, this optimal criterion can be biased favouring one type of mistake (false alarm or miss) over the other, in cases, where both errors differ in cost, or where the a priori base rate of events is not uniform.

Would such optimally biased decision criteria challenge success semantics? One could argue that a bear foraging at a river rich in fish should be credulous rather than sceptic. Its overall utility is maximized if it frequently falsely believes fish to be present rather than impartially or sceptically evaluating visual evidence. Couldn't one now argue that the falsity of the bear's beliefs leads to maximal utility? Although this argument, to our knowledge, has never been proposed, it has been discussed and refuted (Millikan 1989; Godfrey-Smith 1991; Usher 2001). The refutation argument is as follows: A high proportion of false beliefs can lead to maximal utility in case a lot of situations are aggregated (i.e. *in average*). However, in each *specificinstance,* the bear can feed iff its belief about the presence of fish was true. Also, each false alarm in every *specific instance* leads to a loss of energy. The fact that on average false alarms are rare, because the river is packed with fish, thus cannot challenge the close coupling between truth and success of a single representation. This coupling is reflected in the utility matrix, where for each action the utility is maximal iff the representation is true, and the utility matrix itself remains unaltered even if the a priori base frequencies of events change. Consequently, even though frequent misrepresentations are optimal, they are no challenge to success semantics (10.3).

---

[8] Cf. Godfrey-Smith (1991) for the analytic derivation of the optimal decision criterion. An alternative (and equivalent) approach has been described by Bischof (1998).

10.3 False Representations Which Lead to Success

Here, we demonstrate that it is indeed possible that a false representation systematically causes successful actions—not only on average but also in specific instances (building upon an example first presented by Bischof 2009). Above, we have summarized that frequent false representations are optimal (i.e. maximizing average utility) under specific conditions: First, there has to be information loss between the target domain and the biological signals, based on which representations are formed. Second, there has to be present at least one type of asymmetry: Either the a priori frequency of events in the target domain is not uniform or the utility matrix Ω has to be asymmetric. We now argue that in cases where actions are based on indicator representations, success semantics can be violated.

10.3.1 Indicator Representations

In order to describe what we mean with indicator representations, let's sketch out a further example of monkeys in presence of toxic and harmless snakes (cf. Bischof 2009). First, without considering indicator representations, the target domain of the representation is the snake's toxicity, $T_t = \{\text{toxic, non-toxic}\}$. The monkey's representation R_t of the fact also is binary and can cause the actions $A = \{\text{eat, run}\}$ (see Fig. 10.3).

The utility matrix is represented in Table 10.2. Trying to eat a toxic snake is not assumed to be deadly, but to have severe fitness consequences. Eating a snake has moderate gain. Here, success semantics (10.3) is satisfied: For each action, utility is maximal iff the mediating representation is true.

Now, the monkey does not have a poison detector, but snakes come in different colours and nature has it that toxic snakes are coloured bluish, whereas harmless snakes are of a greenish colour. That is, the snakes' colour can be used as an indicator for their toxicity. The snakes' colour T_c is assumed to vary continuously between saturated blue and saturated green, including a completely desaturated (i.e. greyish) colour. Therefore, T_c varies between $-\infty$ and ∞, where 0 stands for grey, negative values for a blue hue, positive values for a green hue and its absolute value for the colour's saturation. As the monkey has a colour detector, the activity of which is denoted X_c (see Fig. 10.3), it still is able to distinguish toxic from harmless snakes and act accordingly. However, again, there is information loss between the actual colour T_t and sensory evidence X_c. Specifically, the sensory evidence for each colour t_c is assumed to be normally distributed around x_c. As a consequence, there is uncertainty, given a specific value of sensory evidence what the snake's colour actually was (see also Fig. 10.4, left panel).

But how does the monkey decide upon its action, given its visual input? First, it forms an indicator representation of colour, R_c, about T_c, based on sensory evidence X_c and finally, depending on the colour representation, a representation of the toxicity R_t is triggered (with blue snakes always being treated as toxic, that is leading to a 'run' action). A snake is assumed to be toxic, if the colour representation indicates

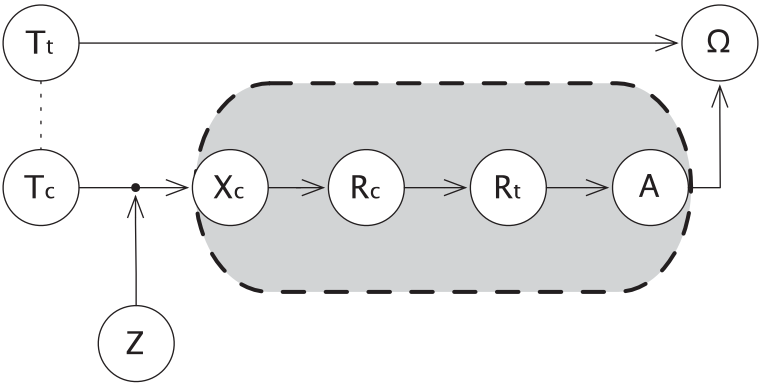


Fig. 10.3 The notation is equivalent to Fig. 10.2. New variables are T_c and R_c which denote a state of the world external to the organism and its representation. The *dotted line* between T_c and T_t indicates at least a correlation, possibly a causal connection

Table 10.2 Utility matrix Ω for the monkey example

Toxicity	Action	
	Eat	Run
Toxic	−1	0
Non-toxic	0.5	0

a blue hue, i.e. for all r_c 0 representation r_t=toxic is formed. For all $r_c \geq 0$, the representation r_t=non-toxic is formed. The whole causal setup is depicted in Fig. 10.3.

This setup differs from the thought experiments presented above in one important aspect: It is no longer assumed that there is a direct causal connection from the fitness-relevant feature (the toxicity of the snake) to the sensory apparatus of the monkey. Rather, a different feature, in this case colour, is picked up by the observer and used as an indication for the success-relevant feature, which is possible if there is a correlation between the indicator feature T_c and the success-relevant feature T_t. The subsequent discussion does not depend on there being no causal connection between T_t and T_c. It could well be that the poison within the snake causes the production of a molecule within the snake's skin, making it appear blue. The crucial difference is that there is no direct causal route from the success-relevant feature T_p to sensory evidence, but only an indirect one via T_c (and thus X_c).

10.3.2 *Plausibility Check*

Before discussing the impact of indicator representations on success semantics, let's first check their plausibility. In fact, there are natural examples in abundance, where indicator representations are used by organisms, because frequently they do not possess sense organs which are directly causally affected by those features that are relevant for the success/utility of actions, such as fertility of a mate, nutritiousness

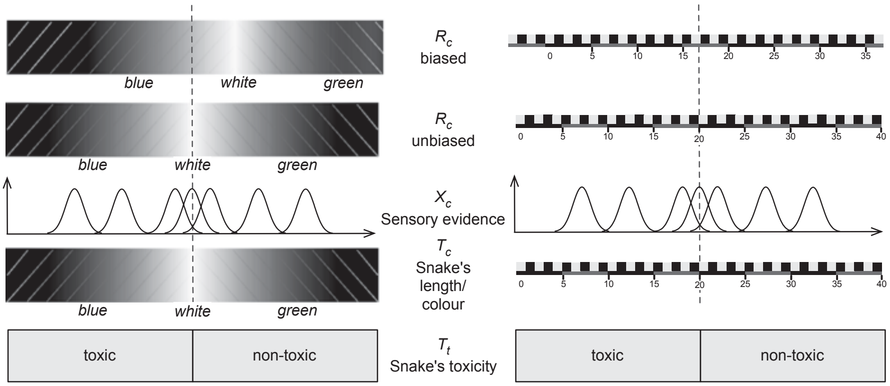


Fig. 10.4 The *left panel* presents toxicity T_t, colour T_c and sensory evidence X_c according to the thought experiment. T_t and T_c are perfectly correlated with snakes of a completely desaturated colour being non-toxic. In the figure, the snake's colour's hue is represented by *texture* and saturation by *luminance*. Toxic snakes have a blue hue (*right tilted texture*), and non-toxic snakes have a green hue (*left tilted texture*). *White colour* in the figure denotes completely desaturated colour of the snake. The *top* of the figure displays two possibilities of how a colour representation R_c can be formed based on X_c: unbiased, where the colour grey leads to representations with content blue and green with equal probability, or blue-biased, where grey mostly leads to representations with content blue. The *right panel* presents an analogous example, where snakes differ in size and snakes smaller than 20 cm are toxic. R_c are again presented for the unbiased and the biased case

of food or danger of a predator. Consider for instance the tiger in the example by Godfrey-Smith (1991): It uses disturbances of the grass as an indicator for presence of prey. Consider alternatively the famous example by Dretske (1986): There exist prokaryotes which have a cell organelle responsive to magnetic north. Those anaerobic prokaryotes with a habitat in the northern hemisphere align themselves to drift north, in the southern hemisphere to drift south (Blakemore 1975; for a recent review see Bazylinski and Frankel 2004). In both habitats, the organisms drift into deep waters containing less oxygen. Here, magnetic north is used as an indicator for oxygen, because the utility of their alignment action depends on the oxygen level of the waters they are drifting into, not on their magnetic properties. Let's briefly touch on a couple of further examples. Frogs and toads are known to snap at small dark moving objects or longish things moving along their long axis, respectively (Lettvin et al. 1968; Ewert 1974; Borchers et al. 1978), indicating presence of prey. Sickle-back males show behaviour of territory defense when confronted with an oval, medium-sized dummy with a red lower half (Tinbergen 1951) and sickle-back females show courtship behaviour for dummies of correct hue, contrast and configuration (Baube et al. 1995), i.e. they use visual configurations as indicators for presence of rivals or mates. Finally, vervet monkey sentinels can issue alarm calls based on which other vervets in the vicinity perform an evading action fitting to the type of attacker indicated by the alarm call. In this case, the reacting vervets have potential sensors for detecting predators, but they use the alarm call as an indication of predators' presence, even though these are currently out of their sensory range (e.g. Seyfarth et al. 1980).

Table 10.3 Summary of biological examples in which successful actions are triggered by indicator representations

Example	Action	T_t: success relevance	T_c: indicator
Monkey	Locomotion	Toxicity	Colour
Tiger	Attack	Nutritional value	Visual configuration
Beaver	Locomotion	Danger of bodily harm	Visual/auditory configuration
Magnetotactic bacteria	Locomotion	Oxygen	Magnetic north
Frog	Feed	Nutritional value	Visual configuration
Toad	Feed	Nutritional value	Visual configuration
Sickle-back male	Defend	Competitor for resources	Visual configuration
Sickle-back female	Court	Fertility of mate	Visual configuration
Vervet	Locomotion	Danger of bodily harm	Alarm call

Each of these examples (summarized in Table 10.3) can be mapped upon the architecture as presented in Fig. 10.4: There is a success-relevant variable T_t for which there is no direct sensor that T_t would directly causally affect.[9]

However, there is a sensor X_c for something else, an indicator feature T_c, which is frequently correlated with T_t. In each of these cases, the class of objects triggering an action ('real things plus dummies') is greater than the class for which the reaction leads to success ('only real things'). Snapping at small dark moving objects does not feed the frog—snapping at flies does. Also, in all of these cases, in the environment of evolutionary adaptiveness, frequent presence of the indicator coincides with the presence of the success-relevant object. Importantly, the success of actions triggered by R_c does depend solely on T_t, nor on T_c. To refer to the main example of this chapter, the success of running away only depends on the snake's toxicity—not on its colour.

In sum, the present setup exemplifies a rather large class of cases, where a success-relevant variable is picked up by an organism only indirectly, using an indicator variable and thus the consequences for success semantics, as developed below, are relevant for a rather large class of cases as well.

10.3.3 Optimal Mutations

As outlined above, in the presence of information loss between the target domain and sensory evidence, and if costs of misrepresentations are unequal or the a priori probabilities within the target domain are not uniform, average utility can be optimally increased when there are frequent misrepresentations. In the setup discussed until now (e.g. Godfrey-Smith 1991; see Fig. 10.2) misrepresentations could be introduced at the transformation of sensory input into a representation. In the presence

---

[9] In the following, we keep the indices c for indicator and t for success-relevant representations maintaining continuity with the main example without implying restriction to that specific example.

of indicator features/representations, there are two possible loci, where misrepresentations can be introduced: first, as in the simpler setup, at the transition to R_t, or second at the transition from X_c to R_c. The relation to success semantics of these two possibilities will be discussed in turn.

10.3.3.1 $\mathbf{R_t}$ Shift

As in the absence of an indicator representation, introduction of misrepresentations between the representation R_c of the indicator T_c, and the representation R_t of the success-relevant domain T_t, can increase average utility: Treating even slightly greenish snakes as toxic may be unsuccessful in each specific instance, but on average, utility can be maximised, because less errors of the costly type (trying to eat a toxic snake) are less frequent than errors of the harmless type (running away from a non-toxic snake). In this case, success semantics still holds, because each specific action is successful iff R_t is true.

10.3.3.2 $\mathbf{R_c}$ (Indicator) Shift

Crucially, consider that there are genetic variants of the monkey, shifting its colour representation R_c. These variants, when encountering a grey snake, would believe it to be of a certain shade of blue, in case R_c is shifted towards a bluish bias, or would believe it to be of a certain shade of green, in case R_c is shifted towards a greenish bias. The saturation of R_c when encountering a grey snake would correspond to the colour bias ranging from some negative to some positive constant. Which one of these variants would have the greatest fitness, the highest number of surviving offspring? Let us first consider a slight greenish bias. Then, slightly bluish snakes ($T_c < 0$) would lead to greenish colour representations ($R_c > 0$). Such a monkey would frequently try to eat toxic snakes—especially more frequently than its relative which has an unbiased colour representation. Second, let us consider a slight bluish bias. Such a monkey would frequently believe a slightly greenish (and thus harmless) snake to be blue, and thus try to evade it. However, it is conceivable that due to this slight misrepresentation of colour, it makes more mistakes trying to eat slightly bluish snakes, as its colour representation R_c amplifies the actual blue colour of the snake (T_c). Even though it thus misses more slightly greenish snakes, overall, it should have a higher accumulated utility than its unbiased relative, because mistaking a green colour for blue has more harmless consequences, than mistaking a blue snake for green. Further mutations, which have a very strong blue bias, would try to evade nearly every snake, except the deep green ones. Although such a monkey would never get bitten by a toxic snake, it also would fail to feed on the harmless snakes. Consequently, there seems to be some slight bias of bluish tint for R_c, which would have the highest fitness value: higher than unbiased mutations and higher than mutations with an even stronger bias of blue. Before quantitatively demonstrating that this argumentative, qualitative reasoning is indeed correct within a computational, evolutionary simulation, let's explore the consequences for success semantics.

10.3.4 Consequences for Success Semantics

For the optimal misrepresentation of colour, i.e. grey having a slightly bluish tint, so far, success semantics still holds, because misrepresentations of R_c in these cases leads to misrepresentations of R_t and consequently to unsuccessful actions, in each specific instance. Crucially, however, as both the indicator feature T_c and its representation R_c are continuous variables, there are cases, where blue snakes are misrepresented as bluer than they actually are, and green snakes as less green than they are. In both of these cases, toxic snakes lead to a successful evasion action and non-toxic snakes are eaten. That is, even though the colour representation is false, the resulting action is successful, in a systematic way—violating the core assumption of success semantics (10.3) that in order to systematically cause a successful action, a representation has to be true.

Let's be even more concrete. The misrepresentation of colour can be quantified with a value of *b,* such that the content of the colour representation is that of the actual colour, shifted by *b:*

$$cont(r_c)=t_c+b,$$

for a representation $r_c \in R_c$. That means, a completely desaturated colour, $t_c=0$, leads to a representation $r_c=b$, which is true in case $b=0$, and which corresponds to saturation corresponding to the absolute value $|b|$ of b of green in case $b>0$ and of blue in case $b<0$. Consequently, colours in the range of $-b<t_c<b$ lead to unsuccessful misrepresentations R_c, because green snakes are mistaken as blue, and vice versa, and consequently toxic snakes are treated as non-toxic, and vice versa. However, for snakes of colour $|t_c|>b$, even though the colour representation is wrong (indeed wrong by the amount b), green snakes are treated as harmless and blue snakes as toxic. The exact value of b of course depends on the probability distribution of T_t, the asymmetry of the utility matrix and the amount of information loss between T_c and X_c. The following section presents an evolutionary simulation of the monkey/snake example in order to validate the qualitative argumentation presented above.

10.4 Simulation

10.4.1 Description of the Situation

The following simulation will be used to illustrate the impact of information loss (e.g. sensory noise), asymmetry of the utility matrix and misrepresentation on the utility of an action. The scenario of monkeys and snakes given above will be used as illustration. Recall that the monkey can encounter *non-toxic snakes,* which can be eaten and increase the monkey's reproductive success. The *toxic snakes* attack monkeys that try to eat them with a poison that makes them infertile for some time. The skin colour can be used as a cue for the type of a particular snake: The skin of healthy snakes has a slight green tint, and the skin of the toxic snakes has a slight blue tint (see Fig. 10.4).

Some monkeys are born with a minor genetic defect that causes them to misperceive the colour spectrum in a way that everything gets a blue tint. In this case, white will be perceived as blue, which in the case of snakes is a conservative perception (the monkey rather runs away from a white healthy snake than getting bitten). The simulation aims at demonstrating the impact of sensory noise and the amount of misrepresentation on the fitness of the monkeys.

10.4.2 Method and Results

As an indicator for the monkey's fitness, the number of expected descendants is computed ('lifetime reproductive success'; Abrams 2012). The following numbers were used for the simulation:

- Maximum life duration of a monkey: 5 years
- Average number of descendants in the monkey population: one per year
- Encounters with a snake (of unknown type): once a year
- Probability of encountering a toxic snake: 50 %
- Trying to eat a toxic snake leads to -1 descendants in that year (resp., in the "strong toxic" case, to -2 descendants)
- Eating a healthy snake increases the number of descendants in this year by 0.5

The simulation had two experimental factors. On the one hand, four subpopulations of monkeys with different sensory fidelity were simulated by adding Gaussian noise with $M=0$ and $SD=0, 0.1, 0.2, 0.3$ to the true colour signal.[10] On the other hand, the level of misrepresentation was varied by inducing a sensory shift in R_c ranging from $b=-0.6$ to $b=0.2$. Negative levels correspond to the induction of a blue tint, leading to the misperception that a whitish healthy snake (with a colour value of 0.05, for example) is perceived as a toxic snake.

For each experimental condition, the life courses of 10,000 monkeys were simulated and the average number of descendants computed as a measure of fitness. The results are displayed in Fig. 10.5. The *x*-axis shows the amount of colour misrepresentation, the *y*-axis the fitness and different line types show different levels of noise. The left panel shows a situation with high asymmetry of utility, and the right panel low asymmetry. The asterisk marks the optimal sensory shift for each level of sensory noise.

One can clearly see that for noise levels 0 the optimal fitness is achieved with a misrepresented colour. Increasing noise levels lead to increasing shifts in sensory perception. Not surprisingly, the overall maximum of reproductive success is achieved with the most accurate perception (lowest noise). Hence, selection pres-

---

[10] The general conclusions do not depend on the specific numbers used in the simulation.

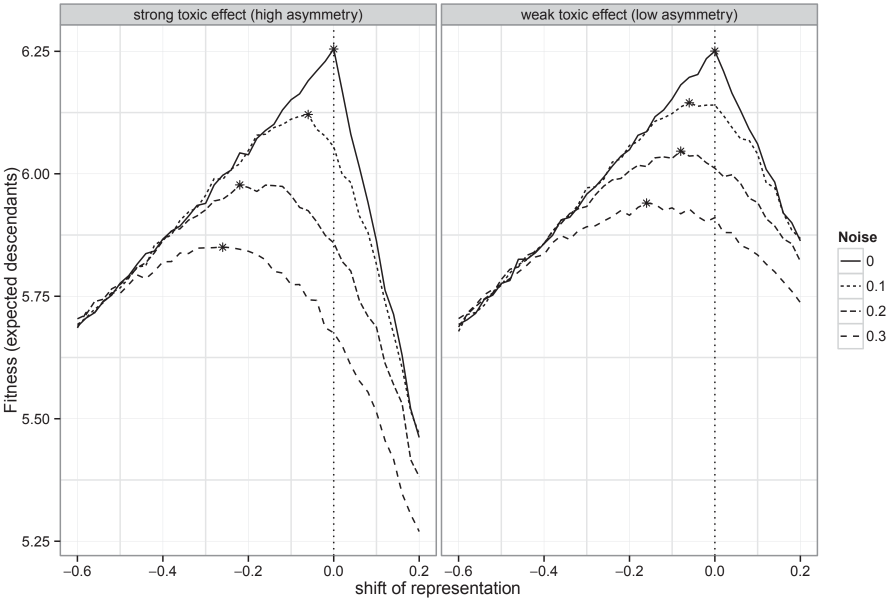


Fig. 10.5 Average fitness of monkeys, depending on sensory accurateness (*noise*), *shift of representation* and asymmetry of utility matrix (strong vs. weak toxic effect). The *asterisk* marks the optimal shift of representation for each noise level

sure should favour better sensors, and better sensors should go along with a more accurate representation.

Given an asymmetric utility matrix and noise, fitness is optimized at some degree of misrepresentation 0. Increasing noise levels are compensated by increasing shifts of representation.

Concerning the *truth* of a representation R_t, however, increasing levels of misrepresentation lead to a decreasing number of correct action, as can be seen in Fig. 10.6. In contrast to the utility, the number of wrong decisions is independent of the utility matrix and therefore symmetric around the zero point. With negative shift values, costly errors (missed toxic snakes) are reduced, but the less costly errors are disproportionately increased, leading to an increased overall level of false representations R_t.

In other words, if the representational shift is treated as the free parameter, one can predict that at least three factors lead to stronger misrepresentations: higher sensory noise, stronger asymmetry in the utility matrix and a higher prior probability of encountering the more detrimental world state.[11]

---

[11] The last point is not shown in the simulation. Increasing the probability of encountering a toxic snake to values greater than 50 % leads to lower overall levels of fitness and to a greater shift of representation. Probabilities smaller than 50 % have the reverse effect.

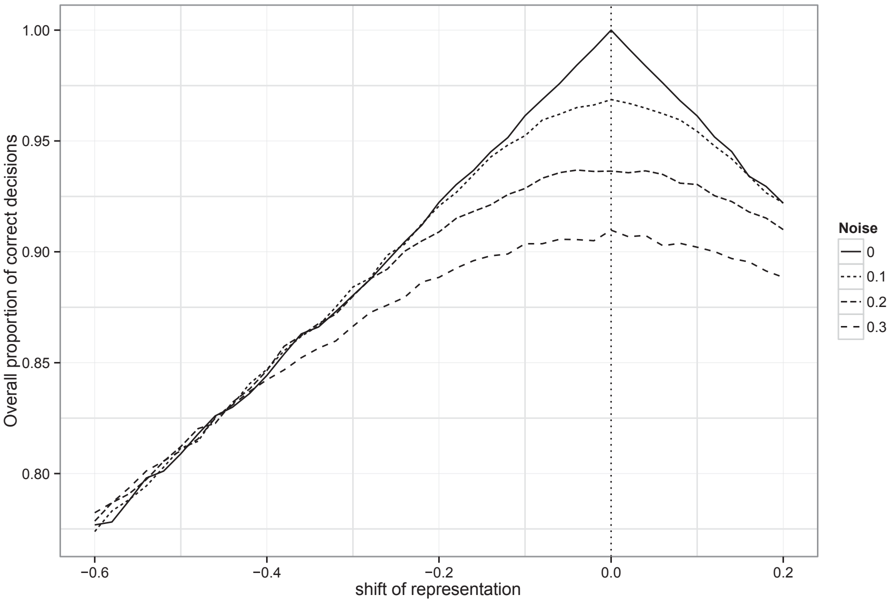


Fig. 10.6 Overall number of correct decisions, depending on the amount of shift of representation

10.5 Discussion

First, we discuss the generality of our proposal, specifically prerequisites and examples from illusions. Second, we discuss the relation to evolutionary epistemology, specifically as presented in Vollmer (1975) and Bischof (2009).

10.5.1 Generality

10.5.1.1 Omnipresence of Indicator Representations

The success of a false representation is not confined to the present example. Rather, it is a general phenomenon based on a generic principle. For the case without indicator variables (Fig. 10.2), Godfrey-Smith (1991) and Bischof (1998) have demonstrated analytically, based on signal detection theoretic (Green and Swets 1966) and game-theoretic (von Neumann and Morgenstern 1947) considerations, respectively, that average utility can be maximised for systematically false representations. These false representations, however, in specific instances are in line with success semantics. However, when the success-relevant feature has no direct causal impact on the organism's sensory system, indicator variables are used to form intermediate indicator representations, and false representations of the indicator variables can cause successful actions.

It seems to be rather easy to find examples, where indicator representations are used, i.e. $T_c \neq T_t$ (Fig. 10.3), compared to finding cases where the success-relevant feature is directly sensed (Fig. 10.2). Even in very simple organisms, for which the complete set of sensors, the set of behavioural repertoire and most of the internal neuronal wiring is known, such as nematodes (*Caenorhabditis elegans*; Faumont et al. 2012), the sensory modalities such as touch to different parts of the organism's shape, chemotaxis (oxygen or carbon dioxide) or temperature are not the success-relevant features for different sets of behavioural actions, such as ingestion, defecation, feeding or escape. The only example we were able to think of, where the success-relevant variable is identical with the indicator feature (i.e., $T_t = T_c$), is phototaxis in photosynthetic organisms (for a review, see Jékely 2009). Phototaxis is a 'behavioral migration response of an organism toward a change in illumination regime' (Hoff et al. 2009, p. 25). Positive phototaxis is a migration towards the light source, which is a successful action for photosynthetic organisms. It seems that apart from photosynthetic organisms, light sensors (such as eyes) rather generally produce indicator representations (similar to sound waves picked up by ears, or odours picked up by olfactory sensors).

10.5.1.2 The Role of Learning

Now, a large number of the indicator examples presented in the present chapter could be summarised as fixed action patterns (instincts) where a sign stimulus or releaser signal acts as an indicator to trigger a certain action which is adaptive for situations that frequently correlate with the indicator in the environment of evolutionary adaptiveness (e.g. Tinbergen 1951; Lorenz 1937; for a review see Schleidt 1962). Beyond such fixed (innate) or acquired fixed action patterns, in associative learning in classical or operant conditioning, arbitrary coupling of any T_t with nearly any T_c (for limits see Breland and Breland 1961) can be generated by an organism. For instance, drooling in dogs is an appropriate (successful) response to the presence of food, but ringing a bell (as an indicator feature; Pavlov 1927/2010) has no influence on drooling's success. In general, learning allows ontogenetic adaptation to the statistical properties of dynamically changing and unforeseen configurations of the environment. Thus, a flexible mechanism would possibly be implemented. In case misrepresentations in such learned contexts are optimal (i.e. in the presence of asymmetric priors or costs), misrepresentations should presumably happen between R_c to R_t. In the size version of our snakes example (Fig. 10.4, right panel), a monkey could learn that small snakes are toxic and treat snakes as toxic up to a size of 25 cm, even though the true cut-off was 20 cm. These cases would be no challenge to success semantics. However, the present chapter aims at demonstrating that there are cases where there is little flexibility of changes in R_c.

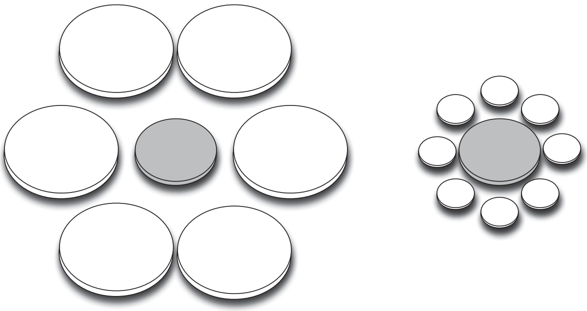


Fig. 10.7 The apparent size of the central disc depends on the size of the surrounding disc, even though it is of the same size in both cases (when measured with a ruler)

10.5.1.3 The Case of Illusions

The examples provided in this chapter mainly stem from the animal kingdom. Thus, two examples of misrepresentations are presented to demonstrate that systematic misrepresentations stably exist in humans for primary qualities (i.e. size and location), and can serve multiple action purposes: the Ebbinghaus size illusion and visual capture of sound.

In the Ebbinghaus (or Titchner) illusion, the relevant physical property is a disk's size S and a corresponding representation of size R_S in humans (see Fig. 10.7). Specifically, the representation of size depends on the central disk's context: In case the central disk is surrounded by large circles, it is represented as being smaller and in case of small disks in the surrounding as being larger. The surrounding changes the size representation in the range of 5–10 % (cf. Franz et al. 2000).

The (mis-)representation of size, R_S, is a multipurpose representation that is causally involved in several different tasks/actions (Aglioti et al. 1995). Concerning perception, the context determines how big humans *see* the disk. When asked to *match* another disk to the size of the central disk ('perceptual matching'), humans over- or underestimate the disk size. The same happens when they are asked to *show* the disk size with their thumb and finger. And finally, when asking to *grasp* the central disk, the maximal grip aperture, that is the maximal opening of thumb and index finger when the hand is en route to the target, is also modulated by context (for a review, see Franz 2001). Physical size linearly affects maximal grip aperture as well (Castiello 2005). Finally, even though the wrong representation of size, R_S, has a causal influence on the grasping action, it is successful in picking up the central disk. Thus, it seems possible to successfully grasp a disk even though the representation of its size is false.

The second systematic misrepresentation in humans presented here concerns the primary quality of location in space. Consider watching TV or a movie where a person speaks. Then, the mouth has a certain location in space. Simultaneously with the lip movements, speech utterances are audible. What usually happens is that the speech is perceived as originating from the speaker's mouth, even though it actually originates from loudspeakers to the left and right of the screen. Here, there is a false representation of the sound's origin in space, specifically, the sound's origin is mislocalized to the location of the moving lips (which also allows ventriloquists to do their trick). In general terms, vision's representation of spatial origin captures the auditory representation of spatial origin (e.g. Pick et al. 1969; Warren et al. 1981). Recently, it has been demonstrated that sound can capture vision if sensory noise in vision is increased (Alais and Burr 2004) in such a way that the integration of conflicting information about an event's location of vision and space are combined optimally, depending on the levels of noise in both modalities (Ernst and Banks 2002).

These examples foremost serve the purpose of demonstrating that systematic misrepresentations are present in humans for primary qualities (such as size and location) even in situations where the respective representations are causally involved in a wider range of different types of actions. However, these examples also serve a second purpose: Based on the conceptual framework of the present chapter, it is possible to ask why such illusions (i.e. systematic misrepresentations) have stably evolved in human evolution. Remember, we could demonstrate that misrepresentations maximise fitness under certain conditions: (1) presence of sensory noise, (2) asymmetry in priors or cost of errors and possibly (3) the presence of indicator representations. Consequently, the framework presented here provides a heuristic for asking, whether these prerequisites are met in the case of the presented illusions. Note that it is possible that the Ebbinghaus or ventriloquist misrepresentations are not themselves maximising fitness but caused by a different organetic or other constraint.

10.5.2 Evolutionary Epistemology

In the previous sections, we argued that there can be situations where our representations systematically deviate from an objective reality. From a first-person point of view, however, it is not evident for a specific situation whether we misrepresent or not, as has been illustrated by several illusions. Given this situation, we can ask: What can we know about the world at all? Or, put in other words: What are the epistemological consequences of misrepresentations?

Evolutionary epistemology[12] (Lorenz 1973; Vollmer 1975) maintains the idea that our perception and representations have adapted to the (hypothesized) real world. From this general point of view, several deductions can be made. On the one hand, our representations can be expected to be quite reliable and objective in domains which are highly fitness-relevant. From the same theoretical framework, one can also conclude that our representations will not be perfect: 'In evolution, that is under competition, it pays to recognize outside objects more or less correctly. But it would not pay to aim at or to reach perfection' (Vollmer 2010, p. 1652). Finally, for domains which are not fitness-relevant at all or where no potentially success-

---

[12] The term 'evolutionary epistemology' has been used in at least two different ways (Bradie 1986). Popper and others used the term to describe the growth of human knowledge by the (non-genetic) evolution of ideas and theories (Popper 1972). In this chapter, we use it only in the sense of Lorenz (1973) and Vollmer (1975).

ful actions can be performed by the organism, no selection pressure existed which would have shaped humans' sensory abilities or representations towards objectivity.

Based on this general framework, several conclusions concerning misrepresentations have been drawn by Vollmer (1975) and Bischof (2009), as will be shown in the next paragraphs.

10.5.2.1 Micro-, Meso- and Macrocosm

Domains of human knowledge which somehow relate to fitness-relevant domains of the external world can be shaped by evolution towards objectivity (Vollmer 1975). There are, however, domains of knowledge which are completely unrelated to the external world (e.g. mathematical symbolic systems). There is no way to falsify or verify representations and beliefs within such symbolic systems by recurrence to the external world. Likewise, there are physical domains of the external world which have not been fitness-relevant in our history of evolution (e.g. strong atomic radiation). Without selection pressure, no detectors could have evolved for such physical domains, even if they have become fitness-relevant nowadays.

Fitness-relevant physical phenomena are predominantly located in a rather narrow range of physical scales: The retina is only sensitive to a small band of electromagnetic frequencies, size estimates below 0.5 mm and above some kilometres are nearly impossible, and time spans of nanoseconds or geological history are hard to imagine.

Based on this observation, Vollmer (1975) categorized the physical phenomena into three 'cosms', (a) *microcosm,* which subsumes phenomena that are on a too small scale to be fitness-relevant (e.g. subatomic structures), (b) *mesocosm,* which describes phenomena on a medium scale and (c) *macrocosm*, which describes very large physical scales (e.g. cosmologic dimensions). Whenever knowledge domains exceed the mesocosmic scales to which our sensory apparatus and representational categories are adapted, these categories might be suboptimal or misleading. As a consequence, our intuitive sense of such phenomena can lead us astray. For example, it is hard to grasp the wave-particle dualism of electrons, even for experienced physicists.

To summarize, Vollmer's framework of micro-, meso- and macrocosm gives some guidelines for where to expect objectivity in human perception and in which domains our adapted senses and categories might be bad guides.[13]

---

[13] Note that an analogue distinction has also been proposed by Bertalanffy: 'The popular forms of intuition and categories, such as space, time, matter and causality, work well enough in the world of "medium dimensions" to which the human animal is biologically adapted. Here, Newtonian mechanics and classical physics, as based upon these visualizable categories, are perfectly satisfactory.' (Bertalanffy 1968, p. 241, see also Bertalanffy 1955).

10.5.2.2 Meta-, Para- and Orthocosm

Bischof shares the general framework of evolutionary epistemology with Vollmer, but he suggested an alternative classification of the ‘cosms’. According to Bischof (2009), it is not necessary to distinguish micro- from macrocosm. Both categories are completely irrelevant to fitness and therefore can be combined into a single category which he calls *metacosm*. The metacosm describes all phenomena which are fitness-irrelevant, beyond their location on a physical scale. For example, the question of whether ‘Beauty and Truth are the same’ is hardly fitness-relevant, and it cannot be expected that encounters with the real world can give any evidence for or against that idea.

Furthermore, Bischof proposed to divide the mesocosm based on the symmetry of the utility surface (given that evolution does not favour ‘over-optimal’ sensors, it is assumed that always some level of noise is present). The *orthocosm* describes all phenomena with a symmetric surface—representational errors on both sides of the ridge have more or less the same costs in terms of fitness. For these phenomena, representations can be assumed to be shaped towards truth, as objective representations are optimal. Although we can never be sure whether we have *reached* objectiveness (Vollmer 2010), we can expect a convergence towards objectiveness. The *paracosm,* in contrast, denotes phenomena with asymmetric utility surfaces. Based on the arguments given above, paracosm representations are expected to converge on a conservative level, shifted away from objectivity, as this is optimal. Bischof locates most social categories in the realm of paracosm, whereas domains which require a physical interaction with the external world, like tool usage, predominantly are in orthocosm. For a graphical comparison of Vollmer’s (1975) and Bischof’s (2009) categories of evolutionary epistemology, see Fig. 10.8.

10.5.2.3 Increasing Objectivity: Multidimensional Utility Surfaces and Measurement Invariance

Both Vollmer and Bischof emphasize that multidimensional utility surfaces (‘utility hyper surfaces’) usually lead to more symmetry and consequently to more objectivity. In our example of monkeys and snakes it is possible that a shifted colour perception (which is optimal for snake encounters) has adverse impacts in other fitness-relevant domains, for example ‘finding edible fruits’. If this is the case, the overall utility hypersurface is a weighted average of all fitness-relevant tasks that make use of this particular sensor. This multidimensional utility hypersurface typically (but not necessarily) is more symmetric than the utility surface of a single dimension.

As evolution shaped our cognitive apparatus towards optimality and not towards objectivity, representations of paracosmic phenomena are not objective. But how can we, at least, approach objectivity for the paracosmic domain? Both Vollmer and Bischof agree that additional sensors can increase the objectivity of a representation. If multiple methods of measurement converge on the same result this would be a sign of *measurement invariance* (Bischof 1966; Vollmer 2010). Our monkeys

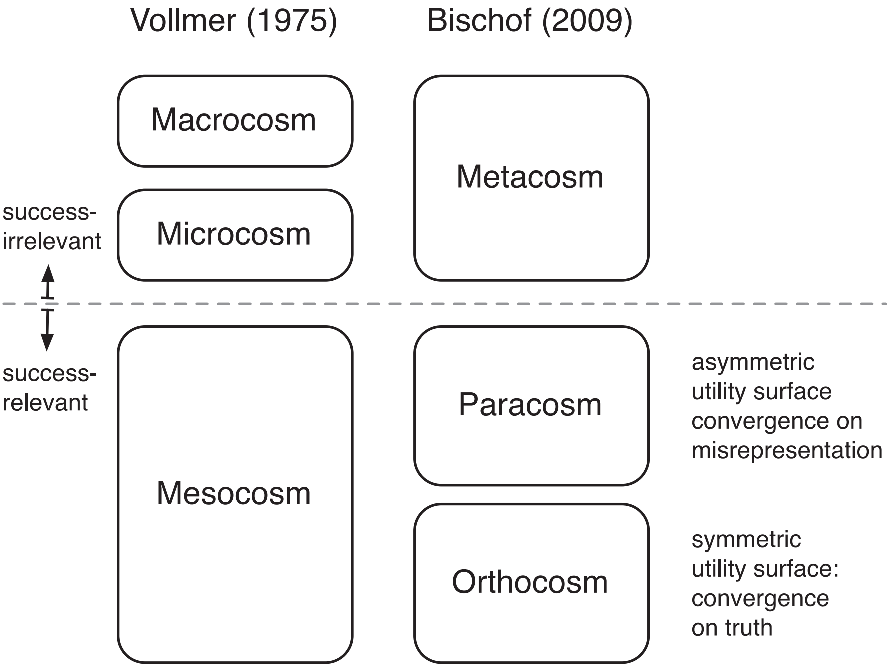


Fig. 10.8 A graphical representation of Vollmer's (1975) and Bischof's (2009) epistemological categories

have variance in colour perception in comparison to other measurement methods: Spectrometer measurements of the snake's colour, or the perception of other species that are immune to these snakes, would not converge with our monkey's perception. For Vollmer (2010), measurement invariance actually is the defining criterion for objectivity: 'A proposition is objective if and only if its meaning and its truth is invariant against a change in the conditions under which it was formulated, that is, if it is independent of its author, observer, reference system, test method, and conventions' (p. 1658).

Hence, probing the colour perception of our monkeys with additional devices, independent of any toxicity associations, can be the tool to assess the objectivity of a representation (Vollmer 2010). Although we can never be sure whether we have reached objectivity, invariance is the touchstone that tests for objectivity.

10.6 Conclusion

As has been argued before, under some conditions, namely information loss plus either asymmetry of the utility surface or prior probabilities, misrepresentations (or criterion shifts) lead to optimal actions. Success semantics state that a representation is true iff it causes a successful action. When misrepresentations are optimal, does that violate success semantics? Not necessarily. Although the average success is optimized for shifted representations, for each single action still holds that only true representations lead to successful actions.

In contrast to that situation, however, we argued that there are conditions in which false representations systematically lead to successful actions—not only in average but also in single instances. Specifically, whenever (beyond information loss and asymmetry/different priors) the organism employs indicator representations as proxies for the actual success-relevant feature, systematically false representations in the indicator variable can lead to successful actions, and this situation indeed *is* a violation of success semantics. We provided examples suggesting that the usage of indicator variables is probably more the rule than the exception in living organisms.

Embedding these ideas in the context of evolutionary epistemology, it can be assumed that humans have knowledge domains that converge to the truth (‘orthocosm’), because the overall utility hypersurfaces are symmetric. However, humans might also have knowledge domains where representations systematically deviate from the truth (‘paracosm’), for asymmetric selection pressure.

Acknowledgments We thank W. Schleidt for valuable comments on previous versions of the manuscript and E. Ratko-Dehnert, A. Hetmanek, J. Jarecki, M. Rausch, B. Schlagbauer, and B. Ruf for fruitful discussions on the topic.

References

Abrams, M. (2012). Measured, modeled, and causal conceptions of fitness. *Frontiers in Genetics, 3,* 1–12. doi:10.3389/fgene.2012.00196.

Aglioti, S., DeSouza, J. F., & Goodale, M. A. (1995). Size-contrast illusions deceive the eye but not the hand. *Current Biology: CB, 5*(6), 679–685.

Alais, D., & Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Current Biology: CB, 14,* 257–262. doi:10.1016/j.cub.2004.01.029..

Bazylinski, D. A., & Frankel, R. B. (2004). Magnetosome formation in prokaryotes. *Nature Reviews Microbiology, 2*(3), 217–230. doi:10.1038/nrmicro842..

Baube, C. L., Rowland, W. J., & Fowler, J. B. (1995). The mechanisms of colour-based mate choice in female threespine sticklebacks: Hue, contrast and configurational cues. *Behaviour, 132*(13–14), 13–14.

Bertalanffy, L. (1955). An essay on the relativity of categories. *Philosophy of Science, 22,* 243–263.

Bertalanffy, L. (1968). The relativity of categories. In L. Bertalanffly (Ed.), *General system theory* (pp. 222–250). New York: George Braziller.

Bischof, N. (1966). Erkenntnistheoretische Grundlagenprobleme der Wahrnehmungspsychologie. In W. Metzger & H. Erke (Eds.), *Handbuch der Psychologie in 12 Bdn. Bd. 1/I: Wahrnehmung und Bewusstsein* (pp. 21–78). Göttingen: Psychologie.

Bischof, N. (1998). *Struktur und Bedeutung*. Bern: Hans Huber.

Bischof, N. (2009). *Psychologie: Ein Grundkurs für Anspruchsvolle[Psychology: A basic course for the ambitious]* (2nd ed.). Stuttgart: Kohlhammer.

Blackburn, S. (2005). Success semantics. In H. Lillehammer & D. H. Mellor (Eds.), *Ramsey’s legacy* (pp. 22–36). Oxford: Oxford University Press.

Blakemore, R. P. (1975). Magnetotactic bacteria. *Science, 190*(4212), 377–379.

Borchers, H.-W., Burghagen, H., & Ewert, J.-P. (1978). Key stimuli of prey for toads (Bufo bufo L.): Configuration and movement patterns. *Journal of Comparative Physiology, 128*(3), 189–192.

Bradie, M. (1986). Assessing evolutionary epistemology. *Biology and Philosophy, 1,* 401–459.

Breland, K., & Breland, M. (1961). The misbehaviour of organisms. *American Psychologist, 16,* 681–684.

Castiello, U. (2005). The neuroscience of grasping. *Nature Reviews Neuroscience, 6*(9), 726–736. doi:10.1038/nrn1744.

Dretske, F. (1981). *Knowledge and the flow of information*. Cambridge: MIT/Bradford Press.

Dretske, F. (1986). Misrepresentation. In R. Bogdan (Ed.), *Belief* (pp. 17–36). Oxford: Oxford University Press.

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature, 415,* 429–433. doi:10.1038/415429a.

Ewert, J.-P. (1974). The neural basis of visually guided behaviour. *Scientific American, 230*(3), 34–42.

Faumont, S., Lindsay, T. H., & Lockery, S. R. (2012). Neuronal microcircuits for decision making in *C. elegans*. *Current Opinion in Neurobiology, 22*(4), 580–591. doi:10.1016/j.conb.2012.05.005.

Fodor, J. A. (1990). *A theory of content and other essays*. Cambridge: MIT Press.

Fox, C. W., & Westneat, D. F. (2010). Adaptation. In D. F. Westneat & C. W. Fox (Eds.), *Evolutionary behavioral ecology* (pp. 16–32). New York: Oxford University Press.

Franz, V. H. (2001). Action does not resist visual illusions. *Trends in Cognitive Sciences, 5*(11), 457–459.

Franz, V. H., Gegenfurtner, K. R., Bülthoff, H. H., & Fahle, M. (2000). Grasping visual illusions: No evidence for a dissociation between perception and action. *Psychological Science, 11,* 20–25.

Godfrey-Smith, P. (1991). Signal, decision, action. *The Journal of Philosophy, 88*(12), 709–722.

Gold, J. I., & Shadlen, M. N. (2007). The Neural Basis of Decision Making. *Annual Review of Neuroscience, 30*(1), 535–574. doi:10.1146/annurev.neuro.29.051605.113038.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.

Hoff, W. D., van der Horst, M. A., Nudel, C. B., & Hellingwerf, K. J. (2009). Prokaryotic phototaxis. *Methods in Molecular Biology, 571,* 25–49. doi:10.1007/978-1-60761-198-1_2.

Hubel, D. H., & Wiesel, T. N. (1974). Sequence regularity and geometry of orientation columns in monkey striate cortex. *Journal of Comparative Neurology, 158*(3), 267–294.

Jékely, G. (2009). Evolution of phototaxis. *Philosophical Transactions of the Royal Society B: Biological Sciences, 364*(1531), 2795–2808. doi:10.1098/rstb.2009.0072.

Lettvin, J. Y., Maturana, H. R., McCulloch, W. S., & Pitts, W. H. (1968). What the frog's eye tells the frog's brain. In W. C. Cunning & M. Balaban (Eds.), *The mind: Biological approaches to its functions* (pp. 233–258). New York: Wiley.

Lorenz, K. (1937). Über die Bildung des Instinktbegriffes. *Die Naturwissenschaften, 25*(19), 289–300. doi:10.1007/BF01492648.

Lorenz, K. (1973). *Die Rückseite des Spiegels: Versuch einer Naturgeschichte menschlichen Erkennens*. Munich: Piper.

Millikan, R. G. (1989). Biosemantics. *The Journal of Philosophy, 86*(6), 281–297.

Nanay, B. (2013). Success semantics: the sequel. *Philosophical Studies, 165*(1), 151–165. doi:10.1007/s11098-012-9922-7.

Neumann, J., & Morgenstern, O. (1947). *Theory of games and economic behavior*. Princeton: Princeton University Press.

Papineau, D. (1984). Representation and explanation. *Philosophy of Science, 51,* 550–572.

Papineau, D. (2003). Is representation rife? *Ratio, 16*(2), 107–123.

Pavlov, I. P. (1927/2010). Conditioned reflexes: An investigation of the physiological activity of the cerebral cortex. *Annals of Neurosciences,17*(3), 136–141. doi:10.5214/246.

Pick, H. L., Warren, D. H., & Hay, J. C. (1969). Sensory conflict in judgments of spatial direction. *Perception and Psychophysics, 6*(4), 203–205. doi:10.3758/BF03207017..

Popper, K. R. (1972). *Objective knowledge: An evolutionary approach*. Oxford: Clarendon.

Ramsey, F. P., & Moore, G. E. (1927). Symposium: Facts and Propositions. *Proceedings of the Aristotelian Society, Supplementary Volumes, 7,* 153–206.

Reeve, H. K., & Sherman, P. W. (1993). Adaptation and the goals of evolutionary research. *Quarterly Review of Biology, 68,* 1–32.

Schleidt, W. M. (1962). Die historische Entwicklung der Begriffe “Angeborenes auslösendes Schema”, und “Angeborener Auslösemechanismus” in der Ethologie. *Zeitschrift für Tierpsychologie, 19,* 697–722.

Seyfarth, R., Cheney, D., & Marler, P. (1980). Monkey responses to three different alarm calls: Evidence of predator classification and semantic communication. *Science, 210*(4471), 801–803. doi:10.1126/science.7433999.

Shea, N. (2007). Consumers need information: Supplementing teleosemantics with an input condition. *Philosophy and Phenomenological Research, 75*(2), 404–435.

Stephens, D. W., & Krebs, J. R. (1987). *Foraging Theory*. Princeton: Princeton University Press.

Tinbergen, N. (1951). *The study of instinct*. Oxford: Clarendon.

Usher, M. (2001). A statistical referential theory of content: Using information theory to account for misrepresentation. *Mind & Language, 16*(3), 311–334.

Vollmer, G. (1975). *Evolutionäre Erkenntnistheorie [Evolutionary epistemology]*. Stuttgart: Hirzel.

Vollmer, G. (2010). Invariance and objectivity. *Foundations of Physics, 40,* 1651–1667. doi:10.1007/s10701-010-9471-x.

Warren, D. H., Welch, R. B., & McCarthy, T. J. (1981). The role of visual-auditory “compellingness” in the ventriloquism effect: Implications for transitivity among the spatial senses. *Perception & Psychophysics, 30*(6), 557–564. doi:10.3758/BF03202010.

Whyte, J. T. (1990). Success semantics. *Analysis, 50*(3), 149–157.

Chapter 11
Human Sacrifice and the Evolution of Thinking: A Critical Assessment of Christoph Türcke's *Philosophy of Dreams*

Andreas Mayer

> *Tyger Tyger, burning bright, In the forests of the night;*
> *What immortal hand or eye, Could frame thy fearful symmetry?*
> William Blake

11.1 Introduction

Imagine our dreams were more than mere messages from the unconscious. Imagine they were windows into a long-ago past and something no archaeologist or palaeoanthropologist will ever find, no matter how modern and precise his radio carbon data are: hitherto-ignored prehistoric remains of how our hominid ancestors thought, what their psychic reality and subjective life was like. No bones and skulls, no tools and artefacts could ever tell us anything about this. What if our own dreams told us something about what was going on in *their* heads when they were awake? What if our way of dreaming resembles their way of thinking? If this were true, then our ancestors' subjective life would still live within us and resurrect every night as soon as we close our lids. Our own dreams, then, would be windows into prehistory, and the analysis of their basic structures and constitutive processes would amount to a kind of mental archaeology that could teach us not only about the processes that gave rise to human thinking but also about the context in which they came into being: in rituals of human sacrifice.

Christoph Türcke suggests precisely this. According to Türcke, the principles and forces that constitute our dreams as described by Freud (1961 [1900]) in his ground-breaking book *Die Traumdeutung* were also at work when the early *Homo* tried to cope with experiences of traumatic shock via traumatic re-enactment in the form of rituals of human sacrifice. In his book *Philosophy of Dreams* (2013), Türcke offers a brilliantly written account of seductive plausibility. Yet, there are

---

A. Mayer (✉)
Institute of Psychology, University of Osnabrück, Osnabrück, Germany
e-mail: andreasfrankmayer@gmx.de

T. Breyer (ed.), *Epistemological Dimensions of Evolutionary Psychology*,
DOI 10.1007/978-1-4939-1387-9_11

points that raise questions and finally cast doubt on the claim that rituals of human sacrifice opened up mental life as we know it. This chapter therefore provides a critique and suggests an alternative pathway in which human sacrifice is not as important for the rise of human mental life as Türcke suggests. Before we focus on any critique, however, a summary of his account is necessary. This summary will also make clear how dreams and human sacrifice relate to each other.

After that, I will try to critically assess Türcke's account in various steps. To begin with, I will try to assess the basic assumptions his account is built on. First, I will reflect on the assumption that only *Homo sapiens* was "nervous" enough to be affected by experiences of traumatic shock (ETS) in a way that forced him to traumatically re-enact and repeat these experiences in rituals of human sacrifice (RHS). Second, I will focus on Türcke's (2011) use and interpretation of the German word *Schrecken* (horror, terror, fright; manifestation of something which terrifies, frightens, shocks, scares).[1] In the next step, I will introduce Michael Tomasello's (1999, 2008) very different view on the origins of human cognition and communication. Yet, although different, both accounts are not incompatible. For example, one might argue that the abilities identified by Tomasello as being specific and constitutive for our species might well have developed in parallel with or even as a result of RHS. I will then suggest an alternative, arguing that ecological changes forced our human ancestors to collaborate in qualitatively new ways when they had to defend themselves against predators in the open savannahs and finally started to hunt big game as well. Finally, I will suggest—and here, I lapse into pure speculation—that the forces at work in our dreams, which Türcke postulated to be characteristic of both RHS and primitive (hallucinatory) thinking, can also be detected and identified in the alternative scenario in which adapting to the new environment of the savannah was of primary importance.

Before getting started, however, be warned. Any theory that tries to describe how mental life came into being stands on very shaky grounds. We are dealing here with speculations rather than empirically testable theories. Türcke himself pleads for the vindication of speculation in science. Since this chapter deals with a speculative account, the following pages are inevitably as speculative as the book they are dealing with. Hopefully, they are also as plausible.

11.2 *Philosophy of Dreams*—A Summary

Türcke's book is full of ideas, hints, speculations and associations—this makes *Philosophy of Dreams* a very exciting reading, but a difficult one when it comes to breaking it down into clear-cut scientific hypotheses. Still, some ideas recur, although they are sometimes slightly varied.

---

[1] Since the German term *Schrecken* captures aspects of both the English terms *fright* and *horror,* I will—to simplify matters—continue to use the German term as well as the corresponding adjective *schrecklich* throughout this text.

First, Türcke (2013, p. 92) argues that traumatic shocks are the strongest experiences human beings can make and that no other experiences have the power to arouse the nervous system in the same way.

Second, he argues that they were the only experiences with the power to arouse the nervous system in a way that forced early hominids to look for ways to cope with them and find relief. Since ETS are disturbing and difficult to get rid of, similar experiences were repeated and traumatically re-enacted—a desperate attempt of the organism to get them under control and to find relief (Freud's *traumatic repetition compulsion*).

Third, Türcke argues that we can only imagine things we have previously perceived as being enacted. In the German version of the book (*Philosophie des Traums,* 2011), Türcke plays with the word *Vorstellung,* which means both imagination and performance and which—in its verbal form (*vorstellen*)—can be read as *vor-stellen,* i.e. to put something in front of someone or something (so that it can be observed). According to Türcke, only procedures and actions that have been repeatedly performed can give rise to their mental representation and, hence, to the realm of imagination. Repetition, however, is not sufficient in his view. Only procedures and actions that are experienced in the same *intensity* as traumatic shocks can, via repetition, become internalized and thus give rise to mental representation and imagination.

Fourth, RHS are controlled ways of repeating and performing ETS.

Fifth, RHS represent the original ETS, which were much more unforeseeable and uncontrollable. For RHS to represent ETS, three steps are necessary. The original experience must be (1) displaced, (2) condensed, and (3) reversed. It must be *displaced*, i.e. shifted from specific places in time and space (where the original ETS has occurred) to the place where RHS take place. There, it must be *condensed*, i.e. the variety of possible ETS is reduced to a single phenomenon (human sacrifice), which can be structured and ritualized. The more chaotic elements of the original ETS event are thus eliminated. Finally, it must be *reversed:* The horror and shock that was experienced is now re-enacted in a safer environment and inflicted upon the sacrificial victim.

Sixth, since RHS is another—although controlled—experience of traumatic shock, early hominids sooner or later had to find a way to also get rid of the ritual itself. RHS, which were originally designed to cope with ETS, had become another traumatic aspect of life that had to be overcome. This happened via internalization and subsequent mental representation. Whereas RHS are very literally an internalization (in the spatial sense of the term) of ETS into the ritualistic centre, the whole course of action and its relieving function must now be internalized mentally and hence be transformed into a mental function. This is to say that the forces that helped RHS to represent ETS became the cognitive abilities that *displace, condense* and *reverse* mental representations. Hence, this form of primitive thinking was made up by exactly the same forces Freud discovered to be at work in the dreams of modern people.[2] Our dreams are therefore windows into the beginnings of thinking, which started with the internalization of repeatedly observed performances (RHS) powerful enough to arouse the nervous system of early Homo. *Vorstellungen* in the sense of performances or performative representations thus became *Vorstellungen* in the sense of mental representations.

Seventh, the first word(s) and thus human speech came into being in the context of ETS. According to Türcke, these words could hardly have been anything else than exclamations like “Here!” and/or “There!”—demonstratives accompanying sudden shock in the face of *Schrecken*.

Summing up, Türcke considers *condensation, displacement* and *reversal* to be not only the formative principles of dreaming—in his view, these principles are key to the evolution of thinking, culture and civilization, and were originally manifest in rituals of human sacrifice. From this perspective, the forces that shape our dreams night by night are the same as the forces that made us human.[3]

11.3 Early *Homo* as a Bag of Nerves

One of the first questions that come to mind when reading *Philosophy of Dreams* is “Why should only humans re-enact ETS by sacrificing conspecifics?” Türcke’s answer is simple and straightforward physiological: Humans were literally a “bag of nerves” and more aroused by traumatic shock because they had a larger brain and hence more nerves than any other species.

The problem with this assumption is, of course, that we do not know much about the early sensitivity and nervous vulnerability of early hominids. Türcke’s evidence comes from Freud’s observation of war survivors in the twentieth century. Freud was forced to postulate *thanatos* as a second drive since he was not able to explain the recurrent nightmares of traumatized soldiers in the terminology of *eros* and his psychosexual theory. But the mere fact that the nervous system of twentieth-century soldiers was so aroused by traumatic war experiences that they were compelled to repeat them in the course of the organism’s attempt to get rid of these memories does not necessarily tell us whether the nervous system of early hominids was as vulnerable as well. Türcke transfers the sensitivity of modern man and posttraumatic stress symptoms onto early hominids, thus ignoring the possibility that traumatic reactions of the kind demonstrated by Freud’s patients might already be—at least partially—the product of a long cultural history and process of civilization, in which

---

[2] This early thinking was, according to Türcke, mainly hallucinatory. Although the following pages will not deal with this idea and the notion of hallucination among early Homo sapiens, it is worth mentioning that hallucinations get in the way of clear perceptions of reality. Therefore, continuous hallucination in daytime can hardly have been adaptive.

[3] “My suspicion is that, taken together, condensation, displacement and the dark spot [Türcke refers here to *reversal;* AM] contain a lot more than the secret of dream construction—namely, nothing less than the secret of the construction of thought and culture and of the development of humanity” (Türcke 2013, p. 38).

sensitivity and vulnerability were gradually increased by normative forces, shared values, religious and humanistic ideas, empathic practices, which had been exercised for centuries and so forth. In doing so, Türcke's starting assumption that early *Homo* was the only animal sensitive enough to be forced by nature to traumatically re-enact ETS is based on nothing more than a physiological fact. Since our forebears' brain volume was bigger than the brain volume of other species, they must have been more sensitive and thus more shocked by terrifying experiences than others. However, although brain size is known to be correlated with, for example, intelligence, there are factors that correlate higher with it, such as the number of cortical neurons and conduction velocity (Roth and Dicke 2005). Yet, since humans possess only marginally more cortical neurons than whales and elephants—animals also known for having complex emotions like grief and psychological pain, but not for ritualistic sacrifices of conspecifics—and since even less is known about the relationship between brain volume and psychological vulnerability, Türcke's starting assumption is not fully convincing. What if it is nothing else than a projection of our current anthropocentrism onto our ancestors? Or, more precisely, our *speciocentrism:* Since we believe ourselves to be the most sensitive and emotionally vulnerable species on the planet, our ancestors must have been as well. But neither the premise nor the conclusion can be taken for granted. Elephants and dogs appear to suffer from traumatic experiences in ways comparable to human beings (Bradshaw 2004; Nagasawa et al. 2012). The *New York Times* recently reported that military dogs show signs of combat stress, indicating that they suffer from traumatic experiences undergone in war zones.[4] Military veterinarians are already speaking of canine posttraumatic stress disorder (PTSD) to refer to patterns of behaviour among dogs that were exposed to explosions, gunfire or other combat-related violence. The symptoms are strikingly similar to those of humans: hypervigilance, avoidant behaviour, changes in temperament, aggressiveness and so forth. According to this article, some dogs are even treated with the same medications used to treat panic attacks in humans. Moreover, dogs also dream, and there is no reason why traumatic experiences should not be repeated in their dreams as well. But we do not know for sure, simply because dogs cannot tell us. Yet, if military dogs should repeat traumatic experiences in their dreams, too, then traumatic repetition compulsion is not restricted to human beings and, hence, not necessarily to our forebears. More importantly, other species that do not sacrifice conspecifics in order to find relief would also be sufficiently "nervous" to (a) suffer from ETS, (b) develop symptoms similar to human PTSD and maybe (c) repeat the traumatic experiences in dreams.

Let us again have a look at the alleged relationship between brain volume and psychological vulnerability. About 4 million years ago, a variety of hominids known as *Australopithecus* lived in eastern and southern Africa. One subspecies, *Australopithecus afarensis,* was relatively closely related to *Homo* ("Lucy" being the most famous one). These early hominids were clearly bipedal, yet their brains "were little better than ape-sized" (DeSalle and Tattersall 2008, p. 118)

---

[4] http://www.nytimes.com/2011/12/02/us/more-military-dogs-show-signs-of-combat-stress.html?pagewanted=all&module=Search&mabReward=relbias%3Ar&_r=0

and “nothing in the record would allow us to infer that in early stages, at least, they had developed a level of intelligence significantly greater than that of our modern apes” (DeSalle and Tattersall 2008). Consequently, they were probably not sufficiently “nervous” to be forced to repeat ETS in RHS. What about more recent members of the *Homo* family? Of all extinct hominids, Neanderthals “are the ones who have left us with the richest record of themselves and their lives” (DeSalle and Tattersall 2008, p. 127). Interestingly, some authors argue that there is virtually no evidence for symbolic behaviour, although Neanderthals had brains as big as ours (DeSalle and Tattersall 2008, p. 129). If the mere amount of nerves should therefore have been responsible for the kind of nervousness that led to traumatic re-enactments in RHS, and if RHS were symbolic and led to mental representation and other symbolic behaviours, then why should there be no evidence for any kind of symbolic behaviour among Neanderthals? What we know about Neanderthals casts doubt, albeit indirectly, on the idea that brain volume—the bag of nerves, so to speak—might have stood at the beginning of Türcke’s developmental trajectory from ETS via RHS to symbolic behaviours and full-blown mental capacities. DeSalle and Tattersall (2008) conclude that “it was evidently not the big brain of Homo sapiens that by itself made the competitive difference; it was symbolic thought, as substantiated by archaeological evidence of symbolic behavior” (p. 193). But, of course, the mere absence of evidence for symbolic behaviour does not prove that Neanderthals were not able to engage in these kinds of behaviours.

On Türcke’s account, everything begins with *Homo* having more nerves and thus being more nervous. At the end of a long historical process, language and culture had evolved. Finally, for the sake of completeness, let me just mention that some support the idea of an opposite trajectory. Hockett and Ascher (1964) argue: “For our ancestors of a million years ago the survival value of bigger brains is obvious if and only if they had *already* achieved the essence of language and culture” (p. 145). Since big brains are biologically very expensive (high percentage of blood supply, skull enlargement, difficulties during parturition, etc.), something must already have existed that led the ongoing enlargement of brain volume to continue. Therefore, the authors argue, language and culture “selected for bigger brains” (Hockett and Ascher 1964, p. 146), and not the other way around.

11.4 *Schrecken* and the Case of Cannibalism

A second critical remark concerns the nature of *Schrecken,* i.e. naturally occurring experiences of shock and horror. When Türcke speaks of the Urstrudel des Grauens (2011, p. 63; “original vortex of fright”, 2013, p. 48), one gets the impression that Türcke conceives of *Schrecken* in almost metaphysical or substantial terms—as an *ens in se,* so to speak. One is reminded of the last words of Joseph Conrad’s protagonist Marlow in *Heart of Darkness:* “The horror! The horror!” But what exactly was *the* horror, what was so frightening that it shook our forebears to the core? Fortunately, Türcke (2011, p. 133, 172, 181) repeatedly provides very concrete examples as well: volcanic eruptions, earthquakes, wild animals—especially snakes with their abrupt movements—thunderstorms and lightning. More generally, what characterizes *Schrecken* is that it occurs all of a sudden and hence completely absorbs an individual's attention (Türcke 2011, p. 181).

Let us take the example of wild animals. How do our closest living relatives, the chimpanzees, react towards dangerous animals? Chimpanzees' behaviour towards predators is highly interesting. Although they are sometimes attacked by predators such as leopards, this does not prevent them from striking back. Hiraiwa-Hasegawa et al. (1986) report about a group of more than 30 chimpanzees in Tanzania observing a leopard den containing a mother and at least one cub, which they eventually dragged out and killed. Apparently, chimpanzees can defend themselves against a predator of up to at least the size of a leopard. Remember that Türcke lists wild animals in his ETS examples. But the present chimpanzees' behaviour does not really support the idea that they are so severely shocked by irregular sudden attacks of predators. If so, would they strike back? Likewise, it might be that our human ancestors reacted in comparable ways, thus rendering the need to cope with ETS in RHS superfluous.

What about present hunter–gatherer societies? Of course, early *Homo* cannot be compared to present hunter-gatherers in almost as many ways as urban people. However, their environment and their resources to cope with nature's *Schrecken are* comparable to a certain extent. Thus, sources for traumatic shock in former times should resemble those in present times in areas where people still live in comparable ways. The San of southern Africa or the Hazda of northern Tanzania live (at least until very recently) as hunter-gatherers in savannahs with dangerous predators, reptiles, snakes and scorpions. They have to face nature's forces without the protecting technology and infrastructure present in urban areas. Nevertheless, they do not sacrifice other human beings in order to cope with traumatic experiences. Also, there is, to my knowledge, no evidence that they suffer significantly more from nightmares in which ETS are traumatically repeated than any other people. But why do these people not suffer as intensely as Türcke claims for our forebears? Türcke would probably give an answer in line with one of his central arguments: They do not suffer as much because—apart from the fact that nature is more under control nowadays—mental representations have an appeasing, calming effect. And like all present *Homo sapiens,* hunter-gatherers have a long history in which traumatic re-enactment has already helped them to dampen traumatic shocks via performative representation (*Vorstellung* in the sense of performance) and finally mental representation (*Vorstellung* in the sense of imagination). If one looks at present hunter-gatherer societies, however, one does not get the impression that nature is as terrifying and shocking to them at all. Looking at both our closest living relatives and at people who still live in a traditional way gives rise to the possibility that the everyday life of our ancestors might not have been as horrible as Türcke suggests.

But let us again focus on the idea that the ability to imagine things and to mentally represent them helped early *Homo* to cope with traumatic shock. What about the opposite effect? What if the level of mentalization our species acquired in the course of phylogeny has made us *more* vulnerable and sensitive? The gradual mentalization and increasing intelligence of *Homo sapiens,* in this perspective, did not only help to "buffer" ETS—it did also increase vulnerability and enhance creativity in ways which gave human beings the power to be even more *schrecklich* as the original natural horrors.

On the one hand, increased mentalization and intelligence might also increase vulnerability via various forms of perspective taking. Perspective taking allows individuals to compare their fate with the fate of others, to view their own lives from a bird's eye perspective and to engage in empathy. All this adds something to the dimension of physical pain. Whereas without perspective taking, an individual suffers only physiologically, perspective taking opens up a more psychological source for suffering which might not simply buffer, but also *add to* the physiological one.

On the other hand, the enhancement of intelligence and mentalization also increased the potential for *Schrecken* by enabling humans to "invent" ETS that are at least as uncontrollable and unforeseeable as the original natural horrors. The organized cruelty in Nazi Germany, the genocide in Rwanda, the experiences of child soldiers, the history of torture methods or the recent reports on torture camps on the Sinai peninsula are just a few random examples to demonstrate that the increased ability to mentally represent what another person experiences can also be used for malicious purposes, i.e. in order to increase *Schrecken* and produce traumatic shocks which are probably at least as arousing as the ones experienced by early hominids (just imagine a skilled and creative torturer).[5]

If we, with regard to the situation of early *Homo,* have to infer the intensity of traumatic shocks and their impact on early hominids on the basis of RHS, which are their traumatic re-enactments and hence expressions of it, then we are also allowed to infer and comparatively evaluate the intensity of traumatic shocks and their impact on the human psyche *in present times* on the basis of the various traumatic re-enactments we can witness today (forms of torture, sadism, violence, etc.). If we do so, we confirm the presence of the traumatic repetition compulsion—yet the idea that the ability of *Homo sapiens* for mental representation somehow softens the impact of ETS in modern humans loses its plausibility. Since today's traumatic re-enactments are not necessarily less cruel than in former times, this would mean that ETS are still experienced as *schrecklich* as ever.

Finally, I will finish my reflections on *Schrecken* with some considerations of a phenomenon that is considered to be a paradigmatically horrific one: cannibalism. Evidence for cannibalism is as old as 800,000 years (Carbonell et al. 2010), much older than any evidence for RHS. Türcke is certainly right in saying that

---

[5] http://sz-magazin.sueddeutsche.de/texte/anzeigen/40203/Im-Reich-des-Todes; http://www.bbc.co.uk/news/world-africa–22575182; http://www.theguardian.com/world/2012/jul/17/eritrea-people-trafficking-arms-sinai.

the killing or sacrificing of conspecifics and cannibalism were not separate from each other in the beginnings of human sacrifice—yet evidence for cannibalism is not an automatic evidence for co-occurring sacrifice. The human remains found in Spain do not show a specific distribution, they appear "mixed with lithic tools and bones of other taxa" and both "nonhuman and human remains show similar evidence of butchering process" (Carbonell et al. 2010, p. 539). In other words, there is no evidence that human remains were treated differently, which we would expect if cannibalism had been integrated into a ritualistic context. Since cannibalism is common among many different animal species—among them our closest relatives, the chimpanzees (Goodall 1977)—it is a practice that certainly preceded sacrifice. Moreover, it is very probable that cannibalism among our human ancestors is much older than 800,000 years. From our present point of view, cannibalism with prior killing is as cruel as the practice of human sacrifice. So what about cannibalism among the hominids and earliest representatives of *Homo*? Should we interpret its repeated occurrence already as a form of traumatic re-enactment, as a compulsion to repeat ETS? With regard to the omnipresence of cannibalism in the animal kingdom, it is more plausible to interpret cannibalism in less mentalistic and more parsimonious terms: Cannibalism was not the rule, but it was part of everyday life and probably not experienced as *schrecklich* as we perceive of it today (which would again support the view that our sensitivity towards ETS might have increased *despite* increasing capacities for mental representation). If this is true, then the appearance of human sacrifice could not have had its catalysing function, since there is no reason to assume that RHS were more horrifying than the already existing cannibalism. Another historical trajectory seems more probable: Killing conspecifics only became a terrible and horrifying experience *after* something has made our ancestors more sensitive. We are left, then, with two possible scenarios:

1. Cannibalism was a traumatic experience for *Homo antecessor,* leading to more ritualistic forms of cannibalism and killing of conspecifics: RHS. As a consequence, *Homo antecessor* found relief and the capacity for mental representation developed.
2. Cannibalism was not a traumatic experience for *Homo antecessor,* and there was no need to find relief in RHS. However, other developments led to a qualitative leap in thinking and sensitivity; thus, the cruelty of the old practice was recognized and needed justification. This led to RHS, in which existing practices of cannibalism were ritualized and endowed with meaning.

While Türcke appears to support the first scenario, I suggest that the second might be a plausible alternative.

My critical remarks on Türcke's conceptualization of early hominids as bags of nerves and *Schrecken* are far from being strong enough to refute his account. Rather, they are suggestions to rethink and fruitfully discuss some of his assumptions. However, they make clear that if we follow Türcke's line of argumentation, we do so on the basis of two starting assumptions that, although plausible to a certain extent, cannot be taken for granted.

11.5 Origins of Human Cognition and Communication from a Comparative Perspective

We do not know when our ancestors sacrificed a conspecific for the first time. Massacres and subsequent cannibalism differ from ritualistic sacrifices. The latter took place at a specific location and the course of events was structured and controlled. The majority had to agree on someone who was going to be sacrificed, they had to collaborate in order to bring the victim under control, and they had to jointly attend to the following events and stick to their ascribed role as either participant, observer, dancer, priest or any other part. Therefore, RHS required some basic abilities, namely, the ability for *joint attention,* the ability for *social reference,* since the group had to refer to "the chosen one", and the ability for *cooperation*. In other words, the group had to be able to work on and bring about a *joint goal*. Suppose that the first ritualistic form of sacrifice took place before *Homo* had acquired a conventional language, then the choice of the victim could only have come about via gestures and gaze following.

According to the developmental and comparative psychologist Michael Tomasello (2008), these are precisely the abilities that distinguish humans from their primate relatives. The greatest difference between young children and great apes lies in children's ability to share attention and to cooperate. Tomasello's account is based on numerous empirical studies in which social-cognitive abilities of human children were compared to those of their closest living relatives like chimpanzees (*Pan troglodytes*), gorillas (*Gorilla gorilla*) and orang-utan (*Pongo pygmaeus*). In a nutshell, the guiding principle of this research program is the following: If all great apes and humans are comparable with respect to certain abilities and differ with respect to specific others, which only humans but none of their primate relatives exhibit, then it is reasonable to assume that these very abilities must have played an important role in the phylogenesis of *Homo sapiens*.

For Tomasello (1999, 2008), the most human-specific activity distinguishing our species from others is our ability and inclination to jointly attend to something in a very specific way. While chimpanzees understand that another chimpanzee attends to something, only humans share the fact *that* they share attention and only humans are willing to collaborate, for example, when it comes to pursuing a goal that was formerly jointly established. This view is derived from a vast corpus of comparative data. In this sense, it has a stronger empirical basis than Türcke's account. Since at least ritualized forms of sacrifice certainly required forming a joint goal as well as joint attention, the question arises whether the species-unique abilities identified by Tomasello did not develop phylogenetically prior to RHS.

Let me briefly summarize some of Tomasello's ideas developed in *Origins of Human Communication* (2008). Tomasello argues that we have to look for the origins of language in gestural communication. Both great apes and humans engage in intentional communication using gestures. Apes use gestures very flexibly in order to request or demand something from others. They use intention movements in order to demand a specific action directly—for example, they approach another individual with their back first in order to be groomed. Moreover, they use attention-getters in order to direct the other's attention. For example, they throw objects or slap on the ground. According to Tomasello, human cooperative communication is more complex because humans do not only understand individual intentionality but also *shared intentionality,* with recursive mindreading being its cognitive basis. Whereas chimpanzees see what others see and understand their gaze as intentional, i.e. as being directed towards something, children go beyond this: They understand that others see themselves seeing. This recursiveness opens up the possibility for sharing. While chimpanzees only use the pointing gesture to request something, young children already use it for more collaborative purposes: They inform others, and they point at something in front of someone else because they want the other to attend as well. Importantly, they do not only point for someone else in order to get their help for an egocentric goal but also because they enjoy the sharing of attention as a means in itself.

Apart from pointing gestures, humans use iconic gestures, i.e. pantomiming, to direct others' imagination. According to Tomasello, more arbitrary communicative and linguistic conventions rely on the same "cooperative infrastructure" (Tomasello 2008, p. 322) as the more natural human gestures (pointing and pantomiming). As neophytes acquire the use of iconic gestures whose iconicity is not fully grasped, they drift, so to speak, to the arbitrary. For example, if someone makes gestures as if he had a spear in his hands in order to communicate to others that some antelopes are at the nearby lake (since they use spears to hunt antelopes), and if children or others who are not familiar with this communicative relationship observe all the other people leaving, they might conclude that the gesture simply stands for the act of leaving and consequently start to use the same gesture in this way. The gesture is thus shifted to the arbitrary and becomes more and more abstract and symbolic (Tomasello 2008, p. 323). Although Tomasello does not overtly speculate about what the first words might have been, Türcke's intuition that they could hardly have been anything else than demonstratives[6] would probably be supported by Tomasello: "The ontogenetic transition from gestures to language demonstrates the common function of (i) pointing and demonstratives (e.g., this and that); and (ii) iconic gestures and content words (e.g., nouns and verbs)" (Tomasello 2008, p. 324).

Tomasello repeatedly stresses that both the skills and motives for shared intentionality "arose initially within the context of mutualistic collaborative activities" (Tomasello 2008, p. 324). Only the cognitive underpinning of shared intentionality, i.e. recursive mindreading, could have led to the formation of a joint goal, which is only truly joint if both collaborators know about having and pursuing the same goal. According to Tomasello, this ability to participate in collaborative activities

---

[6] Türcke seems to suggest that both pointing and exclamatory vocalisations with demonstrative meaning appeared for the first time in the face of *Schrecken* (2013, chapter "Primal words"). Yet, if ETS were really as shocking as Türcke suggests, it is doubtful that our ancestors were still capable of pointing. In a true situation of shock, one might be too paralysed and absorbed by the experience to engage in pointing.

is not present among the great apes (Tomasello 2008). Both pointing and pantomiming "arose as ways of coordinating the collaborative activity more efficiently" (Tomasello 2008, p. 324).

Summing up, Tomasello offers a plausible account of how language might have evolved from nonverbal abilities for the purpose of shared intentionality and joint collaborative action. However, he does not say very much about the phylogenetic origins of shared intentionality. Why were our ancestors motivated to share attention? What made recursive mindreading possible? And what if the species-specific abilities present in young children, but not or less so in other living primates, developed in the course of RHS? One might argue that they were results or by-products of RHS rather than prerequisites. In the same way as unsystematic massacres developed to more ritualistic forms, collaboration, joint goal formation and joint attention became more necessary and might have developed in parallel. How can we know whether the abilities identified by Tomasello evolved in RHS or somewhere else?

11.6 Into the Wild

Our ancestors did certainly not begin to share attention and cooperate out of boredom, fun or sheer curiosity, but because they were forced to do so. Sharing attention and cooperation must have been important for survival. The questions are then:

1. What forced early hominids to jointly attend to some object while—and this is the important part of it—being aware that (a) *both* are attending to it and (b) both are attending to each other's attention (recursive structure of joint attention: I attend to your attention to my attention to your attention, and so forth, ad infinitum; see Tomasello 2008), thus allowing for truly shared attention and joint goal formation?
2. What forced early hominids to collaborate?

As we have seen, shared intentionality is based on the ability for recursive mindreading. Without it, we might have the *same* goal, but not a *joint* goal. Was this ability necessarily a part of RHS? What could have forced our human forebears not only to re-enact ETS, i.e. to sacrifice conspecifics, but also to be aware of others' attention towards one's own attention? Why should recursiveness be important for RHS? According to Türcke, RHS are traumatic re-enactments that provided relief as well as a sense of control. *Schrecken,* normally unforeseeable and uncontrollable, is now enacted in a ritualistic, i.e. more structured way, so that things *can* be foreseen and controlled. If this was the main function of RHS, then the mere observation of and participation in RHS should be sufficient for an individual to find relief. We might also put it as follows: If forming a joint goal is a prerequisite of RHS, then recursive mindreading must have its origins somewhere else. On the other hand, if it did not require forming a joint goal, then recursive mindreading was not necessary and therefore RHS were no acts of shared intentionality—participants had the same goal (observing and enacting something which repeats *Schrecken* in order to find relief), but not a joint goal. And since, according to Tomasello's view, shared intentionality is a key feature of what distinguishes us from other primate relatives, the questions still remain the same: What forced our ancestors to go from individual intentionality to shared intentionality and recursive mindreading? What forced them to collaborate in less individualistic ways than the great apes?

During the East African Miocene, about 7 million years ago, large areas formerly covered by fairly dense and continuous rainforest, dried out due to climatic change, converting tropical forest into open savannah. As a consequence, members of the primate family, including both hominids and apes, had to spend more time on the ground and adapt to open savannahs (DeSall and Tattersall 2008, p. 113). In other words, they were *displaced*. This new situation was threatening:

> Treeless country holds discomforts and dangers. There may not be much food, at least not of a familiar sort. There may be little available water, for the trees tend to cluster where the water is more abundant. And there are fleet four-footed predators, as well as herbivorous quadrupeds big and strong enough to be dangerous at close quarters. (Hockett and Ascher 1964, p. 140)

Adapting to this new environment was not simply uncomfortable and difficult, it was a matter of life and death: "It is a wonder that any of the waifs of the Miocene savannah survived at all" (Hockett and Ascher 1964). In this situation, headquarters were "the safest place to be, if for no other reason than the safety of numbers" (Hockett and Ascher 1964). Later, these places became fortified and turned into locations where things were kept or stored. Many aspects of life, which in the times of hominids' arboreal existence were not restricted to a specific locale, were now confined by a probably circular demarcation: Home was *displaced* and aspects of life *condensed*—they were literally internalized spatially into something we would nowadays call home.

Note that Türcke (2013, p. 120) considers wild animals as sources of traumatic shock as well. The German word for animal is *Tier*. Fierce and dangerous animals are sometimes called *Bestie*. In English, both words are still close to their Latin origins *animal* and *bestia*. Interestingly, the word animal also points to the word *anima,* which is the Latin word for soul. Beasts are usually representatives of *Schrecken*. It is a rather long shot, but what if wild animals were actually both representatives of *Schrecken* and key to the emergence of the soul, or, more precisely, to those mental capacities Türcke considers to develop in RHS?

In the open savannah, our ancestors had to either hunt dangerous animals or protect themselves against them. My suggestion is that the cooperative infrastructure described by Tomasello (2008) might have evolved in this context. We do not know precisely when our human forebears began to hunt big game. About 2 million years ago, hominids were armed primitively, if they were armed at all, and evidence for meat consumption is therefore evidence for scavenging rather than hunting (Isaac 1978; Shipman 1986). But both hunting and protecting probably required a new form of collaboration.

11.7 From Prey to Predator

According to Türcke, *displacement, condensation* and *reversal* are not only the constitutive principles of dreaming, but of early thinking as well. In the scenario outlined above, we have already stumbled upon two different cases of *displacement* and *condensation* than the ones described by Türcke regarding RHS: when early hominids had to survive in the open savannahs. What about the third principle? As Türcke says, wild animals were sources of traumatic shock. If this is true, then hunting animals that are potentially life-threatening can be interpreted as a way of coping or: *reversal*. Although we do not exactly know when *Homo* began to hunt big game, we know for certain that this kind of reversal took place. In Türcke's interpretation of RHS, *Schrecken* was reversed and could lead to mental representation because of the repeated, structured and arousing features of the ritual. Group hunts, too, were regularly done. In the case of big game hunts, these hunts were probably based on a specific strategy and might well have been accompanied by hunting rituals. In other words, they had a certain structure, too. And they were certainly arousing. According to Burkert (1997, p. 27), the ideal man is even nowadays considered to be brave precisely because men must have been brave in former times in order to be a hunter. In an almost poetical style, Burkert reflects on the greatness of the point in time when our ancestors succeeded in taking over the role of the wolf and transformed from prey to predator.[7] Hockett and Ascher (1964) describe the reversion as follows: "Thus the hunted became the hunters, and weapons of defense became weapons of offense" (1964, p. 141). According to Burkert (1997, p. 27), hunters must be able to voluntarily delay their gratification and to be true to their word. The latter idea is central to the question we are still trying to answer: What forced early *Homo* to go from individual intentionality to shared intentionality (recursive mindreading being its cognitive basis) and to cooperate in order to achieve a joint goal? When our human forebears had to adapt to the savannah, they were more exposed to predators. No matter whether they hunted predators or protected themselves against them—for the group to survive, its individual members had to rely on each other in a qualitatively new way. Collaboration might have been the only way to survive.

One possible objection, of course, is that some believe hunting activities of wild chimpanzees to be cooperative as well (Boesch 2005). If this were true, then cooperation could not have been the basis for our human-specific abilities. Whether or not hunting activities of chimpanzees are cooperative in the sense that chimpanzees have a *shared* goal (and not simply the *same* goal) and know about having this same goal, is hotly debated (Boesch 2005; Tomasello 2008). At this point, I want to suggest a potential solution that might be able to explain the different developmental pathways of humans and chimpanzees *even if* chimpanzees' hunting activities were cooperative. The idea is that there is a clear difference between chimpanzees'

---

[7] "Welch ein Erlebnis mußte es sein, als es dem Verwandten des Schimpansen gelang, die Macht des Todfeindes, des Leoparden, an sich zu reißen, in die Maske des Wolfes zu schlüpfen, vom gejagten Wild zum Jäger zu werden!" (Burkert 1997, p. 26).

hunting activities and the hunting of our human forebears, at least to those that had already developed weapons efficient enough for big game hunt.

Chimpanzees hunt smaller monkeys, i.e. prey that is not dangerous to an individual chimpanzee. In contrast, big game hunted by our human forebears was much more dangerous and could easily kill an individual human. Therefore, cooperation and fulfilling one's role in the hunt were much more important than among chimpanzees: They were a matter of life and death. Each hunter had to rely on the other hunter(s). Yet, what is reliance if not a recursive form of expected on-going cooperation? I am confident that you will think of me and help me when I am in danger during the hunt, because I know that you have the same confidence in me when you end up in a dangerous situation. If collaborative hunting (or collaborative defense) was indeed important for the survival of our human forebears in the open savannahs of Africa, then we are justified in assuming that they were indeed forced to cooperate in a much more reliable way than, for example, chimpanzees. Big game hunt usually provides more meat than the hunters can consume. Therefore, meat was shared—again with the expectation that others will share their meat in the future as well. All these abilities belong together. Sharing food might have helped to transform pointing gestures that were used to request something into pointing gestures that were used to share something, for example knowledge, attention or information. Such sharing was crucial for joint hunting activities to be successful.

11.8 An Alternative Story

The aim of this chapter was to critically discuss Türcke's account outlined in *Philosophy of Dreams*. His ideas are highly interesting and original. The aim of this contribution was not to discard his account—on the contrary, it deserves much further attention. At the same time, it is an account that is in need of further clarification, discussion and critical comments. In order to start such a discussion, I have developed some arguments that lead to a different story. In this story, RHS might still have served to cope with ETS, and they might well have been traumatic reenactments. But they are not at the core of what makes us human. Türcke suggests that sensitivity or "nervousness" led to RHS. Bearing in mind the long history of a comparably cruel behaviour like cannibalism, however, it was suggested that—due to specific socio-cognitive abilities—*Homo* became more and more aware of the cruelties that already had been part of everyday life for many thousands of years. I suggested that joint hunting activities or joint defense activities might have provided a plausible context for shared intentionality, recursive mindreading and joint collaborative goal formation to evolve. Yet, as soon as recursive mindreading is practiced, it gives rise to a completely new perception of other individuals. Recursive mindreading allows for forms of inter-individual perspective taking that were formerly impossible. Cruelties that had been practiced up till then are now not only perceived from the outside but from the perspective of the victim as well. Therefore, instead of suggesting that our species-specific abilities evolved in the course of

RHS, I suggest that these abilities had a different origin. They led to the recognition of existing behaviours that were eventually perceived as more arousing than previously. RHS, then, were already attempts to "make it better": whereas cannibalism and prior killing had no meaning, meaning was now urgently needed and projected onto RHS. But what distinguishes us from our primate relatives was already present in our human ancestors before they began to sacrifice conspecifics.

References

Boesch, C. (2005). Joint cooperative hunting among wild chimpanzees: Taking natural observations seriously. *Behavioral and Brain Sciences, 28,* 692–693.

Bradshaw, G. A. (2004). Not by bread alone: Symbolic loss, trauma, and recovery in elephant communities. *Society and Animals, 12,* 143–158.

Burkert, W. (1997). *Homo necans. Interpretationen altgriechischer Opferriten und Mythen*. Berlin: de Gruyter.

Carbonell, E., Cáceres, I., Lozano, M., Saladié, P., Rosell, J., Lorenzo, C., Vallverdú, J., Huguet, R., Canals, A., & Bermúdez de Castro, J. M. (2010). Cultural cannibalism as a paleoeconomic system in the European lower pleistocene. *Current Anthropology, 51,* 539–549.

DeSalle, R., & Tattersall, I. (2008). *Human origins. What bones and genomes tell us about ourselves*. College Station: Texas A & M University Press.

Freud, S. (1961 [1900]). *Die Traumdeutung*. Frankfurt a. M.: Fischer.

Goodall, J. (1977). Infant killing and cannibalism in free-living chimpanzees. *Folia Primatologica, 28,* 259–282.

Hiraiwa-Hasegawa, M., Byrne, R. W., Takasaki, H., & Byrne, J. M. E. (1986). Aggression toward large carnivores by wild chimpanzees of Mahale Mountains National Park, Tanzania. *Folia Primatologica, 47,* 8–13.

Hockett, C. F., & Ascher, R. (1964). The human revolution. *Current Anthropology, 5,* 135–168.

Isaac, G. (1978). The sharing hypothesis. *Scientific American, 4,* 90–106.

Nagasawa, M., Mogi, K., & Kikusui, T. (2012). Continued distress among abandoned dogs in Fukushima. *Scientific Reports*. doi:10.1038/srep00724.

Roth, G., & Dicke, U. (2005). Evolution of the brain and intelligence. *TRENDS in Cognitive Science, 9,* 250–257.

Shipman, P. (1986). Scavenging or hunting in early hominids: Theoretical framework and tests. *American Anthropologist, 88,* 27–43.

Tomasello, M. (1999). *The cultural origins of human cognition*. Cambridge: Harvard University Press.

Tomasello, M. (2008). *Origins of human communication*. Cambridge: MIT Press.

Türcke, C. (2011). *Philosophie des Traums*. München: C. H. Beck.

Türcke, C. (2013). *Philosophy of dreams*. New Haven: Yale University Press.

Index

T. Breyer (ed.), *Epistemological Dimensions of Evolutionary Psychology,*
DOI 10.1007/978-1-4939-1387-9