ProtoSociology
Publications on Contemporary Philosophy

ration ist seit E. Durkheim, den wir als einen Begründer des Fachs Soziologie einstufen, ein relevanter Schwerpunkt des Fachs Soziologie. Um dabei die Weichen richtig zu stellen,

Free Will and Evolution

– Ingvar Johansson –

ziologie nimmt einen besonderen Anschnitt vor, der zu skizzieren ist, da soziale Integration als eine Differenzierung von Mitgliedschaftsordnungen eine bestimmte allgemeine Theorie voraussetzt. Unter „Ordnungen" sind dabei die Regelung der Mitgliedschaftsbedingung und damit die Teilnahme an Kommunikationssystemen in der Ausübung von bestimmten Rollen und Statuspositionen zu verstehen. Die Mitgliedschaftstheorie fasst die System-Umwelt Relation nicht als sinnkonstituiert (Niklas Luhmann), sondern als die selbstreferenzielle Entscheidung über Mitgliedschaftsbedingungen und ihre Selektion, die keine Resonanz in der nicht-sozialen Umwelt hat. Der Verweisungszusammenhang von Sinn, wenn wir das einmal unterstellen, ist in diese Differenzstruktur einzuordnen. Gehen wir von der mitgliedschaftstheoretischen Selbstkonstitution sozialer Systeme aus, so sind soziale Systeme souverän. Damit geht einher, dass die soziologische Theorie die folgenden Annahmen aufgeben sollte:

ProtoSociology

Publicatons on Contemporay Philosophy

Ingvar Johansson
Free Will and Evolution

Translated from the Swedish by Alice E. Olsson with the author

© 2023 ProtoSociology – Gerhard Preyer
Frankfurt am Main
http://www.protosociology.de
peter-t@protosociology.de
Erste Auflage / first published 2023
Herstellung und Verlag: BoD - Books on Demand, Norderstedt
ISBN: 9783743124707
Bibliografische Information Der Deutschen Bibliothek
Die Deutsche Bibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über http://dnb.ddb.de abrufbar.

Alle Rechte vorbehalten.
Das Werk einschließlich aller seiner Teile ist urheberrechtlich geschützt. Jede Verwertung außerhalb der engen Grenzen des Urheberrechtsgesetzes ist ohne Zustimmung der Zeitschrift und seines Herausgebers unzulässig und strafbar. Das gilt insbesondere für Vervielfältigungen, Übersetzungen, Mikroverfilmungen und die Einspeisung und Verarbeitung in elektronischen Systemen.

Bibliographic information published by Die Deutsche Bibliothek
Die Deutsche Bibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available in the Internet at http://dnb.ddb.de.

All rights reserved.
No part of this publication may be reproduced, stored in a retrievalsystem, or transmitted, in any form or by any means, without the prior permission of ProtoSociology.

Contents

© ProtoSociology

Preface

With this little book, I wish to demonstrate that not all philosophers have given up on the belief that humans sometimes have a little bit of free will. All states, events, and processes in the world can, in my opinion, not be explained solely through causal factors, regardless of whether these are seen as completely determining or only determining with a certain degree of probability. It is not the case that everything happens due to necessity or chance. Sometimes, humans have a little bit of freedom of action and—more often yet—a will that is, within certain limits, free. For a long time, this opinion has not been held in high regard among philosophers and scientists, especially not among natural scientists and social scientists of the structuralist proclivity.

After tentatively starting to write this book in the early autumn of 2020, my motivation to finish the work was reinvigorated in an unexpected way later in October, when the three recipients of the Nobel Prize in physics were announced. One of them, Roger Penrose (b. 1931), has long held the door open to a belief in free will. Though he was not given the award for these opinions, of course. He received it for his theories on how black holes are formed in the universe.

To me, it feels as though, toward the end of the 20th century in the academic West, belief in free will was sucked into a cultural black hole. Once matter and radiation has been absorbed by a physical black hole, they can—according to prevailing theories—never escape again (except for any possible Hawking radiation). Yet I allow myself to remain optimistic that a belief in free will may yet escape the black hole by which it has been swallowed.

In *Shadows of the Mind* (1994), Penrose writes: 'This book will not supply an answer to these deep issues [about free will], but I believe that it may open the door to them by a crack—albeit only by a crack' (p. 36). His words brought to my mind a famous line from the song 'Anthem' by Leonard Cohen (1934–2016): 'There is a crack, a crack in everything / That's how the light gets in.'

The most central chapter of this book is Chapter 3, 'Why it is absurd to completely deny the existence of free will'. It can be read independently of the other chapters. The same goes for Chapter 11, 'Free will and morality'.

Two of the chapters are significantly more philosophically finicky than the others, and thus likely to be more difficult for most readers to immediately digest. These are Chapter 7, 'Free will before the Scientific Revolution', and Chapter 8, 'Free will from an evolutionary perspective'. Yet I hope and believe that a quick reading of these chapters may give any reader a sense of

how I want to tackle certain philosophers' arguments as to why free will is an illusion.

For valuable comments on a previous draft of the book or parts of it—both supportive and requiring me to rethink—I wish to thank: Jan Almäng, Thomas Caesar, Rögnvaldur Ingthorsson, Nils-Aage Larsson, Ida Linde, Niels Lynøe, Carl Gustaf Olofsson, Svein Solberg, Christer Svennerlind, Per-Olof Westlund, and Olof Öhlén.

Ingvar Johansson
Lund, Sweden, December 2021

© ProtoSociology

Chapter 1

Freedom of action in human beings

Sometimes when we want to do something, we can just go ahead and do it. But other times, we are prevented by people or other external circumstances. Sometimes, we are prevented by such things as disease. In both kinds of cases—external and internal obstacles—most people would probably consider their restricted freedom of action as consistent with a will, behind the regrettably impossible actions, that nevertheless to some degree is free. If we are unsure of whether we can actually do what we want to do, we sometimes give it a try to see if we can. If we are 100 per cent sure of its impossibility, our will is reduced to a free wish, and we might say to our friends: 'Oh, how I wish I could do this, that, and the other.'

We often look at others the same way, that is, we believe that behind their—perhaps for the moment limited—freedom of action there is a free will and a free wish. The debate usually focuses on free will, but the arguments against it are such that they also lead to a denial of the existence of free wishes.

Thus my book is also a defence of our freedom to create wishes.

If you believe that—in the way I have outlined—human beings have some, albeit limited, degree of freedom of action and free will, but do not believe in a god or some other supernatural phenomenon from which our free will derives, this book is a thorough defence of your position. Herein, I explain free will from an entirely naturalistic, evolutionary, and secular perspective. At some point during the course of evolution, I argue, free will has arisen on our planet. It may be highly limited in its content, yet nevertheless it is not in all respects predetermined by the laws of nature, social structures, and the previous moment.

If, on the other hand, you believe that you and everyone else lack any free will whatsoever, and that this opinion is the only one that is consistent with modern science, I hope this book may disturb your circles. At closer inspection, the view offered by modern science is neither as unequivocal nor as universal as deniers of free will tend to think. Almost everyone accepts evolutionary theory; yet few appear to have properly thought it through.

The will is a mental phenomenon, and as such it differs in its very nature from purely material phenomena. Yet despite this, I will not be defending a dualism like the kind known as Cartesian dualism after its originator René Descartes

(1596–1650; Latin, Cartesius)—that is, the belief that matter and consciousness are of such different nature that they can theoretically exist independently of each other. My stance could be referred to as both non-Cartesian dualism and non-reductive materialism. I argue that, though matter may exist without mental phenomena, the latter cannot exist without a material substrate.

In the next chapter, I will explain why it has become so easy—indeed, all but a given—for many naturalists to completely deny the existence of free will. The absurdity such a total denial actually leads to will be explored in the third chapter. In chapter by chapter, I will then present and defend various positions I hold. These, taken together, lead me to conclude in Chapter 8 that it is fully reasonable to believe that evolution has given rise to a partially free will. In Chapter 9, I comment briefly on the views held by evolutionary biologists—in particular, Richard Dawkins. Chapter 10 is dedicated to exploring freedom of action as it shows itself in our perception, and Chapter 11 to free will and morality. All that then remains for the final chapter are some concluding reflections.

But before I begin, a few words on the not entirely unambiguous term 'naturalist' that I have already used.

I am *not* a naturalist in the sense of believing that knowledge can only be gained through the kinds of methods used in the prototypical natural-scientific disciplines of physics and chemistry. Believing in the fundamentals of the scientific theories about the cosmos and its origins, as well as the biological theory of evolution—as I do—is not the same as believing that all knowledge about our world must be obtained using natural-scientific methods. In particular, I do not believe that the natural sciences offer methods appropriate for all types of logical-semantic reasoning (see Ch. 3) or for providing an adequate description of perceptions as conscious mental phenomena (see Ch. 10). In both cases, the natural sciences have, through abstraction, done away with significant elements of human subjectivity—what is sometimes referred to as the 'first-person perspective'. This often yields successful results, but when the impersonal third-person perspective of the natural sciences is built into our general view of the world, these abstractions have devastating consequences.

I *am* a naturalist in the sense that I believe everything that exists is part of the spacetime unity we call the universe. This means that if a phenomenon in our universe can be explained with the help of one or several other phenomena, the latter must also exist in our universe.

I regard the question about the existence of free will as a purely ontological problem: does it exist or not? All too often, this question is immediately linked to whether people can ever be said to be morally responsible and how harshly they should be punished if they fail to live up to certain moral stan-

 © ProtoSociology

dards. My answer to the ontological question, as I have already indicated, is unequivocally in the affirmative; the reasoning is as follows. My answer to the moral-philosophical question is more loosely sketched. It can be found in the penultimate chapter (Ch. 11).

Chapter 2

Why it seems absurd to believe in free will

According to today's science, both humanity and the planet we live on have a history of origin—neither has always existed. It is not that science believes itself to have finally uncovered the evolutionary details. But it is certain that in both cases there is a long, drawn-out history of origin of some kind. I will initially present some parts of what the prevailing physical cosmology and biological theory of evolution have to say on the subject. I want to stress that these are the views of today. The details will almost certainly change in the future, as they have over the past one hundred years. But this does not affect their undermining of a belief in free will, a fact I will return to in a special section toward the end of the chapter.

Many readers will most likely be familiar with many—maybe even all—of the popular-scientific views I outline herein. For this reason, I shall present the chapter's conclusion already at its outset. Any reader who wishes to may then proceed directly to Chapter 3, in which I explain why—despite the evolutionary origin story thus presented—it is absurd to completely deny the existence of free will.

Before I present my conclusion, however, I must make a distinction pertaining to the philosophy of science. Scientific theories that describe temporal sequences can be divided into two main categories: deterministic and indeterministic. Deterministic theories allow scientists to feed initial conditions into a system that the theory is presumed to describe, such that it is *in principle* possible to predict exactly what the system will look like at later points in time. What happens is assumed—given the initial conditions—to do so by necessity. If the theory is not entirely correct, or the initial conditions do not correspond to reality, the prediction will be more or less incorrect. Indeterministic theories, on the other hand, contain a component which assigns a certain probability to each of the various results that are theoretically possible. In other words, indeterministic theories can be said to formulate probability laws. With such theories, no definitive predictions can ever be made, even in principle. What happens is assumed—given the initial conditions—to happen with a certain degree of probability. But even deterministic theories can be involved in predictions and probability distributions. In a coin toss, the results of each individual attempt are determined by the laws of nature as well as the initial conditions. But if the initial conditions vary by chance, the results of many attempts will still follow a probability distribution. The conclusion drawn from this chapter can be stated as follows:

> All theories and hypotheses on the basis of which evolution is described today are either deterministic, indeterministic, or a combination of the two.

If all individual phenomena associated with indeterministic theories (probability laws) are referred to as random, the same conclusion can also be formulated in the following way: *Everything that happens or has happened in evolution is due to necessity and/or chance.*

If this is true, free will does not exist. In the case of free will—as I choose to characterise it—*all* of the content of said will may per definition not arise out of necessity, chance, or a combination of the two. Some small part must be freely created. Here, I want to emphasise that it is indeed a matter of 'some *small* part', as no one's will can be entirely above all the needs and desires that arise throughout life.

Today, a much abbreviated popular-scientific story about the origins of the Earth and mankind goes as follows.

Originally, the universe was very, very small. But around 14 billion years ago, it began to expand. Whether the universe had already existed for a while or completely lacks a point of origin in time is a question left unanswered. In this sense, today's story differs from that of the Big Bang—with an absolute starting point—that I and those of my generation grew up with, and which is still likely to be prevalent among the general public.

At the start of this expansion, the universe was extremely hot and dense (mass per unit of volume). In this state, there was no difference between various types of particles. For this reason, it is not possible here to distinguish between different types of particles and any interplay between them. When, for some reason, the universe began to expand, it led to a drop in temperature and particles began to form. Initially, this happened at an enormous rate. In just a few trillionths of a second, various kinds of subatomic particles (quarks, leptons, and bosons) were formed, as well as the four types of fundamental forces or interactions postulated by modern physics: gravity, electromagnetism, strong interaction, and weak interaction.

Subatomic particles are not particles in the sense of classical mechanics or everyday life—that is, unities clearly delimited in space. Nor are electromagnetism, strong interaction, and weak interaction forces in the sense of classical mechanics—that is, a *relationship* between particles. Instead, they are considered mediated by one of the particle types: the boson particle. Nor is gravity a force in the sense of Newton's law of gravity; rather, it is considered an effect of the curvature of spacetime.

Subsequently, these subatomic particles came to form atoms—primarily hydrogen and helium atoms. Later still, the atoms came together to form molecules.

The story outlined above is not a distillation of one single overarching physical or physico-chemical theory. Today's standard model for the origin of the universe, as presented thus far, builds on two separate theory fields: the general theory of relativity and the theory (or theories) of quantum gravity. The former is used for calculations when the given distances or intervals in spacetime are not too small (but without any upper limit), while the latter is used in cases of very minute distances or intervals of time, as was the case at the beginning of the story.

The general theory of relativity is a *deterministic* theory. The theory of quantum gravity, however, falls back on an *indeterministic* theory structure—that of quantum mechanics. No physicist has yet managed to create a synthesis in light of which both above-mentioned theories can be considered approximately true. There are two reasons for why these theories are impossible to simply merge. Firstly, the general theory of relativity is, as already mentioned, deterministic, while quantum mechanics is indeterministic. Secondly, the theory of quantum gravity does not equate time and space in the way that is typical of the general theory of relativity.

In one of its formulations, the general theory of relativity (not to be confused with the *special* theory of relativity—though Einstein is the father of both) consists of a system of equations containing ten partial differential equations with a number of physical variables. As such, the system of equations allows for many solutions, and to arrive at a specific solution for a cosmological model of some kind, certain restrictions must first be imposed. These restrictions are based on what physicists think we know empirically about the universe. Einstein believed the universe to be static, and thus made an assumption that later became obsolete when cosmologists discovered new data that led them to regard the universe as dynamic and expanding. With new values for Einstein's so-called cosmological constant, the theory gave rise to models for the universe's expansion, too. Today, cosmologists are only debating the *rate* of expansion. The general theory of relativity is thus alive and well, even though Einstein's original restriction is completely outdated.

The theory can be used to look both forward and backward in time. It was through calculations into the past, using models that—when looking into the future—resulted in an expanding universe, that the first theory of the Big Bang was born. According to this model, the universe appeared to have originated from one single point in spacetime. The problem was that some of the proper-

 © ProtoSociology

ties attributed to this point seemed to be physically impossible—for example, that its matter density should be infinitely high.

Problems such as this disappear if one decides that the general theory of relativity should not be applied to very minute distances and intervals of time. And that is what cosmologists have decided. Instead of a point that is metaphorically described as exploding (in the Big Bang), they now posit a very small spacetime interval that begins to expand.

But back to our story. The atoms and molecules mentioned came to form giant clouds of dust or gas, which the aforementioned forces eventually bound together into galaxies, solar systems, and planets. In the process, black holes also arose (places where gravity is so strong that neither light nor matter that enters can ever escape again, except for any possible Hawking radiation), as did dark energy (a type of energy that is assumed to permeate the universe and could explain its growing rate of expansion), and dark matter (a type of matter that neither emits nor reflects electromagnetic radiation and thus cannot be observed).

Black holes and dark energy derive their theoretical explanation from the general theory of relativity, while the explanations of the origins of galaxies, solar systems, planets, and dark matter are based on theories about gas dynamics. The latter theories cannot currently be derived from either the general theory of relativity or the theory of quantum gravity.

The story of evolution as told thus far is, in other words, based on *three* types of theories: the general theory of relativity (given certain initial conditions), the theory (or theories) of quantum gravity, and theories of gas dynamics. These are either deterministic or indeterministic.

Having noted this, let us continue the story, now taking the existence our own planet as our point of departure. It is assumed to have been formed roughly 4 billion years ago.

At that time, there was—to put it mildly—a very large amount of different subatomic particles, atoms, and molecules on Earth, which, as they interacted, could give rise to new kinds of molecules. Of course, hypotheses about how this happened cannot be tested directly. However, it is possible to prove that a certain hypothesis is not entirely baseless by successfully recreating the process in a laboratory environment.

One very important event in evolution was the appearance of the first organisms. Another was the appearance of the first *sexually reproducing* organisms. Here follows a few words about both:

An organism has a metabolism and the ability to reproduce. It is, for its existence, dependent on interacting with its environment; it needs to absorb

nutrients of various kinds and release the resulting waste products. Its constituent parts are built up of organic compounds, e.g. the DNA molecule. But as of yet no one has successfully produced an organism using solely organic compounds in a lab.

All organisms consist of one, several, or a tremendous number of cells. In every individual multi-cellular organism, all types of cells that contain DNA—which most of them do—have exactly the same kind of DNA. These spatially separated DNA molecules are, so to speak, identical copies of each other. This fact is what makes it possible for the police to tie bodily substances at a crime scene to a particular individual.

Fossil data shows fairly unequivocally that all organisms have arisen in a certain order out of one or a few single-celled ur-organisms. However, we know rather little about the mechanisms by means of which they arose.

It is likely, of course, that the first organism/cell arose through coincidences involving different kinds of organic compounds. It has been speculated that the event was preceded by meteorites impacting Earth or by electrical discharges. If Earth was formed roughly 4 billion years ago, and the first organisms arose roughly 3 billion years ago—as is a common estimate—there was plenty of time for chance to strike. At the same time, there was an incredible number of possible combinations, of course, and so perhaps the probability that an organism would form was very small, after all. Yet even things with an extremely low probability *can* happen. An extremely low probability is not something to base a prediction on; but in this case we are looking in the rear-view mirror.

When it comes to the mechanisms for how today's rich diversity of organisms—everything from single-celled bacteria to multi-cellular organisms such as fungi, plants, and animals—could arise from those original organisms, we know a lot. The model that explains the origins of new kinds of organisms, and the extinction of those already existing, has sometimes been summarised as 'mutations plus natural selection'. Changes (mutations) in the genetic material (the genome) is a precondition for new species arising; and natural selection explains why some mutated species survive and others die out.

Here, the term *natural selection* (first coined in contrast to the 'artificial' selection in the breeding of domesticated animals) is referred to in a very broad sense. It involves not only individuals of different kinds competing for successful reproduction, but also such factors as changes in temperature, the composition of the atmosphere, and light conditions that could kill certain species (think, for example, of photosynthesis in plants). Now that humanity is causing such changes on a large scale, the suitability of the term *natural* selection is, of course, debatable. Nevertheless, I will continue to use it.

 © ProtoSociology

The first organisms were single-celled and reproduced in the asexual manner we can still observe in bacteria today. They simply split in half. First, the bacterium's chromosome battery—that is, the part of the bacterium that carries its genetic material—is duplicated. (One such battery contains several chromosomes; one chromosome can contain several genes; and one gene is a sequence of parts—so called bases—in the DNA molecule.) After the chromosome duplication, the cell is split in two. Now, each new cell has its own chromosome battery and its own cell wall. Because the chromosome batteries—with their genetic material—are exactly alike, the two bacteria will normally also be exactly alike.

This process of cell division is not described using mathematical equations of the kind found in mathematical physics or chemistry. Instead, the process is only given a non-mathematical description where different phases and sub-phases normally follow each other in a certain order. Whether a phase necessarily leads to the next or only does so with a high degree of probability is of secondary interest, which also means that, in this case, the distinction I initially made between deterministic and indeterministic theories is, so to speak, bypassed.

From single-celled and asexually reproducing organisms, multicellular and sexually reproducing organisms such as plants and animals eventually arose. In the former, cells typically lack a nucleus (prokaryotes), while in the latter cells have a nucleus where the chromosome battery is found (eukaryotes). There are a number of different hypotheses about the mechanisms through which simple eukaryotes developed out of prokaryotes, and how more complex eukaryotic organisms that reproduce sexually eventually came to be. Based on the fossil record we have access to today, it appears that the first sexually reproducing organisms arose ca. 1 billion years ago.

Sexual reproduction requires the joining together of genetic material from two different types of sex cells, albeit not necessarily from two different organisms. There are hermaphrodites with the ability to self-fertilise, but for the sake of simplicity I shall leave them out of the story. In the case of sexual reproduction, then, sex cells from two different individuals (in humans: eggs and sperm)—each with their own set of chromosomes (in humans there are 23 chromosomes in each sex cell)—are somehow brought together, after which the new cell (thus with 46 chromosomes) begins to divide. The end result of continued, repeated cell division, during which cells differentiate to take on different functions, is a fully formed multi-cellular organism. The final organism's own genetic material is not defined by either one of the parents', but both.

Mutations in the genetic material can arise in both types of sex cells. They

may occur randomly during the cell division that creates the sex cells (also called the gametes) or be caused by radiation or certain chemical substances. In any given environment, these may be advantageous, neutral, or disadvantageous to the individual. Those that are advantageous or neutral are passed down to the next generation. With the emergence of sexually reproducing organisms, the probability of new species arising increased dramatically. During the so-called Cambrian explosion, ca. 540 million years ago, an abundance of new animal species arose.

If we stick to sexually reproducing organisms, we can, for the sake of this overview, define a species as a group of organisms in which a female and a male individual can in theory produce fertile offspring. Individuals within a species thus defined differ in terms of both their genes (genotype) and appearance (phenotype). To the untrained eye, there may, within any one species, exist sub-groups of individuals that appear similar to each other in various ways, but that differ from other sub-groups in such characteristics as skin colour, hair colour, height, physiognomy, etc.

When it comes to the human species—in current taxonomy known as *Homo sapiens* (and as such classified under primates)—there was during the 18th to 20th centuries a practice of lumping individuals together into sub-groups known as *races*. This had nothing to do with their ability to reproduce or a total lack of sexual attraction between races. Rather, it was based solely on external, observable traits. Sometimes, these race classifications were buttressed by the belief that the mixing of races would always or easily lead to a reduced general human ability in the children—an idea which lacked any empirical support whatsoever. Yet many believed that if the races mixed, it would lead to degeneration.

Through a great number of small mutations combined with natural selection, a new species can arise out of an existing one. Depending on the circumstances, organisms with certain mutations have greater opportunities to reproduce than others, and thus become, *in the given situation*, 'selected' by nature. It is in such situations that the concept of natural selection as a competition between individuals for successful reproduction is adequate. And if whole species are competing with each other for food and other necessities for living, the species which *in the given* type of competition is weaker will be eliminated. The theory of evolution is not an example of a goal-oriented development.

Today, we know a great deal about the hereditary mechanism behind the reproduction of plants and animals. And not just that. On the basis of this knowledge, we have discovered how to intervene directly in this mechanism. Genetically modified crops and possibilities to change genetic sequences in

 © ProtoSociology

embryonic cells in order to, for example, eliminate hereditary diseases are spectacular advances. But the way geneticists describe this mechanism is, within the context of the present enquiry, worth spending a few words on.

The description contains many terms borrowed directly from chemistry—terms that name different kinds of molecules and chemical substances. Proteins, amino acids, and nitrogenous bases play a prominent role. But just as central—in some ways even more so—are certain terms that superficially appear to belong to the fields of linguistics and information technology. There is talk of genetic *information* that can be *transcribed*; a certain gene *codes for* a certain trait; and the genome contains *instructions* for what the finished organism will look like. This should not be taken literally to imply that there is, as in regular communication between people, a sender and a receiver of said information. It is always, at its core, a matter of descriptions of purely chemical processes. It is a matter of correspondences between the makeup and structure of different molecules.

Theories about gene-to-trait correspondences do not fit neatly within the deterministic-indeterministic binary introduced at the start of this chapter. Provided that certain conditions are met, the descriptions of these correspondences display more of a deterministic than an indeterministic character. Yet, as far as I understand, not much would change within theoretical or applied genetics if, instead of 'gene G codes for trait E', one were to say 'gene G is highly likely to code for trait E'.

For the purposes of this book, the biological theory of evolution can be summarised as follows: the origins of new species (through mutations and natural selection) can be explained by necessity and/or chance, and the disappearance of certain species (through natural selection) is explained by necessity and/or chance. I call this summary the necessity-and/or-chance paradigm.

* * *

I have described the currently prevailing evolutionary-biological story about the organisms on our planet. However, there are at least three well-known proposals for qualifications of the story. These are known as *punctuated equilibria*, *epigenetics*, and *evolutionary-developmental biology*. I will now say a few words about each of them.

According to the theory of punctuated equilibria, fossil data shows that the evolution of plants and animals did not take place over a long period of time through numerous small mutations. Rather, the theory claims, mutations bring about new species during a short space of time between two periods

of equilibrium. However, the general framework of mutations and situation-specific natural selection is not contested.

The field of research known as epigenetics challenges the view that all genetic changes arise through mutations in the genome. Epigenetics claims that gene expressions can be modified by certain factors external to the genome itself. Within an individual, all daughter cells of a modified cell will inherit the modification. And if the change occurs in the sex cells (gametes) or in a fertilised sex cell (zygote), the modification becomes hereditary. This opens up for speculation about whether individual behaviours and lifestyles can cause such genome-external factors inside a cell. If the answer is yes, this would mark the revival of—albeit in a significantly weakened form—the pre-Darwinian idea that acquired traits can be inherited. Yet epigenetics does not challenge the more abstract view that evolution is only a matter of necessity and/or chance.

The field of research known as evolutionary-developmental biology, or evo-devo, is a multifaceted phenomenon. Evo-devo recognises that, in the dominant evolutionary narrative, the theory of natural selection focuses on macro-biological/visible (phenotypic) traits in organisms and the groups they may be an integral part of, while the theory explaining the causes of genetic variation focusses on molecular-biological (genotypic) traits. This means that theories of how embryos develop into fully-grown organisms become irrelevant in this context; it is assumed that genes correspond in a fairly direct way to their phenotypic expression. Especially now, as the sequencing of the human genome has shown that the genetic material in humans is surprisingly similar to that of simple animals, an explanation is needed for why we are so phenotypically different. And the development from embryo to fully-grown organism should reasonably enter into the formation of such a theory. But evo-devo theories do not conflict with the view that evolution is merely a matter of necessity and/or chance, either.

* * *

I have divided the story of evolution into two parts: one cosmological and one biological. In concluding the cosmological part, I mentioned that it is based on three types of theories, which are either deterministic or indeterministic. One could also say that the cosmological story is based on the necessity-and/or-chance paradigm. When it comes to the biological story of evolution, it is not quite so easy to delineate a specific number of theory types, but there is also no reason to do so. Here, biological theories about the conditions under which different species can survive sit alongside theories borrowed from both

 © ProtoSociology

micro- and molecular biology. For our purposes, it is enough to note that the biological story of evolution is based on several types of theories, which cannot be integrated into one unifying theory. Yet they are all based on the necessity-and/or-chance paradigm.

I believe that the answer to the question this chapter asks—why is it absurd to believe in free will?—is primarily that modern cosmology and biological evolutionary theory has no room for it. In the story of evolution, there are only the natural-scientific concepts of causal necessity and chance. There is no room for free will within the necessity-and/or-chance paradigm of the evolutionary narrative.

Two subsidiary arguments can be mentioned—though they are no longer particularly common. One is that free will is not observable and science must be based on observations. Thus, the notion of free will must, from a scientific point of view, be rejected. This view will be discussed in Chapter 10.

Another argument, very popular in the 1990s, referred to the work of physiologist Benjamin Libet (1916–2007). In a laboratory environment, Libet proved experimentally that certain characteristic brainwaves (a so called 'readiness potential' associated with action) always *preceded* conscious decisions about making certain finger movements, and the movements were thus only apparently voluntary. The argument insisted that this could be generalised to all human behaviour. Libet himself, unlike many others, did not actually draw this conclusion. He believed that the will is free to say no to a coming decision, but not free to initiate an entirely new one. In other words, the will has a veto right, but not the right of initiative. In the 2010s, Libet's experiments were criticised for not taking into account normal electrophysiological background activity. My argument, however, is independent of Libet's experiments and the criticism they inspired.

As mentioned, the story of evolution builds on a conglomerate of different theories, but this does not mean that it suffers from a complete lack of coherence. Its coherence arises out of the fact that all these theories describe events and processes overlapping in spacetime. Everything that they describe occurs in the spacetime continuum of our universe.

Chapter 3

Why it is absurd to completely deny the existence of free will

It might seem intellectually unproblematic to deny the existence of free will in the way that has been described, and to claim that everything that happens is either necessary, in the sense of being determined by the preceding moment and deterministic laws of nature or social structures, or random, as in determined by the preceding moment and laws of probability. But a closer look reveals that this cannot be so.

Without human beings, there would be no concepts and statements. I consider this view to be part of a naturalistic worldview—one that does not take into account planets in other solar systems and any organisms that might live there. Within naturalism, there is no room for a Platonic world of ideas where concepts can exist on their own, independently of people. This, of course, does not prevent us from using abstraction to do away with the people behind certain statements in many contexts, and discussing the statements and their qualities *as though* they existed independently of people. But not in all situations. To begin with, I will say a few words about when—from a semantic point of view—it is either possible or impossible to analyse statements independently of people. In light of this, I will then turn to the statement 'free will does not exist'.

Many statements can in principle be considered either true or false. When it comes to general empirical statements (such as 'water molecules consist of two hydrogen atoms and one oxygen atom', which is true, and 'water molecules consist of three carbon atoms and two helium atoms', which is false), mathematical statements (such as '2 + 3 = 5', which is true, and '2 + 3 = 7', which is false), and statements of a simple formal-logical nature (such as 'p or non-p', which is true, and 'p and non-p', which is false), the question of the statement's truth or falsehood can be decided independently of who the speaker is. When it comes to statements such as 'I have two legs', 'here is a house', and 'there is currently a lunar eclipse', however, we must know by whom, where, and when each respective statement was spoken before we can determine whether it is true or false. So far, no problems in principle.

Let us now consider a version of the liar paradox. As an example, I will use the statement 'all Europeans always lie'. In and of itself, there is nothing strange about this statement, which could be a regular empirical statement.

© ProtoSociology

Each person could in theory start investigating whether it is true or false. But if the statement is made by a European, a semantic curiosity arises. If what the European says is true, this person has—according to her own truthful statement—lied, and the statement should be considered false. But if the statement is false, it might be true. Put more succinctly: If true then false, and if false then possibly true. A paradox.

Much has been written by philosophers about the liar paradox. My own suggested solution is to view the paradox as an utterance that falsely appears to contain a statement—that is, only seems to have a content that can be either true or false. The utterance 'all Europeans always lie' spoken by a European does not in fact contain a statement. But the solution to the paradox is not relevant to this book. The purpose of mentioning the paradox is to show that semantics is not an entirely problem-free field of knowledge. More specifically, it is a field in which not all problems can be solved by importing methods from the natural sciences.

A similar yet different semantic problem is posed by certain paradoxical utterances known as *performative contradictions*. When the speaker has been removed through abstraction, such statements appear unproblematic. Yet if the speaker is included, a semantic paradox arises of a kind that makes the term 'contradiction' adequate. Let me explain.

Here follows three statements that are, at the time of writing, false: 'Ingvar is dead', 'Ingvar can no longer speak', and 'Ingvar has lost all his English'. If someone makes these three statements in a conversation where I am absent, no paradox arises. The question of whether the statements are true or false can be decided empirically. But if I am participating in the conversation myself—saying 'I am dead', 'I can no longer speak', and 'I have lost all my English'—a kind of linguistic paradox arises that is known as a performative contradiction. The content of my speech act is contradicted by what my speech act as such shows the listeners. In some sense, I am contradicting myself; but the contradiction is neither logical nor semantic in the sense of existing in my statements independently of me as a speaker.

If a bystander were to describe my utterances, it would sound like this: 'Ingvar says he does not exist', 'Ingvar says he cannot speak', and 'Ingvar says he has lost all his English'.

The three performative contradictions I have used are examples that can be described as *context-independent*. To identify the performative contradiction, it is sufficient to understand each statement in relation to its speaker, all of which is made clear by the bystander's description. But there is also another, often neglected, type of performative contradiction that I will term *discourse-specific*.

Such performative contradictions require that the statement is made within a certain kind of discourse. And it is this type of performative contradiction that is relevant to my defence of free will.

Typically, a statement is made within a given kind of discourse. For the purposes of this book, it is sufficient to differentiate between purely narrative discourses and argumentative discourses. Let us consider the statement summarising Chapter 2: 'Everything happens due to causal necessity or chance; therefore, free will does not exist.' If someone makes this statement while giving a lecture on their worldview, no semantic curiosity arises. The audience can try to determine for themselves whether it is true or false. But what happens if the statement is made in a situation where the speaker is arguing with another person about the existence of free will?

Normally, we distinguish between forcing another person to change opinion by using threats or bribes and changing a person's opinion by using arguments. Only in the latter case is the other person really made to think like oneself. The arguments used inspire the other person to change position freely and voluntarily—not just saying that she has changed opinion. In other words, in an argumentative discourse the other's freedom to change or refuse to change opinion is taken for granted.

If a person engaged in argumentative discourse says that 'everything happens due to causal necessity or chance; therefore, free will does not exist', what she is saying thus contradicts the very discourse she is accepting and within the context of which the statement is made. Just like in the case of context-independent performative contradictions, the statement contradicts something that the speaker is showing through her speech act—namely that she accepts the argumentative discourse within which the statement is being made.

If a bystander were to describe the utterance in question, her description would sound something like this: 'She is arguing that there is no such thing as free will, saying that everything happens due to causal necessity or chance, and therefore free will does not exist.' It can be summarised as follows: 'She is arguing that argumentation is impossible.' A performative contradiction is here described.

Argumentative discourse exists beyond the academic seminar room. It exists in all ordinary squabbles about facts and normative issues. In particular, it exists in many medical and engineering contexts. Often, it is not obvious to a group of experts which surgery/therapy or construction is the most appropriate. Expert groups often have to argue it out.

In the mid-1850s, the Danish philosopher Soren Kierkegaard (1813–1855) described with ironic precision another type of discourse-specific perfor-

mative contradiction—one that is often referred to as hypocritical religious devotion:

> In the magnificent cathedral the Honourable and Right Reverend Geheime-General-Ober-Hof-Prädikant, the elect favorite of the fashionable world, appears before an elect company and preaches *with emotion* upon the text he himself elected: 'God hath elected the base things of the world, and the things that are despised'—and nobody laughs. (Soren Kierkegaard, *Attack upon 'Christendom'*, 2020, p. 217; transl. W. Lowrie)

Here, the content of the preacher's speech acts contradicts what the discourse within which the sermon is embedded implicitly shows both the preacher and the audience—that is, that the preacher himself has a normative acceptance of the fact that many of the world's chosen, especially his audience, are not despised. As the quotation shows, discourse-specific performative contradictions do not always present themselves as explicit contradictions. Alas!

What I have now elucidated in detail with the help of the structure of discourse-specific performative contradictions is not an entirely new idea within the history of philosophy. The view that there is something self-contradictory about determinism has been expressed in more intuitive terms numerous times. Epicurus (341–270 BC) of the ancient world appears to have been first:

> He who says that all things happen by necessity cannot criticize another who says that not all things happen by necessity. For he has to admit that the assertion also happens by necessity. (Quoted from Popper and Eccles, *The Self and Its Brain*, 1977, p. 75).

Put in more general terms: determinism makes all argumentation with someone who opposes one's views meaningless.

Here are a few other, similar examples: A defender of free will comments on an opponent who believes that everything is determined. The defender says: 'If no human action can change anything, it must be meaningless for you to try to change my opinion,' or 'If everything is determined, it must be meaningless for you to try to make me think differently.'

Performative contradictions may also arise when one is debating with oneself in complete solitude. Let us consider Descartes and his in solitude formulated implication: 'I think, therefore I am'. Descartes attempted on his own to question everything, but came to the conclusion that he could not question his own existence while thinking. The typical interpretation is that the implication represents a logical-semantic implication—that Descartes, based on the premise 'I think', was able to draw the conclusion 'I am' using some regular logical-semantic syllogism. But a much more likely interpretation is that Descartes, in

his discussion with himself about what he could or could not question, came to the conclusion that, if he allowed himself to question the statement 'I am', a performative contradiction would arise. And the only way out of it is to believe that, when one thinks, one also exists. (This interpretation can be traced back to a couple of essays by the Finnish philosopher Jaakko Hintikka (1929–2015); however, he did not use the term 'performative contradiction'.)

The same monological approach can be used on performative contradictions that are discourse-specific, as well. Let us assume that—like Descartes—you are attempting to question everything, and so you allow yourself to question the statement 'My will is sometimes a little bit free'. But because your questioning constitutes a monological argumentative discourse to your denial, a performative contradiction arises. You are arguing with yourself, while at the same time trying to question whether you have free will. This is not coherent. Instead of Descartes' 'I think, therefore I am', you are forced to conclude: 'I argue with myself, therefore I ascribe to myself a certain degree of freedom'.

In his dualism, Descartes had no problem believing that thinking is free and not subject to determinism. For this reason, he had no need to discuss the existence of free will (but very much how it can interact with the human body). For Immanuel Kant (1724–1804), however, who believed that everything that happens in the world—in space and time—is fully determined, the issue became one of urgency. In his own special transcendental-philosophical manner, he came to the conclusion that the capacity for rational reasoning must be thought of as free, but also that it must exist outside the world of space and time. In his brief *Groundwork of the Metaphysics of Morals*, he turned his position into a heading: 'Freedom must be presupposed as a property of the will of all rational beings' (2013, p. 57). I mention this despite being a naturalist and thus rejecting all transcendental philosophy, as it may interest readers partial to the history of philosophy to note that, within my naturalist framework, I can offer a paraphrase of Kant's very famous quotation: *Freedom must be presupposed as a property of the will of all argumentative beings.*

The absurdity in completely denying the existence of free will consists in also denying the possibility of argumentation. Of course, this absurdity becomes particularly blatant if, at the same time, one acknowledges—as so many do—that it is specifically *scientific argumentation* that has inspired their conviction that there can be no free will.

Kierkegaard claimed that the only way to enter into the true domain of religious faith is to take a completely unfounded leap of faith—any rational argumentation is impossible. I would like to claim that the same is true of the denial of free will. There is no sustainable reasoning that can uphold a natural-

 © ProtoSociology

ism without free will; though you may leap to this position, of course. There is a difference between these two cases, however. The concept of God as an omniscient, omnipotent, and omnibenevolent entity, yet who accepts suffering in the world, is a logical contradiction. The opinion that free will does not exist, on the other hand, is not a logical contradiction. But it gives rise to an absurd worldview due to its latent performative contradiction.

The solution to the dilemma that arises when pitting the conclusions drawn in Chapters 2 and 3 against each other is to show that there is a place for free will in our modern, naturalistic, and evolutionary worldview after all. This is what I will attempt to do next.

Chapter 4

The core of the problem elucidated

To me, freedom of action and free will go hand in hand in the following way: without free will, no freedom of action; but a free will does not guarantee freedom of action. Even if the desired action is freely chosen, many factors can make the action impossible. One might also phrase it this way: free will is a necessary but insufficient condition for freedom of action.

Many philosophers, however—the so-called compatibilists—are not willing to recognise free will as a necessary condition for freedom of action. For them, the will being unfree does not preclude the action from being free. These philosophers consider free will to be fully compatible with determinism. They simply define freedom of action as the combination of being able to take a certain action and actually wanting to do so—regardless of whether the will is caused entirely by a brain state or by the causal interplay between given beliefs and desires. Strangely, they claim that there is no contradiction between the views that humans have freedom of action and that everything is determined. But to hold that an action is free even when the will that brings it about is unfree appears to me like a corruption of everyday language. In ungenerous moments, this seems to me like the type of displacement of meaning that George Orwell (1903–1950) referred to as 'Newspeak' in his dystopian novel *Nineteen Eighty-Four* (1949), and which he described as a language that diminishes the citizens' ability to think freely and critically.

The kind of free will I seek to defend is, in current philosophical terminology, an *incompatibilist* free will. That is, if it exists, it also follows that not everything that happens in this world happens according to necessity and/or chance. For those readers who are not professional philosophers, I am simply defending free will in its traditional sense.

In neither the neuro-psychological nor the belief–desire model of the human will is there any freedom in the sense that I am seeking to defend. However, it is important to note that the will being free does not mean that it is unaffected by the individual's brain state or beliefs and desires—simply that the content of their will is not *completely* determined or random. As I have already claimed in Chapter 2: no person's will can be entirely above all the needs and desires that arise during the course of one's life.

Often, many of us reflect on our beliefs and desires in order to arrive at a decision about what to do—that is, in order to arrive at a determinate will. But sometimes we want to do two incompatible things, for example, both

keep working and leave work early to go out and have fun. In such cases, we normally begin to reflect on these irreconcilable wants in the same way we otherwise do in order to arrive at what we want in the first place. Most people are capable of creating what may aptly be termed a *second-order will*—that is, a will that stands above an already given, first-order will. But this possibility does not solve the problem of free will. Because *either* this second-order will occurs by necessity and/or chance—and if so the chosen will of the first order is also determined by necessity and/or chance and thus not free—*or* a will of the second order can be free, but explaining how this is so is no easier than explaining how it is that a will of the first order is free.

If our will is free, we must be able to explain this at the first level of the will. To claim that freedom only arises on the level of the second order does not help us explain it. It only serves to reckon those people who reflect a lot on their actions to be free, while leaving people who do not to be unfree. With this remark, I move on.

When modern physics was invented in the 17th century, it introduced a new distinction between two types of phenomena and qualities. One type consisted of phenomena and qualities that were assumed to exist on their own in the material world, independently of any observer—and these could be studied by physics. They included, for example, spatial extension, weight, shape, and motion, which were known as *primary* qualities. The second type consisted of phenomena and qualities that were assumed to only exist in human perception. These included, for example, colours, sounds, smells, and tastes, and were known as *secondary* qualities. The primary qualities were assumed to cause the secondary.

Descartes drew an even sharper line between primary and secondary qualities by claiming that the world consists of two completely different kinds of substances: material and mental/spiritual. The nature of the former is to have spatial extension, while that of the latter is to be thinking in a broad sense—that is, to engage in conscious mental activity. In everyday life, we take for granted *on the one hand* that changes in our bodies (such as consuming food, drink, and medication, as well as bodily injuries) can affect what happens to us mentally (such as feelings of wellbeing or discomfort), and *on the other* that our will can affect what we do with our bodies (if I want to bring the food to my mouth, I do so, etc). With Descartes' dualism, this common truth is turned into a metaphysical problem: how can these two essentially different types of substances affect each other? Material substances can, according to Descartes, only be affected causally through impact by collision. But how, then, can a mental/spiritual substance collide with a material one? Often, Descartes writes as though mental/spiritual substances completely lack any spatial extension

and only exists in time. But how is something non-spatial supposed to affect something existing in space?

Today we can leave Descartes' ideas behind. Since his time, physicists have allowed themselves several different ways of thinking about causal relationships or conditions of influence. No physicist of the present era talks about collisions between subatomic particles, between atoms, or between molecules. Yet despite this, many philosophers have retained a Descartes-like problem of causation. They no longer conceive of the physical and the mental/spiritual as a substance dualism, but rather as a property dualism. Mental/spiritual properties are thought of as belonging to a material property bearer, primarily the brain. They are assumed to be caused by and belong to states and processes in the brain. Yet a problem of interaction—albeit not Cartesian—still remains.

Many philosophers take for granted that physical features can only be affected by physical states, processes, and properties. The physical world is understood as adhering to what is called the principle of physical causal closure. Why so? The answer is that if this were not the case, physics would not be a fully autonomous science, as it is assumed to be. This view does not stop you from believing—as the theory of evolution stipulates—that material particles have given rise to mental/spiritual phenomena. But it does not allow you to believe that the mental/spiritual has an effect on the material/bodily. If, according to tradition, we refer to the material as a lower or underlying level and the mental/spiritual as a higher one, the philosophers I have in mind can be said to accept so-called upward or bottom-up causation but reject downward or top-down causation—also known as mental causation. In other words, they reject that something mental/spiritual can change the bodily, but they accept that the bodily can change the mental/spiritual.

Any evolutionary defence of free will must accept some version of upward causation, as the will—with its mental/spiritual component—only arose *after* purely material phenomena in evolution. But also because pain relief is such an obvious example; we all know that the mental phenomenon of pain can be removed with the help of various chemical ingredients injected into the body.

Downward causation—that is, something mental/spiritual impacting something material/bodily—on the other hand, is a problem that the defence of free will must, in some way, resolve. At the very least, free will must be able to affect the brain in order to have anything to do with freedom of action. This problem I will be addressing in the chapters that follow. A few conclusions can be offered even at this stage, however: Downward causation does not exist. Therefore, in order to solve the problem of how the will can affect the body, we need other concepts beside that of causation.

 © ProtoSociology

Chapter 5

Emergence and spontaneity in the natural sciences

Placing free will within a naturalistic, evolutionary framework requires applicability of the concepts of *emergence* and *spontaneity*, which this chapter will introduce. The mental phenomenon of the will has emerged—that is, it has arisen as an entirely new phenomenon—at some point during the course of evolution. And *free* will is an example of spontaneity—that is, of something that is not predetermined. In this chapter, however, I will limit myself to showing how these concepts are currently applied by scientists to the purely physical world. Emergent mental phenomena will be covered in Chapter 6, and spontaneity as a mental phenomenon in Chapter 8.

When something 'emerges', it generally means that it arises out of something else. Sometimes it is a matter of something qualitatively new arising out of something that already exists. This is known as *diachronic emergence*. The theory of evolution is a perfect example of a story about recurring diachronic emergence. In evolution, new species arise out of those that already exist. But something new can also emerge out of its own underlying conditions of existence. This is known as *synchronic emergence*. However, this requires that the unity in question can be divided into two levels—an emergent level and an underlying level. Let me use the linguistic sign as an initial example; there will soon be more to follow.

A linguistic sign—that is, a word—has two levels: the meaning of the word and its underlying sound-image, i.e., the word as written or spoken. The meaning of the word arises/emerges out of this sound-image, yet it is qualitatively different from it. The same meaning can emerge out of completely different sound-images. Take, for example, 'yellow' in English and 'amarillo' in Spanish. But where there is no sound-image at all—as between these quotation marks ' '—there is also no linguistic meaning to decipher. When encountering words in a language I do not know, all I perceive is the sound-image. Here are two examples of words that, according to Google Translate, have the same meaning as 'yellow': 'رفصألا' and 'الىيپ'. Yet to me, no meaning emerges; all I see are two images.

With the concept of synchronic emergence in our toolbox, we can meaningfully argue (rightly or wrongly) that, for example, macrophysical objects, properties, and relationships emerge synchronically out of microphysical con-

ditions. Naturally, at certain moments both types of emergence (diachronic and synchronic) may come into play at the same time: what the emergent phenomenon arises out of may simultaneously be transformed into its future conditions of existence. Here, I am only speaking of *ontological* emergence—that is, what and in what way something exists. The term 'emergence', which has developed a plethora of meanings, does not in this book touch on epistemological questions, such as when a qualitatively new phenomenon can first be predicted.

The concept of synchronic emergence can also be illustrated with the help of an outdated physics. Let us assume that everything in the physical world ultimately consists of tiny, indivisible, indestructible, and eternal atoms that join together in different configurations in spacetime. At different moments in time, the atoms may be found in different positions in space. And at each moment, any collection of atoms takes on a certain spatial configuration. Depending on the atoms' movements, the same collection can give rise to entirely new configurations. But at any given moment, these are still configurations consisting of the same atoms in the same space. In this sense, nothing *qualitatively* new arises in the atomistic world described above, despite the fact that new specific configurations may arise at any given time. There is thus no emergence.

Moving forward, when speaking of a collection or configuration, I will symbolise its borders as / /. Let us assume there is a collection consisting of two atoms shaped like points / · · /, one atom shaped like a line / – /, and one shaped like a curve /) /. These four atoms move about freely in space, forming various configurations. If we observe them, what we see is often precisely that: their configuration. For instance, / · · –) / represents one four-atom configuration and /) · – · / another. Now let us consider the configuration / :–) /. Here, most people would probably, in addition to the configuration, also see a smiley face. The latter is a qualitatively new ontological phenomenon in the sense that it is something more than a spatial configuration of the atoms in question. The underlying configuration of atoms is only a condition of existence for the smiley face, and thus part of an emergent whole.

At this point, some might very well object: 'Of course, no one would deny that emergent phenomena may arise in the eye of the beholder, and in this way be *subjectively* ontologically emergent. Poems, short stories, and novels always contain literary representations that are something else entirely than a mere configuration of words. And perceptual psychology, for example, examines non-verbal perceptual so-called Gestalt qualities, that is, perceived properties that are delimited in spacetime and cannot possibly be considered only the sum and configuration of their parts. Yet, as you have said, the question under

 © ProtoSociology

consideration here is whether emergent phenomena can appear in the sphere of nature—i.e., independently of perception. Am I wrong?'

My answer: No, you are not wrong; that is the question. The example of the smiley face was only meant to introduce the concept of emergence and immediately illustrate that there are undeniable phenomena to which the term refers. Yet my opinion is that ontological emergence not only exists in the sphere of experience, but also in the sphere of nature, which is independent of perception. Let me continue to use the simplified atomic theory as a pedagogical aid.

Now let us assume the existence of two types of atoms / ● / and / • /, and that between these atoms certain atomic forces/laws-of-nature arise due to properties of the atoms themselves. For instance, they may be gravitational forces based on the atoms' mass (Newton's law of gravity), and electric forces based on their positive or negative charge (Coulomb's law). These are forces that act momentarily between the atoms, even though the latter are positioned at a distance from each other. With growing distances, however, the forces quickly become weaker. Let us also assume that the two aforementioned types of atoms are sometimes bound together in a linear molecule by the atomic forces: / •●• /. Let us further assume that, when studying such molecules, it becomes clear that some of their interactions may relatively easily be described mathematically with the help of *intermolecular* forces/laws of nature—that is, a force between two molecules of a given kind. I shall symbolise this with a double-headed arrow:

/ •●• / ↔ / •●• /

One question that arises is whether these intermolecular forces can be considered identical with the sum of the atomic forces at hand.

If the true answer is yes, the intermolecular forces are said to be *reducible* to the atomic forces. Despite this, it may in calculations nevertheless be practically appropriate to use the intermolecular forces *as though* they were fundamental laws of nature. Even if the atomic forces could be used in theory, it would get too complicated. Here, emergence only exists in the eyes of practicians or non-philosophers. This type of emergence is often called *weak* emergence (though the contrast between weak and strong is not entirely unambiguous in the literature on emergence). The classical examples are macrophysical temperature, which is considered reducible to the kinetic energy of molecules and vibrating atoms, and macrophysical light rays, which are considered reducible to electromagnetic waves. If true, the relationship is similar to that between a term and

its acronym. An acronym can streamline conversations; a weak emergent law of nature can streamline both technology and scientific experiments.

If the true answer to the question is no, however, the atoms joining together into molecules have not only given rise to molecules that constitute something more than their given atomic configurations and the atomic forces holding them together. The molecules are now also unities between which a new kind of force/law of nature exists—that is, the non-reducible, intermolecular forces/laws of nature. The molecules must now be considered something more than a configuration of atoms. Thus understood, the molecules and their intermolecular forces are an example of *objective* ontological emergence, often called *strong* emergence. The new types of objects (molecules) and the new intermolecular forces emerge at the same time. Nothing prevents some property or structure inherent to the molecules from also being included in the formulation of this new force/law of nature.

Please note that the concept of emergence does not coincide with the concept of increased complexity. The relationships on the emergent level can theoretically be either simpler or more complex than the relationships on the underlying level. Deterministic natural laws can in theory emerge on indeterministic ones, and indeterministic laws on deterministic. What the emergence looks like can only be determined on a case-by-case basis.

As a possible concretisation of the abstract example above, let me mention the van der Waals force, understood in this context as a force between molecules. According to this law, the electron cloud that surrounds a molecule can assume an asymmetrical shape, which means that one side has a stronger negative charge than the other. If the end of a molecule with such a negative charge happens to be facing the end of another molecule with a positive charge, they are—for a brief moment (as the force is very weak)—bound together by the van der Waals force into one single unity. It is the *molecules as unities that are bound together*, yet the intermolecular force is dependent on the shape of their electron clouds.

In Chapter 2, I made clear that today's story of evolution is based on a conglomerate of different theories. I will now add that current science does not provide even a draft of something that could prove that all of the emergence that the story of evolution appears to contain is weak emergence. Of course, this does not prove my position that it must in some cases be a matter of strong ontological emergence. Yet there is one case that offers good reason to believe in recurring objective strong emergence in evolution. It goes like this.

If all apparent evolutionary emergence turns out to be weak, the theories in question must be reduced to a purely physical theory. This is because the

 © ProtoSociology

story's beginning describes purely physical conditions. The easiest theories to thus reduce ought to be the ones in chemistry. This question has been discussed avidly within the philosophy of chemistry in the 21st century. To me, it appears the non-reductionists have a clear argumentative advantage.

If a reduction of chemistry into physics were possible, all characteristically chemical concepts—not only the van der Waals force, but also such concepts as valence, chemical compounds/substances, and chemical bonds—could in theory be defined in quantum-physical terms. But we are not even close to being able to do so today. Even in relation to the concepts where actual reduction has gotten the furthest, reductionists have still in the end been forced to introduce approximations of the concepts they are trying to reduce. Of course, this does not prove that reduction is impossible, and, certainly, the debate will go on within the philosophy of chemistry. But I belong to the camp that has become convinced of its impossibility.

In other words, I believe that objective strong emergence arises already on the border between physics and chemistry. It should be noted that the existence of such emergence does not make knowledge about the emergent phenomenon's conditions of existence superfluous when designing experiments. Yet the opposite is also true. Knowledge of emergent phenomena can be a big help when designing experiments to determine the properties of objects in what creates the emergence.

If, as I am convinced, strong emergence arises already on the boundary between physical and chemical phenomena, I find it extremely likely that such emergence of various kinds may appear here and there throughout evolution, and in nature today. Now, let us move on to spontaneity in nature.

We often speak of spontaneous actions and people, but natural scientists sometimes find it adequate even to speak of spontaneous processes and events. There is, though, every reason to distinguish between two such types of spontaneity: one that is subject to a law of probability, and one that is not. I will consider them one at a time, exemplifying with radioactive decay and neuronal activity respectively.

Radioactive material disintegrates independently of external forces and factors, which has led physicists to say that it happens spontaneously. Decay consists in the atomic nuclei of the material emitting a certain kind of particle (in alpha decay, one type of particle is emitted; in beta decay another). What happens in the nuclei can be described using quantum-mechanical indeterministic models. Different types of radioactive material disintegrate at different rates. To capture this fact semantically, the notion of *half-life* has been created. The half-life is the time it takes for any amount of a particular kind of material to

disintegrate until only half remains. That is, during a half-life half of the nuclei in the original amount disintegrate and half do not. With respect to a single nucleus, this means that when the process starts its *probability to disintegrate* is 0.5. The decay of a single atomic nucleus is referred to as both spontaneous and random, yet can nevertheless be ascribed a certain numerical probability. It is, I shall say, a matter of *stochastic spontaneity*. After all, the spontaneity is subject to a law of nature, albeit of the statistical kind.

The nerve cells in the brain, the neurons, are part of what could be described as the brain's signalling system. They both emit and receive electromagnetic impulses/signals. Neurologists sometimes speak of spontaneous neuronal firing. Neurons can emit electromagnetic impulses/signals without any external cause, and we do not know what happens inside the neurons during such spontaneous firing. Neither have scientists been able to identify any overarching regularity, such as the one in the case of spontaneous radioactive decay. Today, there is thus no numerical probability for spontaneous firing that can be ascribed to individual neurons. It is, I shall say, a matter of *non-stochastic spontaneity.* This spontaneity is, on its own level, not subject to any laws of nature, even if it requires certain spatiotemporal conditions for its existence.

In the eyes of the neurologists themselves, non-stochastic spontaneity probably only indicates a lack of knowledge. Most probably believe that future neurology will transform non-stochastic spontaneity into stochastic spontaneity. And this is, of course, true within the necessity-and/or-chance paradigm. But if one is pondering the problem of free will, like I am, and proceeds from the position that it is absurd to completely deny the existence of free will, one should also consider the possibility that non-stochastic spontaneity may have in the strong sense of emergence emerged in evolution, too. But the will is only one of many mental phenomena, and all of them must have emerged. That is the topic of the next chapter. If free will exists in the incompatibilist sense I have delimited, it must be as non-stochastic spontaneity in a mental phenomenon.

 © ProtoSociology

Chapter 6

Awareness phenomena and evolution

If consciousness or conscious mental phenomena exist, they must be strongly emergent phenomena. There are no mental phenomena mentioned at the beginning of the evolutionary story, which is described using theories from physics and chemistry—and in these kinds of theories, no mental phenomena are described.

Traditionally, we often speak of *consciousness* as though it were an independently existing object—the mind. This use of the term fits well within the Cartesian dualism described above, but not the opinions I wish to put forth. As the title of the chapter indicates, I prefer to speak of conscious mental phenomena as *awareness phenomena* instead. The first question is: do they exist?

The answer is yes, without a doubt. The difference between having and not having awareness phenomena is one we encounter every day as we move between being awake or dreaming on the one hand and dreamless sleep on the other. For those of us who take a dreamless nap at some point in the middle of the day, the difference becomes even more pronounced. That is, the difference as seen from one's personal perspective, from a first-person perspective. Seen from the outside—from a third-person perspective—sleep researchers claim that the difference is one of brain activity. But from inside the self, the difference lies in *having* awareness phenomena while awake or dreaming, and *lacking* them (as noted in hindsight) when engaged in dreamless sleep. The brain might be able to solve problems even when we are sleeping dreamlessly, but such thought processes are no awareness phenomena. They are better compared to processes like those involved in constructions of artificial intelligence.

For those who do not consider these remarks as proof that awareness phenomena exist, there is a stronger argument yet. It is very similar to Descartes' famous and spurious proof of the existence of an independent, thinking substance—'I think, therefore I am.' Descartes attempted to question everything, but came to the conclusion that it was impossible for him to question that he and his substantial self exist in the moment that he is thinking. If we are more careful with the notion of a self, as there is reason to be, Descartes might have settled for: 'Now I am thinking, therefore an awareness phenomenon currently exists.' This statement is, in my opinion, an irrefutable truth every time someone speaks it or thinks it.

The existence of awareness phenomena is thus, from each first-person perspective, more clearly evidence-based than any other empirical fact. Yet—albeit

with a little less certainty—the evidence that awareness phenomena cannot exist without a neurological basis is overwhelming. In other words: There are no (as Descartes believed) awareness phenomena that are completely independent of a material substrate. Some types of neurological processes function as *conditions of existence* for awareness phenomena, but this does not mean that these phenomena are *identical* with the neurological processes.

There is a difference between the (erroneous) claim that 'A thought *is nothing but* a certain type of activity in the neurons in the brain' and the (correct) one that 'A thought *requires* for its existence a certain type of activity in the neurons in the brain'.

Before such neurological processes arose during the course of evolution, awareness phenomena could not have existed. Today we can ask questions like: When, during the course of evolution, did the first awareness phenomena arise? How did they arise? What functions did they originally fill, and what functions do they fill today? Has there been a gradual development from simple, minimal forms of awareness phenomena to the very complex forms existing in humans today? Yet it is clear that we still lack good answers. In light of our current knowledge of the complicated systems that relatively simple plants and animals have developed for retrieving and processing information from their surroundings, as well as modern constructions of artificial intelligence, it is hard to point out abilities and mechanisms that could *not possibly* function without being tied to some awareness phenomenon. However, these difficulties in formulating an explanation cannot count as an argument for the non-existence of awareness phenomena. Anyone able to take on a reflective first-person perspective realises without a doubt that awareness phenomena do exist. The difficulties in formulating an explanation only point to the need for more research and philosophising into and around these questions.

Determining when awareness phenomena first arose is an extremely difficult task, of course. A different and more doable one would be to investigate what the neurological activity looks like in the brain when we either dream or sleep without dreaming. This can be done in laboratories. When possibilities for observing and localising different activities in the brain have come further than today—with the help of positron emission tomography (PET), functional magnetic resonance imaging (fMRI), and computed tomography (CT)—it will most likely be possible to identify significant correlations between various types of awareness phenomena and their underlying brain activity.

A common statistical-scientific problem nevertheless remains, of course: how does one, based on statistical data relationships, identify an underlying causal mechanism that gives rise to said relationships? And the problem does not end

 © ProtoSociology

there. If, in a lab, one could identify the mechanisms that give rise to dreams and dreamless sleep respectively, it would still not solve the problem of what such mechanisms look like in everyday perceptions and emotions.

Hypotheses have already been formulated, however. I will mention one that, in its general form, appears to me intuitively reasonable: Giulio Tononi's integrated information theory. Summarised in one sentence, the theory suggests that awareness phenomena arise when several of the brain's information systems interact.

The existence of awareness phenomena is, as I have mentioned, an irrefutable truth that no evolutionary theory can deny. Yet what properties do they have that distinguish them from the properties described by physics and chemistry? The general overarching answer is most easily formulated using two philosophical concepts: *qualia* and *intentionality*. Typically, awareness phenomena contain both qualia and intentionality, while phenomena described by physics and chemistry lack both kinds of properties.

'Qualia' is an umbrella term for the type of phenomena and properties that we experience qualitatively through perception, emotions, dreams, and sensations. This includes colours (though not electromagnetic waves), sounds (though not the structure of air compressions), flavours (though not chemical reactions on our tongues), smells (though not molecules in the nose), and the experience of pain and other bodily sensations (though no neuronal processes). None of these experiences are the subject of theories in physics or chemistry, which only offer explanations of how various kinds of qualia may arise.

Intentionality involves directionality, but a different kind than the directionality of physical bodies in motion, or the one represented by vectors in mathematical physics.

Intentionality is the type of directionality that exists in our perceptions, emotions, dreams, and sensations. In everyday language, intentionality often reveals itself in our use of prepositions. I look *at* something, I take hold *of* something, I am angry *at* someone, I am devastated *by* something, I am thinking *of* something, and so on. There is a direction from one thing to another; the sentences contain a from–to structure. In the examples provided, the direction is from an 'I' to something else. But in order to also accommodate further examples, I will say more generally that they contain a structure with a 'from-pole' and a 'to-pole'.

The from–to structure is obvious in perceptions of the outside world. In this case, the from-pole can be identified as a person and the to-pole as an object or fact in the external world. Yet this structure also exists in awareness phenomena without such a division into a bodily inside and outside. In sensations of pain

in the body, we differentiate between an I (the from-pole) experiencing the pain and the pain itself (the to-pole) somewhere in the body. Everyday examples are headaches, toothaches, stomach aches, and joint pain.

This structure also exists when we think of such things as mathematical numbers and fairy tale creatures. There is, on the one hand, a something that is thinking (a from-pole) and, on the other, a something that is being thought of (a to-pole), i.e., the mathematical numbers and the fairy tale creatures respectively. When we reflect on our dreams upon waking, we realise that, even though as a whole they are awareness phenomena created by the brain, they nonetheless have a from–to structure similar to that of ordinary perception. In dreams, there is a seemingly real perceiving subject (a from-pole) and a seemingly independently existing—albeit often very strange—world (a to-pole).

When I perceive something in my surroundings or sense something inside my body, my awareness phenomena are immediately directed at that which I am perceiving or sensing. When, on the other hand, I read about an event that has happened or see an image of it, my awareness phenomena are directed at the event in question *via* the text or image. When I read an illustrated fairy tale, my consciousness is directed at fictive people and events *via* the text and images. A from-pole may thus be directed at a to-pole both directly and indirectly. And the to-pole may be either real or fictive—or, for that matter, a mix of both.

Conscious wants and intentions only make up one of many different kinds of intentional phenomena. Their to-pole exists in the future and is, in this sense, fictive. But if the will/intention is realised, something real comes into existence in spacetime.

The will as an awareness phenomenon per definition contains intentionality. Yet normally it also contains some kind of qualia. When I want something, the will is often experienced in some way. Though sometimes, like in the case of calm conversations, it can be almost qualia-free. In the moment, the will only expresses itself in calm statements of the kind 'I want this and that'.

Of course, the fact that awareness phenomena irrefutably exist and that a conscious will exists as such does not solve the problem of whether this will is free. The will, with its various qualia and features of intentionality, is perhaps always subject to the necessity-and/or-chance paradigm described in Chapter 2. What I have said in this chapter does not prove that the will is free, since I have not yet addressed the problem of downward causation identified in Chapter 5—that is, when something mental/spiritual is claimed to be the cause of something bodily. But in order to deal with this problem, we might do well to consider what the situation looked like before the problem of free will became so acute. Perhaps it might help us broaden our thinking? Sometimes, ideas for

good philosophical arguments can be borrowed from philosophers who, in certain other respects, are completely outdated. In this case, I will be turning to Aristotle (384–322 BC). He thought that body and soul make up a unity.

Chapter 7

Free will before the Scientific Revolution

During the Middle Ages, free will did not trouble philosophers contemplating nature—rather, it was a problem for theologians. If God was an absolute, omniscient being, he must be able to know in advance what decisions people will make. Yet this did not correspond well to the common Christian-theologian view that all sinful acts arise out of the free will of human beings—and are thus not predictable. The fact that human beings have a certain freedom of will and action appears to be evident from the story of creation in the Bible. It states: 'So God created man in his *own* image, in the image of God created he him' (Genesis 1:27, *King James Bible*). As this similarity cannot be of the bodily kind, it is reasonable to interpret the sentence as suggesting that humans—unlike animals—are endowed with some of the same mental-spiritual freedom as God.

If we allow ourselves a little banter, we might say that, while in the minds of the scholastic philosophers the problem of free will arose out of God's supposed omniscience, in the minds of contemporary philosophers the problem arises from the omniscience ascribed to the natural sciences—that is, the view that everything that happens is governed by the necessity-and/or-chance paradigm.

There was natural science during the Middle Ages, too. But back then, it went under the name of 'natural philosophy'. The investigations made were heavily influenced by Aristotle, and did not pose the same problem for a belief in free will as the Scientific Revolution would later do. In this chapter, I will use a few elements of Aristotle's thinking to illustrate why, during the Middle Ages, so many sophisticated natural philosophers did not take issue with the supposed free will of human beings. This took place in the history of philosophy before the theory of evolution entered the stage, of course. But I believe that some of the elements of Aristotle's static worldview can be modified and transposed to our evolutionary one.

Aristotle is of the opinion that every object existing in spacetime is a complex unity, a fusion of at least two aspects: form and matter. This view is known as hylomorphism (matter-form-ism). A form cannot exist without matter. Expressed in the terminology of Chapter 5, form-matter unities are *synchronically* emergent objects, where the form is the emerging level.

The forms that Aristotle talks about of have nothing to do with geometrical

© ProtoSociology

forms/shapes. Rather, they are the identity-giving property of the unity in question. And because Aristotle views such properties as *capacities* or *functions*, the reader may prefer to replace the term 'form' with either 'capacity' or 'function' when reading what follows. In English, Aristotle's concept of form is reflected in the expression that one's childhood is a particularly *formative* period in one's life.

According to Aristotle, there is a hierarchy of form-matter unities. The form typical of a human being is the capacity to think, which for its existence requires a certain matter, i.e., a body. But if we consider this matter (body) in isolation from the capacity to think, we discover that it—in turn—consists of another form and an even more basic matter. Here, the form is our capacity for perception and movement. If we consider its matter in isolation from this form, we find yet another form—the capacity to absorb nutrients—and yet another matter. This non-living matter (earth, water, air, fire) can, somewhat anachronistically, be referred to as 'the physical-chemical elements'. A human being is thus a hierarchy of forms and matter.

The aforementioned hierarchy makes it natural to say that a form-matter unity has two *levels*: an overlying form and an underlying matter. Of course, this hierarchy can also be presented from the bottom up. That would sound something like this: The physical-chemical elements are matter for the capacity to absorb nutrients. This form-matter unity, in turn, is matter for our capacity for perception and movement, the form-matter unity which, finally, is matter for the capacity to think. At the bottom are the physical-chemical elements—without them, no capacity for thought.

Aristotle also had a theory about four kinds of causes. The kind most closely related to today's philosophical cluster of concepts is Aristotle's notion of *efficient causality*. This is a causal relationship existing *between* certain individual form-matter unities. The relationship can by no means be reduced to a strong statistical correlation between the cause and its effect. Instead, the cause is conceived as giving rise to the effect; the cause is assumed to have force and power.

If one were to conduct a radical thought experiment, asking what Aristotle would have said if he had known of and accepted Newton's law of gravity, my guess is as follows. The law says that between two bodies with a mass, there is at every moment a force (F) proportional to (~) their mass (m) and in inverse proportion to the square of the distance (r) between them: $F \sim m_1 m_2 / r^2$. I think that Aristotle would have responded something along the following lines:

> Ok, we have to broaden the concept of efficient causality to also include momentary *reciprocity* between form-matter unities. How inter-

> esting that there can be such exact numerical causal relationships also here in our world, below the celestial sphere. This I had not exactly expected.

Independently of external efficient causality, however, a form-matter unity may also in itself contain another kind of causality: *final causality* or goal-oriented causality. It can be outward-oriented and have as its goal to change—through efficient causality—something other than the unity in which it exists, or it can be inward-oriented and have as its goal to change only the unity that harbours it. Here follows two examples, one of each kind—both come from Aristotle: When a block of marble is transformed by a sculptor into a statue (the goal), the goal-oriented (statue-oriented) cause exists outside the block of marble, in the sculptor. But when an acorn grows into an oak (the goal), the goal-oriented (oak-oriented) cause exists inside the acorn itself. Somehow, the fully grown oak can be said to exist within the acorn, but only as a potentiality and potency.

This kind of potentiality is something more than a general possibility in the trivial sense that, if something has come to exist in the world, before it arose there must have been a possibility that it might. Yet it is also something more than a simple disposition to, in certain situations, become something else. Dispositions are not goal-oriented.

Even though acorns lack consciousness and the capacity to think, they are considered to carry within them a potency to become fully grown oaks. The actualisation of this potency can be prevented by a number of causes, but as long as the acorn or an oak seedling exists, their inherent potency is nevertheless posited to exist. The same is true of human children. Regardless of whether they are thinking about it or not, there is, inside their bodies, a striving to reach a certain goal: to grow into an adult body. But once they have reached a certain age, they can, within the framework of this biologically endowed subconscious goal-seeking, also use their thinking to *freely make decisions* about certain things they wish to do. Such decisions create conscious, goal-oriented causes—we may call them free, conscious, goal-oriented causes. These can be either outward-oriented or inward-oriented.

The view that a form-matter unity that is not prevented *may on its own* give rise to changes concerning itself was not completely quashed in the Scientific Revolution—though it was radically restricted. According to Newton's first law, a body in motion not affected by any outside force continues to, on its own, *change its position in space* with an unchanged velocity along a straight line. But it cannot, of its own accord, change its geometrical shape, size, or mass, nor can it be goal-oriented. A similar idea lives on in modern particle

 © ProtoSociology

physics. A particle in motion is in many contexts assumed to keep moving forward on its own, but not be able to change its inner properties or seek a well-specified, predetermined target.

When Aristotle speaks of goal-oriented causality, it is—as mentioned—not a causal relationship between two unities, but a causality that pervades and *exists within* a certain form-matter unity as a whole. The causality is active until the goal has been achieved (for example, a new statue has been erected, or a seedling has grown into a mature oak), but not beyond. Between form and matter, there is neither efficient nor final (goal-oriented) causality. Neither can form and matter be seen—in the sense of modern mathematical physics—as two partial causal factors to what happens to the unity as a whole. Instead, form and matter are fused into a non-summative unity—a unity where form and matter have in some sense merged.

Aristotle calls a unity's form its *formal causality,* and the unity's matter its *material causality*. But these notions are in no way connected with modern concepts of causality. Aristotle does not say much about how form and matter can merge into form-matter unities. An appropriate philosophical term for their way of being united is not difficult to find, however. Form and matter can be said to be aspects that *constitute* the unity. There is no causal relationship between a form-matter unity and its two aspects, neither in an Aristotelian nor in any common modern sense. But there is a relationship of constitution. The unity and its constitutive aspects are not identical, but share a co-location, and are asymmetric in the sense that the unity becomes constituted by its aspects, not the other way around.

Aristotle's lack of a good explanation of the constitutive relationship might appear like a weakness; yet it is a weakness that Aristotle's form-matter unities share with many unities postulated by the fundamental theories of modern physics. Thus the lack cannot be used as a reason to reject all thinking in terms of form-matter unities. Let me offer two classical and one ultra-modern example from physics:

The smallest bodies subject to Newton's three laws of motion and his law of gravity have extension in three-dimensions, a geometrical shape, and a mass. But the way in which these three properties constitute said bodies, Newton does not care to discuss. At every point in the spacetime continuum, the electric and magnetic fields described by Maxwell's equations have a field strength. But the way in which all these points come together to constitute a continuous field is not explained by the theory. According to the current standard model for elementary particles, an electron has three properties: a specific electric charge ($-1{,}602 \times 10^{-19}$ C), a specific mass (9.109×10^{-31} kg), and a specific spin (1/2).

Yet the electron is not seen as a collection of three independent properties, but rather as a well-integrated unity that harbours said properties. How these properties constitute the particle unity is not something particle physicists are interested in.

I do not find it entirely misleading to say that, in fact, modern fundamental physics implicitly relies on a kind of form-matter thinking where properties function as matter for various kinds of unity-creating forms. In the examples above, this results in the classical particle, the classical field, and quantum particles, respectively.

Aristotle's form-matter unities are of such nature that the form (that is, the unity's capacity or function) can cease to exist if the matter changes. But it can also *remain the same* if the matter changes within certain limits. A single form can be realised in many ways. Using a modern analytic-philosophical term, the form is *multiply realisable*.

But can the *form* change without the matter changing? As far as I know, this is not something Aristotle ever discussed. And I believe this might be because he was not interested in minor changes in form—that is, changes of degree in a capacity or function. The problem can be situated in relation to a well-discussed concept in modern analytic philosophy known as *supervenience*. Formulated in terms of a general relationship between two types of properties, A and B, A can be said to supervene on B if the following is true: no difference in A without a difference in B.

The principle of supervenience mainly builds on two intuitions. One is that, if you think a certain painting with certain colours and patterns (B-properties) is beautiful (A-property), you are illogical if you do not find another painting with *exactly the same* B-properties to be beautiful as well. In other words: no difference in beauty without a difference in underlying properties. The other intuition, defended by many modern philosophers concerned with the mind–body problem, is that *one* particular kind of neural substrate cannot give rise to different kinds of mental phenomena. In other words: no difference in mental phenomena (A-property) without a difference in neural substrate (B-property).

Applied to Aristotle, the principle of supervenience would mean that two different forms cannot be realised by the same matter. Some of his examples may perhaps correspond to this principle, but not all. Yet the principle of supervenience also implies a weaker principle that I would like to call the principle of supervenience for changes: no *change* in A-properties without a *change* in B-properties. As a general principle, I believe this would fit Aristotle better. In relation to form-matter unities, the principle can be formulated as

 © ProtoSociology

follows: no change in a form without a change in matter. I will return to this principle in the next chapter.

In the terms of modern analytic philosophy, the form in a form-matter unity is multiply realisable and subject to the principle of supervenience for changes. The unity as a whole is a synchronically emergent object.

From this characterisation it follows that, when a form-matter unity is affected by external efficient causality, that is, by another form-matter unity, either only the matter changes or both the form and its matter change—*never just the form*. When both change, it should *not* be imagined in any of the following two ways: (i) initially the external cause only affects the form, which in turn affects the matter through downward causation; or (ii) initially the external cause only affects the matter, which in turn affects the form through upward causation. Both aspects of the constituted unity are affected directly and simultaneously. When a form-matter unity (like the acorn) contains an inward- and goal-oriented cause, it directly affects the two aspects of the constituted form-matter unity at the same time. The concepts of downward and upward causation have no place in this line of reasoning.

To anticipate some of what is to follow in the next chapter: this is the direction in which I consider it fruitful to think of awareness-phenomena-and-body unities.

As I see it, what allowed medieval natural philosophers inspired by Aristotle to unproblematically accept the existence of free will was primarily two ingredients. One of these I accept, and the other I reject.

The will and decision-making were not seen as pure mental phenomena, but as aspects of a form-matter unity—that is, the mind-body unity. The modern problem of downward causation simply did not exist. Whether such an Aristotelian understanding of the will can be affirmed today will be discussed in the next chapter—and the answer is yes.

A determined will is per definition a goal-oriented cause, regardless of whether it is a partially free creation or has arisen due to necessity and/or chance. Wanting something is just to harbour a goal-orientated cause. As noted above, the Aristotelians postulated the existence of goal-oriented causes also in objects lacking awareness phenomena. This is not custom in today's science, nor do I think it should be. This part of the Aristotelian worldview must be rejected.

Yet even though I do not defend this view, I shall take some time to explain it. The explanation casts some general light on the problem of free will, and shows that Aristotle and the medieval Aristotelians were not guilty of some simple, childish anthropomorphising.

Goal-oriented *causes* in the sense that I am speaking of must be distinguished

from goal-oriented *constructions* in a technological sense, for instance, guided missiles and automatic temperature control systems. No guided missile or temperature control system contains as an intrinsic property a goal-oriented cause. If a guided missile is not launched from its pre-programmed location, it will not be seeking the intended target. An acorn as seen through an Aristotelian lens, on the other hand, carries its goal-oriented cause within itself, regardless of where it is placed. It should also be noted that, from the perspective of botanical-technological knowledge, it could be helpful to consider and discuss plants *as though* they were goal-oriented organisms. I do not want to ban expressions such as 'plants reach for the light' or 'plants tell us how much light they want'. Philosophy and science should be tolerant of everyday language. We still speak of 'sunrises' and 'sunsets', even though we all know that these beautiful phenomena are caused by the Earth's rotation and not by movements of the sun.

The non-metaphorical concept of *goal-oriented causality* contains two parts: Firstly, once the goal has been achieved, the cause no longer has any effect. Secondly, it is assumed that *a goal can exist in two different modes*, either unrealised or realised. For example, if it is my goal and intention to clean my apartment, my activity ceases when the apartment has been cleaned. Initially, my goal only exists in some of my awareness phenomena as an idea about a possible state of affairs; in this sense, it is a fiction. But when the apartment has been cleaned, the goal has turned into an actual fact in the world.

The Aristotelians did not believe that plants and animals have conscious intentions, yet nevertheless ascribed to them a goal-oriented causality. However, they did not deny the incontestable truth that speaking about goal-oriented causes means presuming that the goal in question can exist in two modes. According to them, one and the same object in the world can in fact exist in two different modes. It can exist either *potentially* or *actually*, and this difference has nothing to do with the human consciousness or imagination.

Plant seeds and animal embryos were believed to contain, as a potentiality, their fully grown counterparts. This potentiality also contained a potency or striving to become fully grown plants and animals. Even the physical-chemical elements—earth, water, air, and fire—were considered able to harbour such causality. Each element had a natural place in the universe. And when a piece of the element was out of place, it harboured a goal-oriented cause to reach it. Yet once it had reached its goal, it would remain there, at rest. For earth, this place was the centre of the universe. This was considered to explain why pieces of matter mostly consisting of earth fell toward the ground—thereby moving closer to the presumed centre of the universe, namely the centre of the Earth.

 © ProtoSociology

The Aristotelian approach probably made it easier than today to hold the opinion that the human will may sometimes be free. The contrast between human beings with awareness phenomena and life lacking awareness phenomena was not as significant as it is today. Therefore, if humans are assumed to have a certain freedom of choice that plants and many animals lack, this difference did not appear to be especially dramatic.

For the sake of deep understanding, perhaps it is also worth noting that in those days the view on necessity and chance differed from that of today. It was not the case that philosophers lacked concepts for a distinction between states and processes that happen by necessity or by chance, respectively, but these concepts were not as embedded in mathematics as they are today. Necessities were not viewed in light of mathematically formulated deterministic theories, and what happened by chance was not seen through the lens of indeterministic theories formulated using mathematical probability. Per definition, free will cannot be defined by any mathematical relationship. Today, therefore, the belief in a somewhat free will comes into conflict with the widely held view that everything in this world can be represented mathematically. Such a conflict simply did not exist in the Middle Ages.

That said, I once more wish to emphasise that I have no intention of bringing back the Aristotelian notion of consciousness-independent goal-oriented causes. Or, for that matter, any kind of vitalist principles as they were formulated in the 19th and early 20th centuries. In my opinion, goal-oriented causes can only exist where there is intentionality—and intentionality, in my opinion, can only exist as an awareness phenomenon. In other words: unrealised goals can only exist as fictive objects in awareness phenomena.

This is an appropriate place to mention the *anthropic principle*, often discussed in relation to the theory of evolution. In a very general formulation, it states that *the fact that human life has arisen in the universe was in some way premised at the very start.* The principle has been given several different specific formulations, but I only see reason to mention two secular versions. One that is true in a trivial sense, which I thus accept. And one that invokes a kind of Aristotelian goal-oriented cause that supposedly existed at the very beginning, which I reject based on what has been argued above.

The unreasonable formulation of the principle states that the fact that human life would arise in the universe due to the laws of nature and random events potentially existed as an Aristotelian goal-oriented cause at the very beginning. The trivially true formulation, on the other hand, states that the fact that human life would arise in the universe due to the laws of nature and random events must have existed as a general possibility at the very beginning.

Chapter 8

Free will from an evolutionary perspective

The medieval Aristotelians believed in a static worldview in the sense that, once God had created the world, no new physical-chemical elements or biological species arose. Aristotle himself did not believe that the world had a point of origin. Instead, he thought that it had always existed, and had always contained the same physical-chemical elements and species. For this reason, none of them struggled with the problem of free will and evolution that we encounter today. The free will of human beings was as old as the world itself.

The general Aristotelian notion of a form-matter hierarchy, however, can easily be grafted from a static worldview onto an evolutionary narrative. Parts of the existential hierarchy described above can without conceptual difficulty be assumed to also describe a temporal order of before-and-after. If an overlying level is dependent for its existence on an underlying level, the overlying level cannot arise prior to the underlying one. The opposite is conceptually possible, however. This creates a possibility to bring form-matter unities into a story where not only *synchronic* strong emergence may occur—as in the Aristotelian one—but also *diachronic* strong emergence. As previously mentioned (Ch. 5), diachronic and synchronic emergence may metaphorically touch in time.

Based on what I have claimed above about the theory of evolution (Ch. 2) and about emergent objects, properties, and relationships (Ch. 5), I shall now argue that the existing evolutionary narrative can be interpreted as a story about diachronic and synchronic emergence of form-matter unities and their properties and relationships. First came subatomic particles, atoms, and molecules, after which gas clouds, stars, and planets emerged. Organisms emerged too—at least on our planet. Initially, they were not conscious, but could reproduce. Where to draw the line and say 'this is an example of diachronic emergence' is not strictly relevant for the purposes of this book. All that is important for the acceptance of my version of the story is that, somewhere in it, strong emergence occurs among entirely material objects and conditions—thus also granting that diachronic strong emergence is not only typical of awareness phenomena. If one accepts, as I do, that new deterministic as well as indeterministic laws may arise during the course of evolution, the acceptance of a new type of strong emergence that within certain limits stands outside of the necessity-and/or-chance paradigm becomes less dramatic and strange.

Eventually, unities with awareness phenomena arose, too. In this case, it appears to me beyond all discussion that it must be a matter of strong emer-

gence (Ch. 5), both diachronic and synchronic. Such unities (for example: me and you, my reader) are constituted awareness-phenomena-and-body unities. Some such unities may also harbour a will, that is, a conscious goal-oriented cause. The question that this book seeks to answer—Can the human will from an evolutionary perspective sometimes have some degree of freedom?—can now finally be answered.

As proponents of free will have often remarked, no one has ever empirically proven that determinism is true on the level of human actions. Total determinism, if only for a small group of people, has never been shown to be a verified scientific truth. Determinism, it must be noted, is a philosophical position that presumes both that the most basic physical level can be described using deterministic theories, and that all ideas about physical strong emergence are false. Today, both assumptions appear to be wrong. But it is not sufficient to point out that some of the modern basic physical theories are of an indeterministic character. Physical indeterminism is not sufficient grounds for explaining the possibility of free will, since it does not address the problem of downward causation—that is, causation from the mental/spiritual to the bodily (Ch. 4)—nor does it typically distinguish between stochastic and non-stochastic spontaneity (Ch. 5).

I have argued that awareness phenomena can only exist as merged with a material substrate (Ch. 6). The previous chapter has provided concepts that now allow me to argue that such amalgamations result in Aristotelian form-matter unities that take awareness phenomena as their form. I am far from certain that the brain is a sufficient substrate, and that the rest of the body can in principle be done away with, but I am going to leave that up to future empirical investigations. For the sake of simplicity, however, I will henceforth write as though the brain is quite independent, which most people today seem to believe. The form-matter unity that I shall discuss, I can thus speak of using 'awareness-phenomena-brain-substrate unity' as a shorthand.

An awareness phenomenon and its brain substrate are not identical, but neither are they entirely separate. They are two aspects of the same unity. The relationship between the aspects is such that one type of awareness phenomenon can be realised by several different types of brain substrate (according to the principle of multiple realisability), but an awareness phenomenon cannot change without a corresponding change in the brain substrate (according to the principle of supervenience for changes). In other words, I consider it to be true that a change in an awareness phenomenon requires a change in some part of the brain substrate, despite the two not being identical and there not being any causal relationship between them. The unity is *constituted* in such a

way that the aspects relate to each other in the aforementioned way—that is, both through multiple realisability and supervenience for changes.

In short, I claim (i) that *between the unity's aspects* there is a *form-matter relationship*, and (ii) that *between the aspects and the entire unity* there is a *relation of constitution*. Unfortunately, it is not possible to address the problem of free will at its core without becoming a little scholastic—in the sense that several concepts are introduced where many philosophers take for granted that a single one is sufficient.

Form-matter unities, with their relationships, can theoretically occur both in deterministic and indeterministic lawlike relationships. I shall say a few words on this before addressing the will and its potential freedom.

When it comes to the treatment of what is classified as psychosomatic disorders, it is easy for psychologists and psychiatrists to apply the following line of reasoning: Let us prescribe psychotherapy as an efficient causality, which will—with a certain degree of probability—affect the patient's psyche, which in turn—through downward causation—may affect the brain and thus the somatic disturbances. This line of reasoning can be symbolised in the following way: therapist–psyche↔ patient–psyche ↓ patient–body. But as the downward causation (↓) cannot be defended philosophically, we should—from a philosophical perspective—use a different line of reasoning.

According to my Aristotelian views, we ought to reason like this: Let us prescribe psychotherapy, which will—with a certain degree of probability—*affect the patient's form-matter unity of mental phenomena and brain structure* and, via the brain, the rest of the body. Which could be symbolised as: therapist ↔ patient's psyche–body.

For practising psychologists and psychiatrists, the difference between the two formulations is probably irrelevant and most likely appears to be purely verbal. But if we are to make philosophical sense of the question of whether free will exists, the difference between the formulations becomes relevant. The question of whether free will exists is, in my opinion, no longer a question of whether the will as a *pure awareness phenomenon* can be free. That question rests on the erroneous assumption that there are pure awareness phenomena. The question that should be asked instead is: *Can will-awareness-brain unities sometimes strive toward a goal in a way that cannot be explained using the necessity-and/or-chance paradigm?*

My answer is: yes, goal-oriented will-awareness-brain unities can, within certain limits, *spontaneously* (in a non-stochastic sense) set a certain goal. This being so, the natural follow-up question is: on what grounds can this answer be justified?

 © ProtoSociology

Here follows the justification. Premise (i): it is absurd to completely deny the existence of free will, thus it must to a certain extent exist. Premise (ii): a will is a will-awareness-brain unity phenomenon. Conclusion: will-awareness-brain unities can sometimes, within certain limits, spontaneously set a goal.

This means that, by necessity, the spontaneity involves both the will-awareness and the brain substrate at the same time—which implies that spontaneity exists at the brain substrate level, as well. Can this belief also be defended? As mentioned (Ch. 5), neurologists sometimes speak of spontaneous neuronal firing. This, of course, proves nothing on its own. In their eyes, it probably only indicates a lack of neurological knowledge. But if we accept that the existence of free will cannot be completely denied, their use of the term 'spontaneity' gains an entirely new significance. It means that spontaneity in entirely material processes is not unthinkable. An acceptance in principle of non-stochastic brain substrate spontaneity is required for an acceptance of a belief in free will. And—as far as I can see—such an acceptance is more reasonable than accepting the performative contradictions that a complete denial of free will gives rise to.

The spontaneity in will-awareness-brain unities, the existence of which I have now argued for, is an objectively emergent property in the sense described earlier (Ch. 5). I have earlier claimed that deterministic natural laws can emerge on indeterministic ones, and indeterministic natural laws on deterministic ones. What I am now claiming is that *spontaneity can emerge on natural laws*. Spontaneity thus exists at a certain level and for this reason does not come into conflict with the deterministic and/or indeterministic laws on underlying levels.

However, I only consider myself to have proven that free will exists in the sense that the contents of a certain will can harbour *some small part* that is not subject to deterministic or indeterministic laws. It is thus a matter of a limited free will, and I have not attempted to discuss the extent of it or its exact limitations. The limitations are sure to vary depending on the person and the situation, but let me nevertheless—for the sake of concretisation—speculate a little.

Despite all the variations between people of the same culture, and despite all cultural variations, I believe it is possible to identify a set of general goal-oriented impulses that the egoistic pole inside all human beings attempts to satisfy. Here is a, in my eyes, possible list of goals that people in different situations can encounter as given internal goal-oriented impulses: pleasant feelings, food, sex, shelter, activity, and social recognition. Moreover, I believe that available empirical data clearly shows that not all of our goal-oriented impulses are subject to an egoistic calculation, and that we sometimes—independently of the egoistic consequences (and what benefits the survival of our genes)—can harbour conscious impulses to help others (show benevolence) and to worsen

the life of others (show malevolence). It seems to me that now and then human beings have impulses that are independent of their egoism—both benevolent and malevolent ones. The latter are relevant for the discussion of criminal law in Chapter 11.

Regardless of whether my anthropological speculations are correct or not, I find it almost incontestable that, in every situation, the degree of free will that a human being has operates within a given structure of goal-oriented impulses. What is more, nothing in my reasoning rules out that children and animals may have some degree of free will in certain perceptual situations. I have used argumentative discourse only to prove that it is absurd to completely deny the existence of free will, not to prove that it only exists in organisms that can participate in argumentative discourse. But it is reasonable to assume that the free part of the will is greater in beings with language of the kind humans have developed, and with the help of which we can easily imagine situations beyond what we perceive in the world.

I do not think it is possible from a philosophical perspective to say anything specific about the scope and limitations of free will. In this book, I have only sought to prove that the limitations created by the impulse structure of humans cannot always reduce our free will to zero. There is, to use a versatile expression, a vast difference between making a doorway smaller and walling it up completely.

© ProtoSociology

Chapter 9

Evolutionary biologists and free will

I have now presented the crucial parts of my defence of the views that the existence of free will cannot be completely denied and that it has arisen in the course of evolution. That is, not only does it exist, it also does not need to be explained with reference to something transcendent, such as God in traditional Judaism and Christianity. But what do the evolutionary biologists themselves have to say about free will?

My impression is that they prefer to avoid the matter entirely. But if they cannot, they often claim that it is a philosophical question, with the subtext that the difference between science and philosophy is such that philosophy can never question a scientific consensus. However, some of them expressly refer to so-called philosophical compatibilism—that is, the view that there is no contradiction between believing that everything happens because of necessity and/or chance and accepting that some degree of free will exists. As I have already said, I consider this to be a complete corruption of language (Ch. 4). Compatibilist thinkers take an existing everyday expression and give it a completely new content, whereupon they declare this to be its true content. From my perspective, both kinds of answers given by the evolutionary biologists mean that they consider the will in its traditional, incompatibilist sense not to be free. The evolutionary biologists appear to me to accept the necessity-and/or-chance paradigm as being universal.

The remainder of this chapter I shall dedicate to one evolutionary biologist in particular: Richard Dawkins (b. 1941). He is by far the world's most renowned champion of the biological theory of evolution—and for good reason. In the last fifty years, no one has defended the theory of evolution with such passion and superb pedagogical ability. So, what does he say about free will in particular? In most of his books, he avoids the question, but numerous interviewers and opponents have pressed him for an answer. Here is a summary of his most relevant comments:

With the book *The Selfish Gene*, Dawkins became an instant international celebrity. The first edition was published in 1976, with a second, expanded edition following in 1989. The latter contains two new concluding chapters and a significant number of new explanatory notes. The book has been translated into many languages and reprinted numerous times. I will begin with a longer quotation, which concludes the first edition. It is hard to read it without drawing the conclusion that Dawkins here ascribes to human beings some

freedom of will and action in the same way that I do, and that he discerns the same difference between argumentative discourse and other kinds of influence that I have emphasised in my discussion of performative contradictions (Ch. 3). This is how Dawkins concludes the first edition of his book (emphasis mine):

> One unique feature of man, which may or may not have evolved memically, is his capacity for conscious foresight. Selfish genes (and, if you allow the speculation of this chapter, memes too) have no foresight. […] It is possible that yet another unique quality of man is a capacity for genuine, disinterested, true altruism. I hope so, but I am *not going to argue* the case one way or the other, nor to speculate over its possible memic evolution. The point I am making now is that, even if we look at the dark side and assume that individual man is fundamentally selfish, our conscious foresight—our *capacity to simulate the future* in imagination—could save us from the worst selfish excesses of the blind replicators. We have at least the *mental equipment* to foster our long-term selfish interests […] We have the power to defy the selfish genes of our birth and, if necessary, the selfish memes of our indoctrination. We can even *discuss* ways of *deliberately cultivating* and nurturing pure, disinterested altruism—something that has no place in nature, something that has never existed before in the whole history of the world. We are built as gene machines and cultured as meme machines, but we have the power to *turn against* our creators. We, alone on earth, can *rebel against* the tyranny of the selfish replicators. * (*The Selfish Gene*, 1986, p. 200f)

Using ordinary semantics, I personally find it impossible to interpret this as saying that everything in the world happens due to necessity and/or chance. And I believe this would be true for most readers. In my opinion, the parts rendered in italics speak for themselves. This is not the words of someone who has fully embraced the opinion that free will does not exist. The text clearly opens up for the existence of people who—thanks to their 'mental equipment'—are able to freely 'simulate' future alternative possibilities, 'discuss' these, and potentially 'rebel against' the tyranny of selfish replicators. The entire text invites the acceptance of a certain freedom of will and action in human beings. By further pointing out that he does *not* plan to argue for or against the existence of a genuine, disinterested, true altruism, Dawkins implies that he otherwise considers himself to have argued for his opinions. Unfortunately, he does not see clearly how the conclusion, in a paradoxical way, appears to contradict what he has stated previously in the book. But others have been quick to point this out. This brings me to the footnote in the second edition, which the asterisk that concludes the paragraph quoted above refers to. In it, Dawkins writes:

 © ProtoSociology

> The optimistic tone of my conclusion has provided scepticism among critics who feel that it is inconsistent with the rest of the book. […] I *think* [they] accuse [me] of eating our cake and having it. Either we must be 'genetic determinists' or we believe in 'free will'; we cannot have it both ways. But […] it is only in the eyes of [… them] that we are 'genetic determinists'. What they don't understand (apparently, though it is hard to credit) is that it is perfectly possible to hold that our genes exert a statistical influence on human behaviour while at the same time believing that this influence can be modified, overridden or reversed by other influences. […] And no more is it dualist for me to advance rebelling 'against the tyranny of the selfish replicators'. We, that is our brains, are separate and independent enough of from our genes to rebel against them. As already noted, we do it in a small way every time we use contraception. There is no reason why we should not rebel in a large way, too. (Ibid., pp. 331 and 332)

I cannot understand this supposedly explanatory note to imply anything other than Dawkins changing his mind between the first and second edition and now submitting fully to the necessity-and/or-chance paradigm. To say that the brain as well as our genes can exert a statistical influence on human behaviour is normally not at all the same as saying that we, with the help of our 'capacity to simulate' and our 'mental equipment' can go against our genetic and cultural conditioning.

Let me, after offering these quotations from 1976 and 1989, refer to parts of a video recording from 2012 available on YouTube; at the time of writing, it has been viewed more than 100,000 times. In the recording, Dawkins debates with the theoretical physicist and cosmologist Lawrence Krauss in front of a big audience (See YouTube: Free Will – Lawrence Krauss and Richard Dawkins).

A member of the audience asks whether Dawkins believes in free will. At first, he facetiously replies: 'I have no choice.' As expected, his answer causes laughter to break out. I find this entirely natural. Dawkins' answer is paradoxical, and such repartee often gives rise to laughter. It is paradoxical because, in the middle of the public argumentative discourse that Dawkins—known to be very argumentative—is participating in, he suddenly positions himself outside of the discourse to say that he is forced to believe in free will. Yet even though Dawkins steps outside the given discourse through his answer, in some way he still remains inside it. When he accounts for what his brain forces him to think, his answer still appears almost like an argument because of the discourse he is embedded in. He comes very close to committing what I have called a discourse-specific performative contradiction (Ch. 3)—that is, arguing that argumentation on the topic of free will is impossible.

When the laughter has died down, however, Dawkins quickly turns serious.

He admits the question is one that he dreads and for which he does not have a well-prepared answer, referring instead to the work of philosopher Daniel Dennett (b. 1942). As Dennett is a compatibilist in the sense explained above (Ch. 4), we must also assume that Dawkins was a compatibilist at the time—and, as far as I can tell from subsequent interviews, still is. Thus, like most evolutionary biologists, he does not believe in free will in a traditional incompatibilist sense.

I think Dawkins should think it through one more time, and then return to the view expressed in the conclusion to the first edition of *The Selfish Gene*.

© ProtoSociology

Chapter 10

Freedom of action in perception

Based on a free will, we can sometimes also carry out the intended actions. These actions follow neither by necessity nor probability from the current moment and the laws of nature or social structures. If my argument is correct, it seems reasonable to believe that, at the very least, the acting subjects themselves should be able to perceive the moment of their free actions. Yet this view is often denied. The issue was formulated most powerfully by Ludwig Wittgenstein (1889–1951) in the early 1920s:

> Where in the world is a metaphysical subject [e.g., a free will and action] to be found?
> You will say that this is exactly like the case of the eye and the visual field. But really you do *not* see the eye.
> And nothing *in the visual field* allows you to infer that it is seen by an eye.
> (*Tractatus Logico-Philosophicus*, aphorism 5.633, [1921] 1963, p. 117)

This is not at all what I 'will say'. The relationship between a free action that one executes and perceives oneself to execute, on the one hand, and its cause, on the other—that is, free will—is not the same as that between the field of vision and the eye. Making sense of the issue requires two distinctions, both of which are missing in the *Tractatus* and in the writings of many famous philosophers. One is the distinction between the from-pole and the to-pole in awareness phenomena that I described in Chapter 6. The other is a distinction between foreground and background in perception.

A common view is that it is fairly unproblematic to correctly describe one's own perceptions and sensations. But just like the founder of phenomenology, Edmund Husserl (1859–1938), I consider this to be utterly wrong, and—in several philosophical cases—a fateful mistake. We cannot remain in what he calls the *natural attitude* ('natürliche Einstellung' in German) when doing phenomenology. Many things might be overlooked or incorrectly described if we erroneously believe that describing perceptions is no different from the task of describing perceived external objects and processes or internal bodily sensations. Describing everything in a perception or sensation as a whole is not the same as describing the object at the centre of the perception or sensation. Without a special approach, we will only be describing the object instead of the entire from–to structure that exists in awareness phenomena. This is true of conscious perceptions as well as conscious emotions, dreams, and sensations. Let me illustrate with an example.

The expression that someone 'cannot see the forest for the trees' is quite common. A person to which this expression applies may from a distance have seen a forest, but when asked 'What do you see?' stepped closer and came to the conclusion that what she was seeing was in fact a group of trees. However, the latter is not a description of the initial perception *as a perception*. Rather, it is a description of the parts of the object being perceived. Visual perceptions can typically—if a little vaguely—be divided into foreground and background. If something is at the centre of the field of vision and can be seen relatively clearly, it is in the foreground. But the field of vision often extends horizontally as well as vertically, taking in the background to what is in focus. The background is very much visible; seeing something out of the corner of one's eye is a completely adequate expression. But what exactly it is one sees in the background can be unclear. In fact, a correct description of a perceived background *should* describe it as unclear. A person who does not describe it so fails to note the uncertainties of perception, presumably because of a misdirected longing for clarity.

Now, it is often possible to subsequently focus on and clearly see what first appeared as background. It is no stranger than perceiving something as a forest from a distance, but as a cluster of trees from up close. However, this is not always the case.

When we perceive actions that we are executing, there is—in the background of these perceptions—a kind of awareness of a from-pole. At least this is true for me; and I would argue that those who do not agree have not considered and accepted the distinction between foreground and background that is characteristic of all perception. This remark also applies when the action only involves sitting around thinking or daydreaming. The problem is that even though we are able to consider the background in hindsight, we can never turn it into the foreground of an ongoing perception. In other words, *by necessity, the from-pole of an awareness phenomenon always only forms the background in perceptions and other awareness phenomena; by necessity, the from-pole only ever exists as background.*

I am not the first to note this fact, though we are a small crowd. I shall cite two action-focused philosophers whose writings quoted below were initially published in the 1930s. The first is one of the great thinkers of American pragmatism, George Herbert Mead (1863–1931). He makes a distinction between two aspects of the self: the 'I' and the 'me'. The 'I' represents a person's action-pole, the subject that in each moment truly acts, while the 'me' represents a person's view of him- or herself, a kind of object for the subject. Mead writes:

 © ProtoSociology

> If you ask, then, where directly in your own experience the 'I' comes in, the answer is that it comes in as a historical figure. It is what you were a second ago that is the 'I' of the 'me'. It is another 'me' that has to take that role. You cannot get the immediate response of the 'I' in the process. The 'I' is in a certain sense that which we do identify ourselves. The getting of it into experience constitutes one of the problems of most of our conscious experience; *it is not directly given in experience*. (*Mind, Self, and Society*, [1934] 1967, p. 174f; emphasis mine)

A little earlier in the same book, he writes:

> The 'I' of this moment is present in the 'me' of the next moment. There again *I cannot turn around quick enough to catch myself*. I become a 'me' in so far as I remember what I said. The 'I' can be given, however, this functional relationship. It is because of the 'I' that we say that we are never fully aware of what we are, that we surprise ourselves by our own action. It is as we act that we are aware of ourselves. It is in memory that the 'I' is constantly present in experience. (Ibid., p. 174; emphasis mine)

The other philosopher I wish to quote is a young Jean-Paul Sartre (1905–1980), later renowned as one of the frontmen of existentialism. He writes:

> Finally, what radically prevents the acquisition of real cognition of the ego is the very special way in which it is given to reflexive consciousness. *The ego never appears, in fact, except when one is not looking at it*. The reflective gaze must be fixed on the *Erlebnis*, insofar as it emanates from the state. Then, behind the state, at the horizon, the ego appears. It is, therefore, never seen except 'out of the corner of the eye'. As soon as I turn my gaze toward it and try to reach it without passing through the *Erlebnis* and the state, it vanishes. (*The Transcendence of the EGO*, [1936] 1966, p. 88; emphasis mine)

This relationship is not unique to visual perception. It also applies to proprioception—that is, non-visual perception of the relationship between one's body parts as well as the body's position in the surrounding space. This type of perception helps us keep our balance. Let me exemplify this with a dance in which, for a brief moment, the dancers close their eyes. Even here, with the field of vision deactivated, there is always—in the background—an awareness of some kind of from-pole that functions as a centre of movement, to which the movements relate.

The remarks made show how we consciously perceive our own actions and our freedom of action. Not as something that appears in the foreground—as an object or fact perceived through our visual, auditory, olfactory, gustatory, tactile, and proprioceptive capacities—but as something that can only take the form of a from-pole in the background to the objects and states of affairs

in the foreground. By necessity, one's own actions *as actions* only ever have a background existence.

This fact, of course, does not in itself prove that freedom of action exists. Just like in the case of external perception, a dream argument can be made. What we normally perceive as external objects that exist independently of ourselves can also appear in our dreams, where—per definition—they do not exist outside of our own consciousness. And in the same way, the freedom of action that we normally perceive in the background of our actions may also appear in our dreams. Despite actually lying in our bed, we believe ourselves to be carrying out a variety of actions. But disproving the everything-is-a-dream argument is not within the scope of this book. Unless you, my reader, believe that you live encapsulated in a dream world, and can more or less be considered a brain in a vat, there is no reason for you to believe that you could never perceive, in the background, your own freedom to act in the world.

© ProtoSociology

Chapter 11

Free will and morality

The ontological question of free will is often obscured by being tied all too quickly to the moral question of how people who have committed crimes should be punished if the will is thought of as free. Ontological intuitions in favour of free will are too quickly interwoven with moral-philosophical intuitions that undermine a belief in free will. More specifically, I suspect there is often a hidden rhetorical figure of thought of the following kind: *If you believe in the existence of free will, you will end up defending an inhumane legal system. And you wouldn't want that, would you?* No, I certainly would not; but that is not where I end up. This figure of thought is a fallacy. And that is what the present chapter is all about.

In order to settle the matter, one must conceptually separate two types of criminal punishment: *preventive* and *retributive*. Punishment such as regular imprisonment, imprisonment by leg iron, or fines can be justified in two ways, with reference to prevention or retribution, but also—very importantly—a combination of the two.

In purely preventive punishment, the penalty is seen solely as a means to prevent crimes of the same nature from being committed in the future. Normally, its goal is both to prevent the criminals themselves from committing the crime again (individual prevention) and to prevent others from doing so (general prevention). General prevention can work both through pure deterrence and through the penalty inspiring people to reflect and come to the conclusion that the criminal act is indeed reprehensible. When it comes to individual prevention, the criminal can also be sentenced to care and treatment.

Neither type of preventive punishment relies, for its conceptual meaningfulness, on lawmakers believing that criminals and the public have free will. Lawmakers can view the actions taken both by criminals and the general public as entirely determined by necessity and/or chance. The preventive punishment is then only thought of as adding another causal factor to the web of causes that are assumed to determine all human acts. (Only in some extreme cases do those who deny the existence of free will wish to stop punishing criminals; one such case is G. Caruso, *Rejecting Retributivism: Free Will, Punishment, Criminal Justice* (2021).)

In purely retributive punishment, the penalty is seen as a means to administer a particular kind of justice known as retributive justice. It is really a matter of *retribution*—a retaliation that is perceived as justified revenge for the criminal

act. If a person has committed a crime, they should suffer for it in the name of justice, that is, during a period of time they should have their quality of life reduced—the victim of the crime should be avenged. Retributive punishment assumes for its conceptual meaningfulness that criminals have some degree of free will and are responsible for their actions. The criminal is not merely a wind vane in the gusts of life. In other words: free will is a necessary condition for retributive punishment. From this fact, however, it does not follow that as soon as anyone with a free will has committed a crime, they ought to be punished retributively. Free will is not a sufficient condition for delivering retributive punishment. But believing in the existence of free will normally means accepting retributive punishment in some cases. And I belong to this camp.

Preventive punishment is conceptually compatible with the view that humans lack free will, but retributive punishment is not. This is so because it is only possible to avenge free actions. Those who completely deny the existence of free will must, as a consequence, believe that punishment can only be justified as a preventive measure. And philosophers who deny the existence of free will are in this sense usually consistent. (See, for example, the writings of the American philosopher Derk Pereboom.)

Personally, I embrace what one might call *penal pluralism*. Retributive punishment can, of course, be combined with deliberations concerning both individual and general prevention. In my opinion, both preventive and retributive considerations should be taken into account when devising laws and in their application. Here, I am only speaking of punishment sanctioned by a state's legal system. I am completely against private as well as clan justice, be it preventive or retributive. I think it is good if democratic states have a monopoly on violence. I think the vengeful impulses that naturally arise in many victims of crime often can benefit from an imposed pause.

Superficially, it might seem as though I always want stricter punishment than the free-will deniers. They only seek preventive punishment, while I want both preventive and retributive. Does it not follow, then, that I must always favour stricter punishment than those who deny the existence of free will? The answer is no, not at all! Let me explain—but first, a historical comment.

During the Middle Ages, people took no issue with free will, and medieval punishment was often very cruel. But it was not purely retributive. The intention of the cruelty was just as much—or probably even more—to deter others from committing the same crime.

The idea of preventive punishment contains no inherent limitation on how strict the punishment can be. Instead, it all depends on what the consequences of the punishment's enforcement are deemed to be. The idea of retributive

 © ProtoSociology

punishment, on the other hand, contains a limitation that is built into its very concept. Any retribution must be proportionate to the crime committed; sometimes, this principle is known as *lex talionis* or the law of retaliation. According to it, purely retributive punishment cannot be more serious than the crime. And if no crime has been committed, no punishment can be administered.

The proportionality of lex talionis can be understood in a few different ways, however. One is that the punishment should affect the criminal to the same degree as the crime has affected the victim—an eye for an eye and a tooth for a tooth, as the saying goes. But today, this kind of proportionality no longer appears reasonable. The scientific disciplines of psychology, sociology, neuropsychiatry, and neuropsychology (listed in their order of historical appearance) have proven that the human freedom of will and action is not as great as previously believed. From a layman's perspective, there are practically always mitigating circumstances to take into account. The criminal may have internalised other norms than those currently prevailing (immigrants), have a poor understanding of the consequences (children and young people), suffer from compulsive neuroses (kleptomaniacs and pyromaniacs), or lack the capacity for empathy (psychopaths). As I have stressed a number of times, I believe that our free will is always limited to one degree or another.

However, this science-based view on mitigating circumstances does not cancel out lex talionis—it only proves a need to rephrase it. A punishment should not be directly proportionate to how the victims have been affected by the crime. Rather, the relationship should be such that crimes where the victims were severely affected should be punished more severely. This, I think, is a very reasonable principle. And I believe the public legal consciousness would agree.

Toward the end of Chapter 8, I briefly dabbled in anthropological speculation, claiming that humans can harbour egoism-independent impulses of both benevolence and malevolence toward their fellows. It is the latter—that is, a will to worsen someone's life—that is relevant in criminal contexts. If all victims of crime were angels—that is, if they only had benevolent impulses and always willingly turned the other cheek after experiencing injustice—I would not argue that there is any need to discuss retributive justice. There is no general moral-philosophical duty to try to administer retributive justice. But there are very few angels in this world, if any. Most victims of crime feel a need to *get even* somehow, even if they are willing to have their revenge be executed by the legal system. Plenty of people have said 'I'll see you in court' to their perpetrator, and in newspapers I often read victims say that they are happy their perpetrator will go to prison.

As I put the finishing touches on this book, a strange trial is taking place in Germany. A 100-year-old man who used to work in one of the Nazi extermination camps is standing trial accused of being complicit in 3,518 murders. *Dagens Nyheter*, one of Sweden's biggest daily newspapers, writes the following about two of the children whose fathers were executed (7 Oct. 2021):

> Antoine Grumbach and Christoffel Heijer lost their fathers in the extermination camp Sachsenhausen, 30 kilometres north of Berlin. Now, almost 80 years later, they are taking part in the trial against a 100-year-old man who worked in the camp.
> 'I just want him to look at us and for him to feel guilt,' says Antoine Grumbach.

A humanist—as I consider myself to be—must be willing to take seriously all natural human needs. In case of conflict, one must try to find a good solution for what needs should be satisfied first or the most. And I believe that the need and will to seek revenge in certain situations are as natural as wanting to eat when feeling hungry. But I also believe that the need for revenge takes on more reasonable proportions if forced to cool down a little. This happens automatically if the state has the monopoly on violence; my discussion assumes that such a monopoly exists. In order to illustrate my thoughts on the reasonableness in sometimes allowing a certain degree of retribution, I will discuss the case of rape against women.

In the 21st century, rape has been discussed (at least in Sweden) in editorials, op-eds, and letters to the editor much more intensely than before. Women have made repeated demands for harsher penalties. This debate does not typically distinguish between preventive and retributive punishment, though my general impression is that these women are seeking harsher retributive punishment—that is, revenge. At first glance, this might appear strange; but at closer look it is entirely natural.

It may appear strange in part because the demand goes against an earlier (at least in Sweden) established trend of making penalties milder and milder, and in part because it conflicts with the gender stereotype of women as more likely than men to be kind and forgiving. But it becomes natural if you consider what is at the heart of all retribution—that is, the specific type of suffering experienced by the victim.

As mentioned, I believe there is a science-based mitigating view on crime. But I do not believe that to be the entire explanation for why the public legal consciousness was for a long time demanding milder and milder penalties. Just as important, I believe, is the development of the insurance system. This means that many victims of crime are not affected as severely as they would

 © ProtoSociology

be if insurance did not exist. Using the term 'suffering' in a broad sense, I shall concretise my thoughts with the following three statements: If a person is robbed, she will suffer less if she has insurance that covers theft. If a person's home is vandalised, she will suffer less if she has home insurance. If a person is abused, she will suffer less if she has health insurance.

Concisely put, I believe that the more extensive insurance systems we have, the milder the demands for punishment in the public legal consciousness. The Swedish Crime Victim Compensation and Support Authority's payments of criminal injury compensation are likely to have the same mitigating effect. I see this development as a positive one. But not all crimes are of such a nature that monetary compensation or free healthcare can lessen the victim's suffering. Rape is one such example. There is no insurance against rape. And I find it hard to see how such insurance could even be formulated if some people felt they wanted to try their hand at selling it.

If women who have been raped are denied the right to get even somehow, one could say that they are forced into revenge celibacy. If they live in a turn-the-other-cheek environment, which many religious women do, they are doubly affected—as many have pointed out before me. First, they are subjected to the rape itself, then to an informal condemnation of their inability to forgive their perpetrator. I see no other conclusion than that there ought to be a place for retributive considerations in criminal law.

Thus, a well-reasoned humanism should, in my opinion, allow criminal law to contain a reasonable amount of retributive punishment. Being a humanist means seeing human beings in all their complexity—that is, recognising all their types of goal-oriented impulses and their tiny sliver of goal-oriented free will.

I have pointed out that arguments for a complete lack of free will in human beings lead to performative contradictions. But this does not imply that thinkers and lawmakers who are against all forms of retributive punishment must be guilty of a performative contradiction when they say that all punishment should be preventive only. After all, they can view themselves and their colleagues as individuals with a certain degree of free will participating in rational discussion, while viewing criminals and the general public as entirely governed by necessity-and/or-chance factors. They can, in short, see themselves one way and regular folks another.

In the middle of the 19th century, this possibility was brought to light by a man who would later become famous for holding a certain deterministic opinion. His opinion was that the structure of capitalism is such that, by necessity, it will abolish itself—and his name was Karl Marx (1818–1883). But in

his early adult life, Marx argued that determinism cannot capture the whole truth of the human condition. In the third of his eleven so called *Theses on Feuerbach*—written in 1845 and published in 1888 by his close friend and main collaborator Friedrich Engels (1820–1895)—Marx makes the following insightful observation:

> The materialist [and deterministic] doctrine concerning the changing of circumstances and upbringing forgets that circumstances are changed by men and that it is essential to educate the educator himself. This doctrine must, therefore, divide society into two parts, one of which is superior to society.

However, I do not believe that Marx's aphorism is entirely true for the legal elite of today. This elite—I imagine optimistically—does not live completely cut off from the public legal consciousness. And the latter is not permeated by the opinion that free will is an illusion and punishment can thus only be justified preventively. At least in the Swedish criminal law of today—and I do not dare to be concrete with respect to other nations—there is also a degree of quiet retributive thinking, even if not explicitly expressed in terms of revenge and retribution. I find it in the concept of *penal value* ('straffvärde' in Swedish).

Since 1989, the term occurs in the Swedish Penal Code (BrB 29:1). The penal value determines how objectionable each crime is considered in and of itself, and does not refer to any arguments about prevention. For each crime, lawmakers set a penal value interval, which determines the minimum and maximum penalty for the crime in question. What the court then seeks to determine is the specific penal value in each individual case. The penal value is said to be a function of how harmful or dangerous the crime is. I believe that it largely reflects a demand for retributive justice. The penal value reflects the need in the public legal consciousness for getting even—and that need differs for different types of crime.

In this chapter, I hope to have clarified that a belief in free will does not, in my opinion, imply that the cruel punishments of the Middle Ages should be reinstated. However, I do think that retributive justice should be discussed openly and called by its true name.

 © ProtoSociology

Chapter 12

Concluding summary and hopes

There is an old mode of argumentation known as *reductio ad absurdum*. According to this principle, any opinion that leads to absurd conclusions must be dismissed. Applied to our discussion, it means the following: the opinion that free will does not exist comes with absurd consequences, and thus it must be dismissed. Completely denying the existence of free will implies performative contradictions. If there was no free will at all, argumentative disciplines such as philosophy and science, as well as everyday disputes about facts and norms, would be cognitively equated with such practices as playing music and the non-theological aspects of religion.

The degree of freedom of will and action may vary from one person to another, and from situation to situation. But something goes awry if this difference of degree is transformed into a binary difference of kind between having either complete freedom or none at all. As most people recognise that our will is never completely free, this thinking almost automatically leads to the fallacious view that free will is nothing but an illusion.

In my explanation for why the existence of free will must be postulated, and how we can accept that it has arisen during the course of evolution, I have made use of the following, perhaps somewhat unusual, concepts: *performative contradiction, emergence, spontaneity, intentionality, form-matter unities,* and *background existence*. In my experience, people who are interested in finding a worldview display a great deal of openness when they encounter concepts that are rare and new to them, if these are borrowed from physics or speculative physicists. Such concepts include *quantum entanglement* (the notion that quantum particles at any distance from each other in the universe can, momentarily, be intertwined) and *multiverse* (the notion that there is an infinite number of universes, of which ours is only one). It is my hope that the same openness will be extended to the concepts I have used to show that free will exists and has arisen during the course of evolution.

References

For several reasons, I have chosen not to encumber my exposition with footnotes and lots of references. A good introductory text to the problems discussed—with some classical references, of course—is Thomas Pink's *Free Will: A Very Short Introduction* (2004). Those seeking a more substantial and recent overview rich in references to the relevant literature can read the entry on 'free will' in today's most well-cited philosophical encyclopaedia, the regularly updated online resource *Stanford Encyclopedia of Philosophy* (https://plato.stanford.edu/).

Books that have helped me stick to my belief in free will over the years include: Karl Popper and John C. Eccles' *The Self and Its Brain* (1977), John Searle's *Rationality in Action* (2001), Jonathan E. Lowe's *Personal Agency* (2008), Helen Steward's *A Metaphysics for Freedom* (2012), Thomas Pink's *Self-Determination* (2016), Robert Lockie's *Free Will and Epistemology* (2018), Christian List's *Why Free Will is Real* (2019), and Jessica M. Wilson's *Metaphysical Emergence* (2021). As this list illustrates, it has become easier and easier for a philosopher like myself to hold to this opinion. The above-mentioned books—as well as a long list of philosophical journal articles—have also contributed to shaping my particular so to speak evolutionary-Aristotelian defence of free will, of course.

Some of the arguments presented herein hark back to my *Ontological Investigations: An Inquiry into the Categories of Nature, Man and Society* (1989, 2004). In that book, I defended an emergentism known as 'irreductive materialism' as well as the notion that, ontologically, there is nothing fundamentally unsound about the category of *spontaneity*. With the help of the latter, I also defined the concept of *agency*.

Other ideas in the present book I have previously discussed in a number of journal articles. Performative contradictions (Ch. 3) are discussed in 'Performatives and Antiperformatives' (2003), *Linguistics and Philosophy* 26: 661–702; my opinions on the structure of awareness phenomena (Ch. 6) and the perception of actions (Ch. 10) can be traced back to 'Triple Disjunctivism, Naïve Realism, and Anti-Representationalism' (2014), *Metaphysica* 15: 239–65; and form-matter unities (Ch. 7–8) I have discussed in, for example, 'Identity Puzzles and Supervenient Identities' (2006), *Metaphysica* 7: 7–33. Both my book and the essays are available to read on my website (www.ingvarjohansson.se).

© ProtoSociology