

Becoming a Self

The past, present, and future of selfhood

David L. Thompson

2023



This work is licensed under Creative Commons BY-NC-ND 4.0.

Table of Contents

| | |
|---------------------------------------|----|
| Preface..... | 4 |
| Introduction..... | 5 |
| Dualism and its Dilemmas..... | 6 |
| Problems of Mind..... | 9 |
| Alternative to Dualism..... | 10 |
| Outline..... | 11 |
| Part I: Origins..... | 13 |
| Chapter 1..... | 13 |
| Simple Life..... | 13 |
| Unity and Correctness..... | 13 |
| Amoeba..... | 15 |
| Function..... | 17 |
| Chapter 2..... | 23 |
| Evolution..... | 23 |
| What is Detected?..... | 23 |
| Aboutness..... | 25 |
| Co-evolution..... | 31 |
| Systems..... | 33 |
| Conclusion..... | 34 |
| Chapter 3..... | 35 |
| Culture..... | 35 |
| Working Together..... | 37 |
| Cooperation within Tribal Groups..... | 38 |
| Contract Theories..... | 43 |
| Group Identity..... | 45 |
| Imitation..... | 48 |
| Emotional contagion..... | 49 |
| Music and Dance..... | 51 |
| Religion..... | 53 |
| The Context of Group Norms..... | 55 |
| Normative status..... | 57 |
| Cultural World/Umwelt..... | 59 |
| Conclusion..... | 62 |
| Chapter 4..... | 64 |
| Language..... | 64 |
| Defining Language..... | 64 |
| Emotion and Music..... | 66 |
| Linguistic Meaning..... | 68 |
| Adamic theory..... | 69 |
| Mentalism..... | 70 |
| Contextual meaning..... | 71 |
| Truth and Lies..... | 73 |
| Syntax..... | 74 |
| Evolutionary explanation..... | 77 |

| | |
|--|-----|
| Talking to Oneself..... | 78 |
| Neural Processes..... | 80 |
| Linguistic Umwelt..... | 82 |
| Conclusion..... | 83 |
| Part II: Selfhood..... | 85 |
| Chapter 5..... | 85 |
| Self-identity..... | 85 |
| The Problem of Self-identity..... | 86 |
| Inadequate Solutions..... | 89 |
| Memory..... | 89 |
| Narrative..... | 90 |
| Strengths and difficulties of the narrative account..... | 92 |
| Chapter 6..... | 96 |
| Normative Selfhood..... | 96 |
| Binding..... | 97 |
| Character..... | 98 |
| Freedom..... | 104 |
| Reputation..... | 108 |
| Summary..... | 110 |
| Chapter 7..... | 112 |
| Interiority and Belief..... | 112 |
| What More?..... | 113 |
| Belief..... | 115 |
| Animal beliefs..... | 115 |
| Self-attribution..... | 117 |
| World/Umwelt..... | 119 |
| Experience..... | 122 |
| Reflection..... | 125 |
| Chapter 8..... | 128 |
| Are Selves Real?..... | 128 |
| Reality..... | 128 |
| What is a Context?..... | 131 |
| Systems..... | 134 |
| Norms..... | 137 |
| Status of my Account..... | 143 |
| Chapter 9..... | 146 |
| Selfhood: Past and Future..... | 146 |
| Is there a “True Self”?..... | 147 |
| When did Selfhood come about?..... | 148 |
| Will Selfhood Survive?..... | 149 |
| How might Selfhood Survive?..... | 151 |

Preface

What it is to be a human being, what mind and consciousness mean, and what makes something a self, are questions that have intrigued me for decades. This book explores these issues from the perspectives of evolution, of biology and of individuals. It concludes that the life of a self is based on norms—as is any other kind of life. It is not the physical matter that makes up the body nor the neural processes that occur in the brain that make us who we are. Rather it is the biological, cultural and narrative contexts within which these take place that explain what it is to become and to be a self.

In developing these ideas I have been greatly helped by discussions with friends, colleagues and students, to all of whom I am immensely grateful. John Cook, Evan Simpson, Walter Okshevsky, Chris Dennis, Suma Rajiva, Dan Stewart, and other participants in reading groups have inspired me over the years. In particular, I am very thankful to Evan Simpson, Francis Thompson and John Scott for reading earlier drafts of this book and offering valuable feedback. I am particularly grateful to Ariane Wilson for the artwork on the cover.

Introduction

*Who am I? That is the fundamental question for all humankind.
Manshi Kiyozawa (Pure Land Buddhism, Higashu Honganji, Japan.)*

You are a self. I am a self. These claims are incontestable: to contest them, we must assume them to be true. There can be no debate between non-selves. If we were not selves, you would not be reading this book, and I would not be writing it.

On the other hand, for thousands of years there have been mysterious injunctions: “Be thyself;” “Become who you are;” and so on. How could I not be who I am? The Delphic oracle told Socrates, “Know thyself.” Could I not know who I am? Shakespeare's Polonius advised, “To thine own self be true.” Is there a *true self* to which I might be false? Contemporary psychologists sometimes claim to help people to overcome their inauthentic self, but what exactly is authenticity?

Why is being myself important? What is it about my selfhood that I consider valuable? I want to be free to make up my own mind and choose my own actions. I want to have my own values, adopt my own projects, and take responsibility for my beliefs, my behaviours, and my relationships. I want to view the world my way. I want my thoughts and emotions, my feelings and obligations, to be unified into one coherent personality. I want to preserve my identity, as opposed to being fragmented like Dr. Jekyll and Mr. Hyde. I want to be valuable: valuable to others, to society, to history, but, above all, valuable to myself. I want my life to be meaningful.

Is my belief in my own value not hubris? In the order of nature, my self is an insignificant speck in an almost infinite cosmos. Does science not show that freedom is an illusion? If there is no freedom, then I have no responsibility for any of my actions, including the action of “becoming myself.” So becoming the self that I want to be seems to be impossible. Is the matter not out of my hands?

Behind all these questions is the central issue: what does it mean to be a self in the first place? The aim of this book is to examine these and related questions. In the coming

chapters I will lay out a view of the nature and origin of selfhood that will offer answers to some of these questions.

Dualism and its Dilemmas

Before I present my own position, however, we need to examine the most important challenge to it. This challenge is a way of thinking that has dominated Western thought for some centuries: causal mechanism. Causal explanation has greatly influenced our contemporary ways of thinking about nature and about selfhood. While I will present it as sympathetically as I can, the idea that the only way to explain things is by understanding their causes is an approach that I *reject*. In particular this approach makes it impossible to understand what it is to be a self. The rest of the book will present my alternative account of understanding, of nature and of selfhood.

“Modern Science”—the approach of people like Bruno, Copernicus, Galileo, Boyle, and Newton—saw itself in the sixteenth and seventeenth centuries as rejecting occult Medieval ways of explaining nature. It opposed anthropomorphic and theological modes of understanding which, it believed (not always fairly) relied on the intentions of God to explain the natural world. “Modern science” rejected explanations such as: the sun rises in order to heat the earth; diseases are punishments for sin; good harvests are a gift from God. Since “the purposes of God are inscrutable,” modern science set out to reject speculation about the purposes and values of events. Instead, science should discover the natural causes that explain the world.

While Ancient and Medieval thinkers had referred to various kinds of causes, seventeenth-century science focused on what had been called “efficient” causes. A cause in this sense is a natural object or event which made the “effect” happen. Every event has a cause. The simplest image of a cause is direct impact. Imagine billiard balls, for example. When one ball collides with another we can just “see” how the movement of the first ball causes, that is, forces, the impacted ball to move. Each ball, as a physical object, is without any purpose of its own, so its movements are always effects of external causes.

Causation is not a one-off phenomenon; causes obey invariant laws. Unlike the whims of gods, spirits or leprechauns, natural causes are predictable in that they obey “laws

of nature.” Appeal to the influence of conscious beings is seen as anthropomorphic and unscientific. Worse, such appeals to wilfulness do not allow for prediction or control, which is one of the motivations for modern science. Causal laws are universal and apply in the same way everywhere in the physical world.

The aim of science is therefore to discover universal natural laws. It is the founding principle of “Modern Science” that the only way to explain any natural event is by pointing to the cause which made the event necessary: it was pre-determined and could not have happened otherwise. I will use the term *mechanism* to refer to this project of explaining the world by appeal to universal causal laws. I will use *mechanist* to refer to anyone who think this is the only way to understand the world. This seventeenth-century approach to understanding – “causal mechanism” – still haunts our thinking today.

Note that the term “cause” is often used in a general way to refer to *anything* that explains an event. For instance, “It was political ideology that caused me to lose my job.” I will *not* be using it in this casual, common-sense way. Throughout this book I will be using the term exclusively in the narrow, restrictive mechanistic sense: a “cause” is that which brings about an effect in accordance with a universal natural law.

The universality of mechanistic laws means that every effect is the law-governed outcome of previous causes which determine that this effect necessarily had to occur; it could not have been otherwise. Eclipses, for example, are predictable because planets obey natural laws. Laplace hypothesized that a perfectly intelligent mind, knowing all the natural, mechanistic laws, and given the exact position and velocity of every particle in the universe at a given moment, could predict every future event (and explain every past event). Everything has to happen the way it does happen; it has been ordained from all eternity. Mechanism implies determinism.

The doctrine of determinism, however, presents us with a problem. It applies to everything in the natural world, and since humans are part of the natural world, all human events are as predictable as eclipses. Humans therefore have no freedom: whatever they do has been written in the stars since there have been stars. Obviously this doctrine has moral and political implications, and, in particular, it implies that the kind of free, self-determining self that we want to be is not a possibility.

Apart from the moral issues, however, determinism poses a dilemma for “Modern Science.” Humans, including scientists, as citizens of nature, are causally determined: what they write or say depends on forces beyond their control. A parrot's training might cause it to say “ $2+2=4$ ” but it could just as well be trained to say “ $2+2=5$.” What it says is not guided by truth. If mathematicians’ brains are governed by natural causes, they do not have the freedom to pursue truth; they too might be caused to claim that $2+2=5$. If what scientists say is determined by natural causes, then we should no more trust what they say to be true than we can trust the squawking of a trained parrot. Scientists are like machines whose output has been determined from the beginning of time. Worse still, they cannot trust their own thoughts, since these too are controlled by causes. Causes are blind to truth, so determinism undermines the very project of science.

It is not just science that is undermined; in a deterministic world any search for truth is impossible. Even the claim (or belief) “everything is determined” is no more trustworthy than the parrot's “ $2+2=5$ ” or “ $2+2=4$.” The doctrine of determinism is self-defeating: if it is true, no one could ever know it to be true. Claiming that truth is unattainable cannot be a solution, because if the claim is true, then, since it is itself a claim, its truth too is unattainable.

This dilemma was already understood by many proponents of “modern science” in the sixteenth and seventeenth centuries. If humans are part of nature and all events in nature are causally determined then science is impossible. The only way out of the dilemma is to claim that humans are not part of nature. At least human thinking has to escape determinism by mechanistic causes. This is the “dualist” solution to the dilemma adopted by many in the seventeenth century.

While there are many kinds of dualists, one of the most influential is René Descartes, so I will be using him as the representative of dualism in general. Writing in 1641, Descartes claims that one belief is certain: “I think, therefore I am,” his famous *Cogito ergo sum*. No cause, not even a Divine one, could make him believe that this thought is false. To think it is false, he must be thinking and therefore existing and so it cannot be false. He calls this existing thing a “mind.” One's body is a part of nature – a material, physical thing governed by causes – but one's mind escapes determinism because it is not part of the order of nature: it

is non-physical, immaterial, that is, “spiritual.” Each human is made up of two things – a body and a mind. This is the doctrine of dualism, an option that many cling to even to this day.

Descartes' dualism has two crucial strategic advantages for early modern science. First, it posits a complete system of mechanical nature governed by discoverable, deterministic laws. This is the nature that natural science sets out to explain without appeal to occult principles, or to human or divine purposes. Secondly, dualism allows the existence of scientists who are themselves autonomous, beyond the reach of such mechanistic determinism, and who possess the freedom necessary for grasping truth.

The dualist strategy for rescuing our freedom, our capacity for scientific knowledge and our human dignity is to place the mind in an immaterial realm beyond the reach of mechanistic, deterministic causal laws. This became the default position for the following few centuries of Western thought. Mind and matter, thought and causality, consciousness and physical reality were seen as irreconcilable dichotomies. Frequently they still are in popular opinion and even in some scientific approaches.

Problems of Mind

The dualist solution, however, has gradually become less plausible over subsequent centuries for a number of reasons. The brain was poorly understood in the seventeenth century. Since then, cognitive neuroscience has revealed the interdependence of the mind and the brain. Studies of psychiatric pathologies have shown that brain processes, which are material and causal, affect mental functioning.

Similarly, psychology has taught us that the mind arises from a long development through infancy to adulthood. It is hard for traditional dualists to conceive of a spiritual mind as coming about gradually. Descartes' *Cogito*, for instance, is an all-or-nothing phenomenon: there cannot be half a mind. He thought of the mind—or “soul”—as created by God in one instant at birth (or maybe conception, or at animation.) Developmental psychology challenges this notion of an innate mind that is fully present from the beginning of a human life. It is not simply that during life the mind comes to contain new ideas; the mind itself comes into being only gradually.

Most crucially for my account, however, is the challenge of evolution. Darwin, in the nineteenth century, discovered that *Homo sapiens*, like all other species, arose in an evolutionary process from earlier biological species. We now know that there were many other hominin species over the last three million years. We can no longer accept human exceptionalism. Humans are not unique creatures with spiritual minds; they are a species of animal; they are biological organisms. Darwinism has raised a question that Descartes couldn't even conceive of: how did human life and selfhood evolve from that of earlier primates? In particular, through what unique evolutionary history did human selfhood come into existence?

As a result of these developments, the claim that our human mode of being can be explained by an immaterial mind that floats free from neural, hormonal and other bodily processes is no longer scientifically credible. Since Descartes took some inspiration from the religious idea of an immortal soul, many scientists and other thinkers have simply rejected the idea of an immaterial mind as a prescientific remnant of religious doctrine.

I think such a simple rejection is a mistake. Whatever its inspiration, dualism is a rational attempt to solve the perplexities of a particular understanding of nature – causal mechanism. Simply rejecting the notion of a non-material mind while not challenging the underlying mechanistic conception of reality to which it is a response leaves the relationship of the scientific world to selfhood, freedom, and responsibility unresolved. The persistence of debate about freedom and determinism into the twenty-first century, even between people who reject spiritual minds and religious souls, indicates that the problem is not so easily solved. The doctrine of a non-natural mind and the mechanistic understanding of nature go hand in hand. Each one feeds on the other. I agree that we should drop the concept of an immaterial mind, but in order to do so, we must also abandon the way of thinking about nature that leads inexorably to that concept.

Alternative to Dualism

Traditional dualism is inadequate for understanding the life of the self, or any kind of life for that matter. The account of selfhood I present in this book requires that we move beyond the dualist approach altogether. I will therefore drop the notion of a spiritual *mind* and

speak instead about the *self*. But we must also abandon the exclusively causal, mechanistic notion of nature. My positive alternative will be laid out as the book unfolds.

My alternative will present a way of looking at organisms and selves in terms of norms and systems rather than physical facts and causes. I will show that evolution results in unified, self-directing entities that are to be understood in terms of norms rather than mechanistic determinism. My account will offer a contextual understanding of the world, one based on systemic relationships. Rather than conceiving of the mind as an isolated thing, we will see that the self is more like a node in a historical and social system. Underlying my exposition throughout will be a claim about the nature of *understanding*, about what it means to understand organisms, selves, or actions. Such understanding has an essentially temporal aspect: evolutionary history for organisms and narrative history for individual selves.

These dense statements are undoubtedly impenetrable at this point. The goal of this book is to unpack them, explain their meanings and offer convincing reasons for them.

Outline

The book is in two parts.

Part I: Origins discusses the evolutionary developments that prepared the way for selfhood. One goal of this part is to introduce key concepts that will be important for later parts: unity, function, context, norms, worlds, and systems.

Chapter 1: Simple Life investigates the nature of early, simple organic life. How do single-cell organisms unify themselves and gain some primitive autonomy?

Chapter 2: Evolution discusses the development of biological functions in evolutionary history and the distinction between normal and abnormal behaviour in the life of organisms. Darwinian “survival of the fittest individuals” needs to be modified in light of the systematic nature of organisms and their *Umwelten*.

Chapter 3: Culture moves from the genetic transmission of ways of life to transmission by learning, that is, by tradition. Early hominins rely on group cooperation as their evolutionary strategy. Cooperation depends on emotional identity with the tribe. Group identity is inculcated by common bodily behaviours, shared tools and survival

techniques, music, dance, and ritual (early religion). Norms for different cultural roles are enforced in various ways.

Chapter 4: Language explores how communication based on cultural norms increases the fitness of the group. Language, whether gestural or vocal, facilitates higher levels of cooperation both within and between generations. Later, syntax increases the flexibility and richness of communication.

Part II: Selfhood focuses directly on the nature of the self and defines it as a normative entity within the context of a system.

Chapter 5: Self-identity examines what it is for a self to be a unified entity. How can one be the same self over a lifetime? Memory is not enough. The narrative that one creates offers a more plausible account of self-identity, but ultimately proves inadequate.

Chapter 6: Normative Selfhood. Individuals become selves by adopting certain social norms and creating their own personal norms. These norms form the *character* of a self. *Freedom* is acting in accord with character. *Reputation* is the necessary social parallel of character.

Chapter 7: Belief claims that a belief is not an object in an interior mind, but is a normative state about the world: it is how a self refers to its objects in its world. Beliefs are not private, but accessible to others. A self can form beliefs about its own beliefs—reflection. A self can believe that it has a stream of beliefs about the world. The Myth of Interiority misinterprets this stream of beliefs as a private stream of consciousness.

Chapter 8: The Reality of Selves rejects the view that selfhood is an illusion by examining and defining the concepts of reality, context, system, and norm.

Chapter 9: Selfhood Past and Future summarizes my account of selfhood as an entity defined by the norms of the self-world system. It examines the idea of a True Self. It speculates on the future of selfhood and discusses whether future selves could be digital.

Part I: Origins

Chapter 1

Simple Life

Selfhood does not spring full-blown from the head of some mythical god. It comes into being as the product of a long evolutionary history and cannot be understood without examining this development. This chapter offers an account of the early biological history on which selfhood is founded. It presents an interpretation of the nature of simple organisms that will allow us to explore some crucial prerequisite concepts needed to make sense of the self: unity, homeostasis, normalcy, functionality, and the role of evolutionary time.

Unity and Correctness

The origin of biological life on earth remains mysterious, despite recent scientific advances. So let me offer a “just-so” story, heavy on speculation.

A simple cell needs three components: an energy metabolism; an information/control mechanism; and a membrane to isolate these two from the wider environment. I have little to say about the first two, so let me focus on the membrane. Pohorille and Deamer offer an hypothesis about how impermeable membranes could have self-assembled on the early earth 3.5 to 4 billion years ago. Amphiphilic molecules—molecules (like soap, for instance) with parts that interact with water and other parts that interact with lipids (oils and fats)—could have been formed by inorganic chemical processes. Amphiphiles tend to congregate on interfaces where water and lipids meet. They orient with one end towards the water and the other end towards the lipid. These oriented molecules bind together and then tend to form a vesicle—a sphere with a hollow for contents. I will call this vesicle a “protocell.”¹

A real cell would need to contain DNA, RNA, and other informational and metabolic elements. Without a membrane, however, these elements would be open to disruptive

interferences from processes in the wider environment. The membrane isolates an internal environment that is partly independent of outside factors.

The membrane allows the protocell to maintain various equilibria. For instance, assume that in a single protocell there is a sodium-water feedback loop that maintains saline concentration within certain limits. Some chemical process in the cell responds to a low sodium concentration by triggering a mechanism that opens receptors in the membrane to accept more sodium from the external environment. If there is too much sodium, the cell opens receptors for water to dilute it. In any cell there are many such stabilizing feedback loops – homeostases.²

Homeostasis enables a cell to exist as an enduring unity. Its persistence depends on various parameters, the concentration of sodium for example, being held within certain limits. There is a “correct” level that a cell needs to maintain. If the level become too incorrect, the homeostatic equilibrium disappears and the cell dies. In this way, homeostasis sets a standard for saline concentration (and other metabolic levels). That is, homeostasis differentiates between the range of levels that are correct and other ranges that are incorrect.

“Correct” in this sense just refers to the range that permits the cell to survive as an enduring unity. The processes involved in the maintenance of the cell's life and in the cell's death are governed by universal causal laws. Whether a cell lives or dies is biologically important, but it is irrelevant from the viewpoint of physics or chemistry. The laws of chemistry apply equally to a living cell or a dead one. Chemistry alone cannot define a “correct” sodium concentration: any level of sodium whatsoever can be explained chemically. The chemical laws governing the feedback systems are indifferent to whether the cell survives or not. In the absence of cells, there is no basis for distinguishing a correct from an incorrect concentration. Mechanistic scientific laws are value-free, so chemical causality is blind to the distinction between correct and incorrect concentrations.

The standard for correct concentration is not based on some mystical Vital Force, nor is it miraculous or magical. The standard is not preordained from all eternity; it is not written in the stars. Only when cells came into being as homeostatic unities some billions of years ago did a particular concentration get established as “correct.” This correlation between unity and

standards of correctness will be a recurring theme in my exposition, especially later when I discuss selfhood.

Amoeba

When cells initially developed, they depended primarily on passive chance encounters to absorb the nutrients they needed to survive. As the ability to move develops over the couple of billion years of evolution, some organisms became more active. So let us now imagine a more sophisticated organism, one with discrete organs and the ability to move. I will call this hypothetical organism an “amoeba.” (My apologies to real biological amoebae – my amoeba is a theoretical, imaginary construct! Think of it as a thought experiment or an intuition pump.) My amoeba gets its energy by absorbing other organisms, say, bacteria, plankton, and diatoms. It detects these nutrients by having receptors that react to different chemicals—scents—given off by these various energy sources.

While initially each receptor might have been connected to its own unique response process, a more efficient, and so evolutionarily adaptive, organization would connect all the nutrient receptors to a single pursuit response. I will call such a unified organ a “detector.” Similarly, the amoeba unifies receptors for noxious substances, say iron, ammonia or alcohol, into the one detector whose output is avoidance behaviour. Assuming that these “rewirings” of the internal organization give the amoeba an evolutionary advantage, then, after many generations, surviving amoebas will have such unified detectors linked to their appropriate response mechanisms. Note that what unifies the many receptors into the one organ—a “detector”—is that they provoke equivalent responses.

In this, I am assuming the simplest model of Darwinian evolution: adaptationism. In any generational system in which there are entities with descendent that inherit characteristics, any mutation in these characteristics that enhances reproductive capacity is likely, if there are limited resources, to crowd out descendants of those without the mutation. The species that is the fittest for its environment is the one that survives. (We will see later, in chapter 2, that this is not the only or best way to understand evolution.)

The arrival of biological structures with discrete organs is the emergence of a new unity. These organs are incorporated into the organism so that the amoeba acts as a unified

whole. In facing its environment as a unit, the organism as a whole is no longer simply the plaything of brute physical causes. The organism regulates its own response. Let me unpack this claim.

In the case of purely physical, non-biological processes, events are governed by physical causes. For instance, given the appropriate energy conditions, when molecules of carbon bump into molecules of hydrogen, methane is formed. The impact-event is the cause of the existence of methane as the effect. Let me call this “direct causation.” With unified organisms, in contrast, what behaviour a stimulus will provoke is insulated from the direct effect of the stimulus as cause. Beyond direct physical causation, with its regular, universally predictable effects, the hypothetical amoeba adds a new dimension. Its internal organization redirects the input in such a way that the final output is modified by what the organism needs. The organism is to a certain extent self-directing, self-determining. When an amoeba's receptor for the scent of plankton encounters a chemical that stimulates it, what happens next no longer depends just on the chemical, but on how its reception is processed within the organism. Similarly to the way a cell's membrane homeostatically isolates its contents from the processes outside the membrane, the amoeba's internal organization give the organism a certain limited independence from its external environment.³

That's what it is for any organism to be a unity as opposed to a random heap of parts. Causes may impinge on an organism, but it is not the part of the organism that the cause impinges on that reacts. Rather, the input is delayed; it is deferred and rerouted through the organism's structure before the organism reacts as a whole. Instead of the direct physical cause leading ineluctably to its inevitable effect, the organism places its stamp on the process and determines what the response shall be on the basis of its own structure. External stimuli are no longer permitted to automatically unleash their effects in accordance with a universal law: they are redirected by the internal organization of the organism. External heat, for example, causes a rock's temperature to increase, but external heat applied to a mammal is moderated by the mammal's internal homeostatic organization, so that the organism's temperature remains stable. If I drop a stone, gravity will cause it to fall downwards, but if I drop a bird it may dive or soar depending on factors internal to itself.

The movement of a dust particle on a stream of water is the direct effect of the force of the stream. But in the case of my amoeba, the force of a stream of water does not necessarily carry the organism along with it. If the amoeba needs to move in a different direction to engulf its food, the final movement cannot be predicted solely on the basis of the water flow. The stimulus is assigned a role by the organism as a whole, so that the output of the stimulus needs to be understood as much on the basis of the internal structure of the organism as by the nature of the input. In this very limited way, the amoeba, as a unified whole, has a certain independence from the direct external causes; it is partly self-directed. In a very weak sense, it can be said to “act,” to be an “agent.”

Function

As a unified whole, my hypothetical amoeba is a unity. But what kind of unity? It is a functional unity. Its parts may in fact be in a close spatial arrangement within a boundary, but it is not this fact that constitutes the unity of the organism. It is a unity because its components—its detectors, response mechanisms and so on—are *organs*. The activity of each organ, whether provoked by external objects or by something internal, contributes functionally to the life of the organism as a whole. The various receptors are unified into one food detector not because of their chemical similarities—they may not be similar—but because they serve the same function: food detection.

What do I mean by “function?” The word has many uses in English. There are mathematical and computational functions. A university convocation is a function. Chairing a committee is a function. These are not the uses of the word that I want.

But even when we close in on the sense I need, there is a cluster of equivocal meanings. Even if, during cardiac arrest, a human heart stops “functioning,” that is, stops pumping blood, it remains the case that the “function” of the heart is to pump blood. Similarly, the function of my house alarm is to protect me from break-ins. I’ve had no break-ins, so it has never functioned. Except for one annoying night when the alarm went off due to a malfunction, it has carried out its function perfectly, that is, by not functioning! The word is confusingly used in two different ways in these sentences. In one sense it refers to the actual operating of the mechanism, the active sounding of the alarm or the current beating of the heart. In the other sense it refers to the role that the house alarm has—to protect me from

burglars—or to the role the heart has in the body. In analyzing my toy amoeba scenario, and throughout this book, I will be using the word exclusively in this second sense—the role that something plays.

In this second sense, there could be more than one physical mechanism that could carry out the function. Functions have a certain kind of independence from the stuff they are made up of. My old house alarm was electrical; my new one is digital. Artificial hearts are not made of biological cells, but they are hearts nonetheless. It is not the underlying physical mechanism by which the amoeba's various receptors get unified into the one food detector that defines it as a detector. If amoebas had evolved on some other planet, it is conceivable that their food detectors might have relied on silicon-based chemical processes rather than carbon-based. What defines any organ in an organism is the role it plays in enabling the life of the organism. This is true of the saline receptors on a bacteria's membrane, of an amoeba's food detectors, or of a mammal's heart. It is not the material composition that determines the function of an organ; rather it is the functional role of the organ that determines which material mechanisms will enable the organ to carry out its function.

A function, as a role, is therefore in one sense general. A role is like an empty container that could be filled in various ways. The role of food-for-amoebas could be filled in one case by this diatom, or in another case by a different diatom, or perhaps by a bacterium. In principle, if the vagaries of evolution had been different, it would have been possible for some alternative entities to have become food-for-amoebas, even though, as it happened, this in fact did not turn out to be the case. In the future, amoebas might come to pursue different nutrients which would then fill the role of food-for-amoebas. To put it another way, a diatom cannot be food all by itself. It is not the intrinsic properties of a diatom that makes it food (though it must have some relevant properties). It is the relationship between the diatom and the amoeba with its food detectors that defines the diatom as food-for-amoebas.

This is a *contextual* relationship in the sense that it cannot be explained just by the physical processes involved. Physical laws are universal, while biological relationships only emerge within the chance biological history of the earth during its first few billion years. If history had been different, as it may have been on Mars, then an amoeba-to-food relation would not have come into being. Contextual relations do not violate causal natural laws;

indeed the evolution of a new relationship depends on the ability of organisms to mobilize causal processes for their own needs. Since the appearance of such biological relations depended on the particular historical conditions on early Earth, however, they could not be explained by, nor predicted by, physical laws alone.

The relationship between the amoeba and its food is not a causal one. In the strict sense of mechanistic causation, the food-for-amoeba does not cause the amoeba's pursuit. There is no universal law to that effect, and indeed the amoeba may on occasion not pursue a particular particle of food. It might be satiated, for instance. Nor is the relation a statistical one. It is not the frequency of pursuit that enables us to understand why an amoeba gives pursuit. The relation is to be understood functionally: it is the function of the food detectors to detect food and the function of the food to-be-detected, and to-be-pursued. A functional understanding is neither a causal nor a statistical one.

It is not only the detectors or other organs of the amoeba that must be defined functionally; the unity of the whole amoeba is a functional unity.

It is not simply that, as a matter of fact, some chemical processes in the amoeba are activated in the presence of food: it is the function of the food detectors to detect food. The functionality of the food detectors makes sense only in the context of the whole organism. If one excised the detectors from the amoeba and set them up in isolation in a laboratory dish, the chemical mechanisms would still respond to the same stimuli; the factual processes would remain unchanged. But they would serve no function. Without the context of the organism, they would no longer be organs. Functionality is not intrinsic to the mechanisms themselves: it depends on the roles they play in their context. A function is a relational property that vanishes if these relations disappear. It is only in the context of the whole organism that the food detector is what it is.

Historical context

The context of the current organism, however, is not enough. What is actually happening at the present moment is not in itself a function. One of my central claims is that a biological process can only qualify as a function within an historical context. In the first amoeba that develops a unified detector, perhaps by genetic mutation, the internal structure

just happens by chance to respond uniformly to the various substances that will constitute food-for-amoeba. This is just a fact, a random causal event; it is not yet the function of this structure. Only after many generations, when later amoebas inherit this capacity precisely because their ancestors replicated successfully due to this new process, does the structure take on the function of being a food detector. This becomes its role, its goal. Functionalization is an historical process. To refer to the “function” of a biological structure is shorthand for referring to the long evolutionary narrative that explains the presence of this structure. That a detector is detecting food is a current, momentary fact; to say it has the function of detecting food is to place it in the temporal context of its evolutionary genesis. Functionality is a sedimentation of the past, a congealed, crystallized history.

My claim is not simply that past historical events cause the amoeba's current organs to have a function. It is that functionality itself makes sense only in the evolutionary context. Consider for analogy a word in English: “sea” for example. The same physical sound expresses a different word in French or in Spanish. It is only in English that “sea” is a word that has the meaning “ocean.” This is not a causal relationship. The English language is not an antecedent event that causes “sea” to mean “ocean” as an effect. Rather, “sea” has its meaning—is the word it is—only within the context of English. Without that context, there may be sound vibrations mechanistically caused by vocal cords, but they only qualify as the word “sea” when placed in the linguistic structure of English. Of course, when we simply use the word “sea,” the English linguistic context remains in the background; it is implicit—“forgotten” as some put it. What I'm claiming is that while there are various factual, causal, mechanistic processes within the amoeba, they are not “food detectors”—they do not have this (or any) function—except in the implicit context of the evolutionary history of the amoeba's species.

The historical nature of biological functionality means not only that the past defines the present, but also that the future has a role in defining the past. Consider a different example. Codfish can survive in very cold waters because they have a circulating glycoprotein that acts as an anti-freeze. About twenty million years ago some cod underwent a chance genetic mutation that produced this glycoprotein. At that point, this fact had no functional significance. However, ice ages and ocean temperatures over the following few million years resulted in the survival of only those cod with this mutation. Hence, over time,

the glycoprotein took on its functional role as an anti-freeze. Only in retrospect, after the occurrence of the ice ages, does the initial mutation come to be of significance as a prefiguration of the emerging function. The future confers on the original mutation its functional destiny.

It is not only organs, such as detectors, that are functionally defined. Actions too are functional unities. If we are careful to dissociate the concept of action from explicit, conscious goal-setting—which the amoeba obviously does not have—we can use the term to distinguish an amoeba's self-directed movements from those that are just the passive effects of external forces. An action is a set of physical movements that have a function in the context of the amoeba as a unified organism. For instance, it is the functional significance of moving up a scent gradient to engulf food that distinguishes pursuit as an action from the passive movement of being carried along by a current of water. Actions in this sense are defined by their function, so it is possible that various different physical movements could still be the same action. An amoeba might pursue food in one way today and another way tomorrow, but both action would still qualify as pursuit. What defines a particular action—which kind of action it is—depends on the role it plays in the life of the organism. The amoeba's action is not “moving physically closer to a bacterium;” it is “pursuing food-for-amoebas.” The various physical movements laid out sequentially in time are unified into the one action from initiation to completion by their functional significance, not by any characteristics visible to physics. Actions are functional unities within the context of the organism's life.

Within this context, entities, events, and actions are divided into two categories: normal or abnormal; functional or dysfunctional. Normalcy is a functional notion. The dichotomy between what is normal and what is abnormal reigns throughout all of life, both biological life and, as we will see, the life of the self. When the saline concentration in a single cell varies outside the homeostatic range, it becomes abnormal and the cell dies. For an amoeba, identifying food-for-amoebas and responding by pursuit is “normal functioning;” it is exactly what its detector-pursuit structure is supposed to do. Pursuing poison-for-amoebas would be a dysfunctional action. If a cod's glycoprotein allows its blood to freeze, it has malfunctioned. In mammals the function of a heart is to pump blood; if it does not pump blood it has failed—it is operating abnormally.

Physical and chemical laws involve no such dichotomy. It is the evolved unity of an organism as a whole that establishes what is normal operation for each of its organs and what is pathological. Whether the heart pumps blood or does not, it still obeys all physical laws. Taken in isolation from bodily life, it is irrelevant whether the heart is pumping blood or not. Indeed, as a functionally defined entity, a “heart” that is unrelated to an organism is a “heart” in name only: it may look like a heart, but it doesn't qualify as an organ. Only the context of the whole amoeba and its species history allows us to make sense of the normal/abnormal distinction, a distinction that, as we will see, is central to the evolution of selfhood.

In summary, the unity of an organism, its organs, its actions, and the relations between these are all of the order of functionality. That is, they are governed by the normal/abnormal dichotomy. Biological life is not the shadowy presence of a parallel Vitalist process or soul over and above physical mechanisms, as some in the nineteenth century thought. Life should not be understood as an extra object or process. Rather, to understand life we must distinguish between analyzing entities and events in terms of physical causal mechanisms and analyzing them as functions in the context of organic unities and within the temporal context of their evolutionary history.

Chapter 2

Evolution

In this chapter, I continue my discussion of the evolutionary steps that precede selfhood. I flesh out the implications of the approach in the previous chapter for the notions of thing, world, aboutness and goal. I introduce an account of evolution that takes us beyond the fitness of individuals to how systems define their components. This sets the stage for the origin of selfhood, though the notion itself will not be examined until later chapters.

While the previous chapter focused on my hypothetical amoeba as a unified, functionally defined organism, in this chapter the focus is on the environment—the things in an organism's world—and the organism's relationship to these things. This leads to a modification of the traditional understanding of Darwinian evolution.

What is Detected?

My toy amoeba, as an organism, has organs for detection. What exactly do these detectors detect? What is their target? The detectors' function is to aim at food-for-amoebas. It is food they refer to; it is food that they are about. That which they are about is itself functional. The detector targets the diatom only as food. That is, if a detector responds to a particular diatom, it does so not because of the particular characteristics of the diatom, but only because it fulfills the function of food. It is under the general aspect of food that the detector responds to the individual diatom. There is nothing in the diatom as an isolated thing-in-itself that provokes the detector's response. If the diatom did not qualify as food, the detector would not respond. Its qualifying as food, however, is not up to the diatom itself. Its functional role as food is a relational feature, not an intrinsic property.

Note that the detector refers to food even when it gets it wrong, such as when it responds mistakenly to poison. Indeed, the detector can be incorrect only *because* its function is to be about food. Without reference to its function, the mechanism's response to a noxious substance would not count as incorrect.

The aboutness relation between detectors and their targets is neither causal nor statistical. It is a matter of defining “what they are.” Detector and food mutually define each other. If there were no food detectors, then there would be no food. Without food, the mechanisms (even if they existed) would not be *detectors*. I don’t mean that if, during a temporary period of scarcity when no food could be found, the detectors would cease to be detectors. The detectors’ function comes from evolutionary history, not from the present instant, but detectors could evolve only in a world in which food is possible. If there were nothing to detect, the notion of a detector would make no sense. Detectors are about food, but what food-for-amoebas is is just what is possibly detectable by amoebas’ detectors.

In the last chapter, I claimed that my toy amoeba is a new emergent entity. Its emergence, however, is paralleled by the emergence of food-for-amoebas, a new kind of thing. To make sense of this claim, I need to first clarify what I mean by a “thing,” for I will be using the term in a special, technical sense.

In everyday language, we use the word “thing” to refer to that which endures over time and is of a particular kind. For instance, imagine finding a raw ruby. As a discrete thing, the ruby must be bounded: it includes some parts of space and excludes other parts, such as other minerals to which it is attached. To be a thing, it must also have temporal boundaries: it must endure for some period so that it can participate in different events as the selfsame thing. These boundaries make it an individual thing that is differentiated from other individual things. It is this ruby, not that ruby. An individual thing must also be of a specific kind. This ruby is of the category “ruby.” It is of the same kind as other rubies, but it is not, for instance, an emerald. A “thing” is therefore defined by sameness and difference in two ways. First, as an enduring individual: this is the same ruby that I showed you yesterday. Secondly, as a kind: this ruby is the same kind of gem as that ruby.

Our everyday use of the word often takes it for granted, thirdly, that what a thing is depends solely on the thing itself. This is particularly true for physical things. We are inclined to think that there were rubies that were different from emeralds long before the existence of visual systems that distinguished red from green. We assume that, even if by a miracle everything else in the universe vanished, a ruby could continue to exist, to be the same individual and to still be of the kind “ruby.”

We don't always think like this, however. The everyday use of words is flexible. We have no trouble speaking about a dollar as a thing even though a dollar does not depend solely on itself. A dollar is a dollar—it only has monetary value—in the context of its social function of exchange. Abstracted from its context—taken by aliens to an exoplanet, for instance—a \$1 coin is nothing but a piece of metal. A dollar is the same thing whether it is made of metal, paper or digital bytes. What it is as a thing depends not on what it is made of, but on its social function. A dollar is an example of a functional thing.

My claim is that food-for-amoebas is a thing since it meets the criteria for sameness and difference that define the boundaries of individuals and kinds. First, the piece of food an amoeba consumes is the selfsame piece that it pursued and a different one than a piece it failed to detect. Secondly, what it pursues today is of the same kind—food-for-amoeba—as it pursued yesterday, as opposed to things poisonous to amoeba, which are of a different kind.

Food-for-amoeba, however, is a functional thing, for it does not depend only on itself for what it is. When amoebas evolved, they did not find a kind of thing, food-for-amoebas, that had been sitting around in reality for billions of years, waiting for amoebas to evolve and discover it. As a functionally defined thing, food-for-amoebas only makes sense—is only what it is—in a context in which there are amoebas with food detectors, the skill to pursue prey, and the need to do so. It is the same evolutionary development that, in unifying the amoeba as a single detecting and responding organism, thereby also defines food-for-amoeba as a thing. As a functional thing, what it is depends on the vagaries of evolution, not on some eternal essence.

Aboutness

To my question above, “What do the amoeba's detectors detect,” I can now answer, “They detect a thing,” namely food-for-amoebas. This is a semi-logical relationship: food-detectors would make no sense if there were no food to detect; food-for-amoebas makes sense only if there are amoebas to detect it. I will label this relation “aboutness.” The detectors are *about* food. The detectors *refer* to food-for-amoebas.

Note that aboutness is not a causal mechanistic relationship. Even when the detectors are not at the actual moment detecting food, they are still “about” food-for-amoebas.

Just as my house alarm retains its function even when not sounding, food is still what the detector refers to, even when it is not currently being detected by an amoeba. Nor is aboutness a factual relation: In the case of error, when the food detector in fact mistakenly responds to poison-for-amoebas, it remains the case that it is food the detector is about.

To put this a different way, it is the thing referred to that sets the standard for correctness of the detector. The detector is supposed to respond when the thing is food, and to not respond when the thing is poison. That's what it is for the detector to be *about* food: it is the nature of the object, it's being food-for-amoebas, that makes the detector's response correct (or incorrect, if the object's state is poison.) To say the detector is *about* a thing is to say that it is the state of the thing that sets the standard for correctness for the detector.

Aboutness is a major problem for dualists, like Descartes, who follow the "way of ideas." Descartes holds that, since the mind is immaterial it can only be in contact with itself, not things in the world. All the mind can grasp are ideas that represent things in the world. How could thoughts, being only about ideas, refer to the world? Descartes' own solution was to resort to the miraculous goodness of God, who, having created the ideas in our minds and also the things in the world, arranged that the ideas *represented* the things. Aboutness was therefore an exclusive feature of divinely created conscious minds. Amoebas, of course, having no immaterial minds, are therefore, from this dualist point of view, excluded from the realms of aboutness.

Many who reject this mind-body dualism, however, still cling to one of its relics: representation. So they are tempted to analyze aboutness by saying that an amoeba *represents* food-for-amoebas. That is, inside the amoeba there is something like an image that looks like food-for-amoeba, just like a photo of a rabbit looks like a rabbit and thereby "represents" it.

I reject this understanding of aboutness. It is not that the amoeba represents food, or is conscious of some physical objects as food. It doesn't think of diatoms and so on as food. An amoeba doesn't think. It has no mind! What the amoeba detects is outside the amoeba; it is external to the organism. Of course, there are structures within the organism—the unified detector-pursuit release organization—that are related to the unified food outside the organism. But it would be misleading to refer to these internal structures as "representations"

of the food; they have no resemblance to food whatsoever. This is a relic of Cartesianism. Aboutness should not be understood as a factual similarity between an internal representation—an image or idea—and an external thing.

Even if these internal structures were similar to food—which they are not—the amoeba's detectors would not be equipped to detect them. The detector's function is to detect food-for-amoeba in the world, outside the amoeba. The detectors do not detect anything within the organism. Talk about representations misses the point: the relation between internal detectors and external things is a functional aboutness relation.

Instead of appealing to representation, it is better to understand aboutness in the context of fitness. A species must be fit for its environment, otherwise it faces extinction. Organisms that lack the ability to detect food are unfit for their environment. The food-for-amoeba that the amoeba needs to survive must be the same thing that its detectors refer to. There is no need to appeal to a divine miracle to understand this aboutness correlation. That the amoeba's detectors are about food-for-amoeba is a result of the evolutionary history that led to the amoeba being fit for its environment.

Some are surprised at the fit between organisms and their environments and think of such fitness as miraculous, that is, as requiring divine intervention to explain it. But a species must be fit for its environment, otherwise it couldn't survive. Fodor claims that this is a “rigged game.” He points out that ... if a creature fails to fit an ecological niche exactly, it follows that that isn't exactly the creature's ecological niche.”⁴ That an organism fits its niche, he claims, is not a surprising empirical finding, but a tautology: it is true by definition. Fish are not fit for life in the trees around the lake nor are birds fit for swimming in the deeps. The fit appears only when we relate the organism to the specific aspects of the environment that are relevant to the species.⁵

I agree with Fodor's insight that the organism and its niche are two sides of the same coin, but he is still thinking of the “environment” as physical reality. Uexkull points out that, even in the same physical setting, each species lives in its own world. One of his numerous examples is the tick:

The constitution of the tick, which is blind and deaf, is composed only to the end of allowing every mammal to appear in its environment as the same carrier of meaning. One can describe this carrier of meaning as a radically simplified mammal, which has

neither the visible nor the audible properties by which species of mammal are differentiated from each other. This carrier of meaning for the tick has only one smell, which comes from mammals' perspiration and is common to all. Furthermore, this carrier of meaning is touchable and warm and can be bored into for blood withdrawal. In this way, it is possible to find a common denominator for all these animals, so different in shape, color, emission of noises, and scent expression, such as we have before us in our environment.⁶

What Uexkull contributes to our understanding is that the world which organisms live in is not the physical world defined independently of the organism, but the world as it is *for* the organism. The things in an organism's world are those that are unified by the perceptual and behavioural powers of the organism. The tick detects a radically simplified thing that includes all warm-blooded animals without distinction, reminiscent of my amoeba's unification of many things into the one food-for-amoeba. What gets treated as the one kind of thing for the amoeba is analogous to the many mammals that get treated as one kind of thing for the tick.

An organism does not relate to things simply because they happen to be in the vicinity. An organism's perceptual organs are not about things defined by physical terms however carefully Fodor selects them for relevance. An organism refers to things-for-the-organism. An organism's niche is composed of emergent things functionally defined by the skills and perceptual systems of the organism. So fitness is no miracle: it is inherent in the pragmatic nature of the things-for-an-organism that make up its world.

Since the world-for-a-species is made up of things in so far as they are useful for the survival of the species, these things are value-laden. Food-for-amoebas has the value "that which should be pursued;" poison-for-amoebas has the value "that which is to be avoided." It is a founding principle of the mechanistic approach, as we saw above, that the objects of physics must be value-free. The sun "rises," for instance, not because this movement is useful for humans, but because of the effects of gravity and momentum on the solar system. In contrast to this mechanistic approach, I am claiming that the things in the amoeba's world are things of value: they are beneficial or noxious to the organism. To mechanists, gold is just one element among others. They interpret the fact that a human considers gold more valuable than lead to be a purely subjective overlay that is solely in the eye of the beholder. It is not in the nature of gold-in-itself. In the absence of humans, gold would continue to exist, but would have no value, positive or negative. This is the fact-value distinction, so dear to mechanists.

Food-for-amoebas, however, is valuable to amoebas in that without it they would not survive, but this value is not like the value of gold as mechanists understand it. The positive value of food-for-amoebas is not a superficial add-on, a subjective overlay on a thing-in-itself. Its value *defines* food-for-amoebas as a kind of thing in the first place. In the absence of amoeba, food-for-amoebas would not simply hang around, valueless: it could not be food-for-amoebas without amoebas, their needs, and their skills. Food-for-amoebas is a contextual, functional thing, so its value is essential to what it is. Without its value, that kind of thing could not exist.

Our common-sense notion of things as dependent only on themselves makes it difficult to find words for such contextual things. Uexkull labels them “carriers of meaning.” He says for instance that the tick’s world is made up of “carriers” of scent. Tønnessen prefers to call such carriers of meaning “appearances” or “phenomena” and therefore interprets Uexkull’s work as a kind of “phenomenology”—a description of how things appear to an organism of a particular species.⁷

While this “phenomenological” approach is tempting and takes us beyond Fodor and Uexkull, I worry about the terms *appearance* and *phenomenon* since they have connotations of experience, consciousness, and subjectivity. Analogies and metaphors risk offering explanations whose plausibility depends on dragging in associated concepts inappropriate to the field towards which they are retargeted. To speak of an amoeba’s phenomenal experience successfully distinguishes the things in its world from objects as defined by physics, but I fear it succeeds by covertly attributing consciousness to amoeba, an attribution I find absurd. We have some notion of what experience is for humans; we have no notion of what the word could mean when applied to amoebas. Such attributions are fraught with dangers of dragging in anthropomorphic paradigms which may give us the illusion that we already understand processes that actually need detailed analysis. I think it is safer to stick with the non-anthropomorphic notion of function: things in the amoeba’s world are contextually defined by their functional role in the amoeba-system.

Uexkull himself takes the German word “Umwelt” (surrounding world) and gives it a special technical sense: an Umwelt refers to the things in an organism’s world in so far as they are perceived and acted on by the organism. This distinguishes them from objects in the

world of physics. I will adopt Uexkull's technical use of "Umwelt" from here on. So, in my own example, I will say that food-for-amoebas is an "Umwelt thing" in the amoeba's world, or in the amoeba's Umwelt.

An amoeba's Umwelt is not established by each amoeba as an individual organism. The world and its things are defined by the historically evolved species. It is the species that adapts in evolution and so it is the species that establishes the functional roles of both the organism's organs and the things of its world. It is the species that develops a specific form of unity and passes this organization on from generation to generation. To qualify as a member of the species an individual organism must have this organization. What is normal functioning for an individual's organs and what the correlated values of the things in its Umwelt are is established by the species.

Species, of course, change. Aristotle held that species are eternal, though individuals come and go, and the notion that each species defines a stable essence lives on in everyday language and common sense. Darwin's adaptationism, however, undermines this notion. We need to think of a species as a kind of temporary resting point in the ongoing evolutionary process. In our particular time-slice, we may conceive of an organism as a bird, but in some previous time-slices its ancestors were dinosaurs. A species is a repository for the accumulated experience of the clade's past, a crystallized history that, for the moment, defines what normal life is for an organism at a particular period in time.

This doesn't mean that, since there are no stable species, the distinction between normal and pathological disappears. This objection is a relic of the essentialist view. It is true that the normal-abnormal distinction is not an absolute one. In any particular case, it is a temporary distinction dependent on the chance developments of evolution. Don't say that because we can't have it all, we have nothing! Just because species are transitory doesn't mean there is no normal; it just means that what is normal changes as evolution moves on. During the period that a species is relatively stable, it defines a mode of being that individual organisms should emulate: if they do not, then they are no longer part of this mode of being, no longer members of the species.

Co-evolution

That species are adapted to their environment is the central thesis of Darwinism. In the competition for survival, the fittest win out. While this thesis is, on a fundamental level, correct, it is now time to face up to some of its limitations. First, adaptationism is an overly simplistic theory of evolution: its organism-centred view needs to be replaced by an account based on systemic relations. Secondly, it focuses too exclusively on competition, when cooperation is just as important. To explore these alternatives, let me introduce a new toy scenario, different and later in evolution than my amoeba scenario.

The perceptual apparatus of bees includes ultraviolet detectors. Bees use these detectors to track down ultraviolet-coloured flowers for their nectar. They pick out ultraviolet-coloured flowers as a class and respond in a similar way to all of them—by flying to them and sucking their nectar. As I've been putting it, an ultra-violet-coloured flower is one kind of thing in the world-for-bees, in the bee's *Umwelt*. For the bees, these flowers are functionally categorized by the motor value they have: they are to-be-flown-to.

From one perspective (that of the bees!), we can offer a simple adaptationist explanation as we did in the case of the amoeba: given the presence of ultraviolet-coloured flowers in their environment, bees with better ultra-violet detectors have a selective advantage. They can gather more nectar and so survive and reproduce better than other bees. Over a period of generations most surviving bees will be the ones with better ultraviolet detectors. The evolutionary history of bees has established a world-for-bees, a world that includes ultraviolet flowers as things-for-bees. Since the flowers are what the ultraviolet detectors are about, we could say that the ultraviolet-flower kind is unified as one class of things by the bees' detectors. (Since humans have no ultraviolet detectors, ultraviolet flowers are not a kind of thing in the perceptual world of humans.) We then have an organism-centred perspective on the presence of ultra-violet flowers-for-bees.

There is another perspective on the situation however: the flowers'. To reproduce, flowers need insects to pollinate them and, in an environment with ultraviolet-perceiving bees, those flower species that have the best ultraviolet pigments have a selective advantage. Over time, more of the surviving flowers will have ultraviolet pigments. The bee species is part of

the environment to which the flowers have adapted. From this perspective, bees are functionally defined things in the flowers' world. Bees are things-for-flowers.⁸

Both of these alternative perspectives are based on an overly simple, adaptationist, interpretation of evolution. Such adaptationism assumes that what evolves is a species which is selected for its fitness relative to a given environment that is held static while we explain the species' adaptation to it. The point of my hypothetical scenario, however, is that the environment is not static; it too evolves.

A more sophisticated account—"co-evolution"—proposes that what evolves is neither the bees nor the flowers, but the interdependent system of their interaction. The two components are in a mutual, positive feedback loop. It is not just that bees rely on flowers and flowers on bees; the adaptive development of each provokes further adaptive development in the other. This is not a matter of two perspectives, but of a mutual "enaction" as Varela calls it. Discussing the bee-flower case, Varela makes the point clear:

This insistence on the codetermination of mutual specification of organism and environment should not be confused with the more commonplace view that different perceiving organisms simply have different perspectives on the world. This view continues to treat the world as pregiven; it simply allows that this pregiven world can be viewed from a variety of vantage points. The point we are making, however, is fundamentally different. We are claiming that organism and environment are mutually enfolded in multiple ways, and so what constitutes the world of a given organism is enacted by that organism's history of structural coupling.⁹

Since the flowers also co-evolve with organisms in the soil that deliver nutrients to their roots and which in turn feed on decaying flowers in the soil, and since the bees have their own co-evolvers, what we have is a complex mutual relationship in which each component receives positive feedback from many other components. The simplistic explanation based on a dichotomy between a privileged active, adapting organism and its passive environment—of categorizer and categorized—must be dropped in favour of a wider interaction of multiple components that co-define each other.

From the viewpoint of this more sophisticated evolutionary theory, we can say that it is not only the bees themselves that adapt; nor is it just the flowers and soil. Rather we have a dynamic set of relationships, a system that evolves as a whole. It is this historically developing system that establishes the functional roles of the entities within it. It is not just

that flowers-for-bees are functional objects in the bee's world or bees-for-flowers in the flowers' world: both kinds of objects are defined by their functional role within the ecosystem as a whole. While it is true that bees have perceptual organs, it is an oversimplification to say—as I did above—that it is these organs that define and categorize the flowers.

This co-evolutionary approach corrects a common misconception of evolution. Thinking of Darwinian evolution as “survival of the fittest” and imagining it solely in terms of competition is only half right. Far from competing, bees and flowers rely upon each other for survival. Not only do they currently work together with each other, the evolutionary history of the two species is dynamically intertwined: each has evolved only because their presence enhanced the fitness of the other. Of course, bees with superior ultraviolet detectors out-compete bees with inferior detectors, but focusing on this feature of the relationship obscures the essentially cooperative nature of evolutionary systems.

Systems

In the amoeba scenario, I proposed that it is the unification of the food detectors that establishes food-for-amoeba as a kind of thing. We can now see that this is a limited understanding of the situation. While, within the system composed of amoebas and their things there are different functional roles, no individual component has the unique privilege of defining these functions. The ultraviolet perceptual organ of the bee, on the one hand, and the flower's attractive colour on the other, are functionally defined entities only when we place them in the context of the overall bee-flower system. Organisms by themselves, isolated from their world, cannot establish any functionality. Organisms do not have the exclusive capacity to assign functional roles to their organs or to the things of their worlds: organisms are as much functional entities as are Umwelt-things. The ultimate basis for the functional significance of all components of the system—organism, organs, things-for-organisms and so on—is the evolved system as a whole.

This systemic account preserves some major features that the amoeba scenario allowed us to recognize. While naturalistic, the account remains non-reductive: entities within a system are functionally defined. To understand how the components of a system interact, we must look not to their mechanistic properties, but to their functional roles. Different flowers may use different pigments for their ultraviolet colour, that is, pigments governed by different

chemical and physical laws. But this is irrelevant for understanding which kinds of things are defined by the system. Similarly, it doesn't matter to the functional relations which physical processes are involved in the bees' capacity to detect ultraviolet light. The functions are defined by their roles in the system: any process that could carry out the same function would do. In principle a biological system could rely on different physical processes than those on which it has in fact depended. Life based on silicon chemistry rather than carbon, for instance, is theoretically conceivable.

Conclusion

The position this chapter has defended is that biological life should be understood systemically. This account is as much about the nature of understanding as about biological life. While biological processes depend on universal, physical, causal laws, they cannot be understood by appeal to such laws. Organisms and their worlds emerge from the chance events of evolution and can only be understood in terms of the functional interrelationships that originate from this history. It is the system that differentiates between what is functionally normal and what is abnormal. Each system as a whole establishes organisms as unified entities, defines the functions of their organs, determines the perceptual and action skills of these organisms and thereby defines the things that make up the organism's world, its Umwelt. Only within the system are its component parts what they are.

We will see that the evolution of humans and, later, of selfhood can also be best understood by this holistic, systemic mode of understanding. These first chapters have introduced key concepts which, as we will see later, enable us to analyze the nature of the self. A self too is an emergent unified whole that is insulated from the direct causal effects of the environment. Selves are embedded in history, both their own narrative history and the history of their culture. They, too, are normative entities, constituted by values, functions, and commitments, though in a different way than amoebas or bees. They live in a world of things that their thoughts and words are about. Unlike purely biological organisms, however, humans live in a context of cultural and linguistic systems. It is to the development of these systems that we will now turn in the next stage of our journey towards selfhood.

Chapter 3

Culture

In the last two chapters I have offered a speculative story that could explain the emergence of biological unities that are structured functionally—organisms. While this evolution is a crucial step in our journey towards selfhood, we still have a long way to go. Amoebas or bees, whatever proto-agency we might attribute to them, are far from the autonomous, language-equipped and reflective selves to which we aspire. How could we have got here from there?

Selves evolved gradually through the long evolution of the genus *Homo*. Well over two million years ago hominins started to develop larger brains, invented and came to rely on tools, and adopted ways of life that depended on individuals working together. They survived on the basis of skills preserved and accumulated across generations. Our experience of communal life today is so intrinsically linguistic that it is hard for us to imagine how cooperative groups could have evolved in the absence of our contemporary, verbal, syntactically structured language, yet given the paleontological evidence, imagine it we must.

In the next chapter, I will discuss language in more detail. In this chapter, the task is to explore the emergence of an intermediate stage between purely biological life and the mode of human existence I call selfhood. How might it have been possible, without relying on language, for early hordes of hominins to form cooperative groups with traditional and distributed knowledge and skills?

While the evolutionary journey from amoebas to hominins took hundreds of millions of years, chapters 1 and 2 have already covered some of the most crucial conceptual tools that we need for understanding it. We have already examined the relation of systems to their parts, the partial independence of more complex unities from their underlying processes, and the normativity that functions exhibit. In this chapter, building on these concepts as I have presented them in my toy scenarios, I propose that social groups of hominins can be seen as

new emergent systems within which individuals and their cultural worlds—their *Umwelten* — come into being.

I will proceed, as usual, with another simplified scenario. While I will occasionally refer to empirical finding in paleontology, the scarcity of hard information in this area means that my approach will be mainly a rational reconstruction of what might have happened. The aim, after all, is to investigate the nature of selfhood and overcome our initial difficulty: from a dualist and mechanistic perspective the evolution of free, autonomous persons from purely biological origins seems in principle unimaginable.

I hypothesize an early hominin society that, even prior to language, has found ways of unifying into small groups based on the transmission of skills from generation to generation. This cumulative inheritance of cultural learning requires a division of labour which would only be possible if individuals cooperate for the common good of the group.

That cooperation is a crucial component of hominin evolution is widely accepted, but the term “cooperation” can be seriously misleading. In ordinary English, cooperation has the connotation of freedom, of individuals voluntarily agreeing to work together. A slave forced to work for a master would not typically be said to be “cooperating.” I will use the term in an extended sense to refer to the behaviour of any organism that increases the likelihood of survival and reproduction of other organisms. Mutualism between species, such as the co-dependence of bees and flowers, therefore qualifies as cooperation in my sense of the term, though there is no voluntary agreement involved. Cooperation need not be contemporaneous: if fungi leave minerals in the soil that is used by future plants, that too qualifies as cooperation—cooperation across time. In the context of my thought-experiment investigating the origin of selfhood, there is a danger of misunderstanding the term “cooperation” anachronistically by attributing to early hominins the very characteristics—autonomy, rationality, responsibility, and self-consciousness—whose origin this investigation is attempting to explain. While it might be better to use some other term (mutualism, for instance), “cooperation” is so well established in the literature that I will, after this warning, continue to use it, but only in the extended (and etymological) sense of “operating together,” that is, stripped of its common-sense notion of conscious voluntariness that we usually associate with the term.

My claim, in this chapter, is that early human cooperation—in my extended sense—is based on the learned patterning of bodily behaviour that makes up group identity. It is the contextual norms that bind individual group members together that explain, even without appeal to rational convictions and explicit linguistic communication, how cooperation evolves.

In the first section of this chapter, I will discuss how cooperation could have originated. I will then argue that group cooperation should not be understood in an intellectual or rational manner. In the second section, I will present my alternative: group identity is a patterning of bodily and emotional behaviour rather than a matter of thinking. The last section explores the implications of group unity for the normative structure of individual hominins and of their worlds.

Working Together

In my amoeba scenario, the amoeba's behaviour is primarily dependent on its genes. A distinctive feature of humans, however, is their dependence on learned behaviour. “Learned behaviour,” however, is an ambiguous term. Almost all individual animals are capable of learning during their lifetime, but hominins have come to also rely on behaviours that were learned by their ancestors and that were passed on from generation to generation. To distinguish it from individual learning, let me call such transmitted learning “tradition.” Tradition is a kind of software DNA: techniques for surviving are passed on by the inheritance of skills that have proven successful over generations. Adaptation by genetic mutations is slow: the ability of hominins to transmit learned behaviour from generation to generation in a cumulative way, that is, “culture,” has enormously increased the speed of evolution.

Focusing on change or on the origin of new species may incline one to think of evolution as primarily a matter of change. The fundamental requirement of evolution, however, is preservation. Once a way of life that promotes survival is found, it must be passed on relatively unchanged to future generation. How to make a cell membrane, for instance, has been conserved for billions of years. Successful genetic mutation are therefore relatively rare. The vast majority of mutations are noxious to the cell, the organism, or the species.

The evolutionary need for preservation of successful modes of survival carries over into human traditional lifeways based on cumulative learning. Early hominin cultures must therefore have been very conservative, passing traditional ways of life faithfully to descendants for centuries or millennia.

How is this accomplished? First, the children of hominins are born immature, and so must remain with their parents for longer periods than is the case with many other species. As a result there is more opportunity for transmission of traditional skills from parents to children. Transmission is enhanced further when offspring are raised in a group setting so that the child is exposed to the skills of a variety of adults. The preservation of skills is also enhanced by a group setting since the danger of a skill not being passed on—that is, of it going extinct—is reduced where numerous individuals act as depositories and transmitters.

While culture is primarily conservative, on the rare occasions when new survival skills are discovered, they can be added cumulatively to the tradition and preserved for descendants. Cumulative learning, however, should not be misunderstood simply as the addition of new skills. As in the case of the amoeba, we must be wary of an overly simplified adaptationism. It is not that there is a fixed environment within which skills that increase fitness out-compete those that don't. The environment within which skills compete is itself a cultural construct that has evolved from the history of previous cultural evolution. The use of one type of stone tool, for example, may out-compete the use of another type, but only within a cultural environment where the practice of using tools is already established. Once adopted, the new tool in turn changes the cultural environment within which a future tool-modification must survive or become extinct. The cultural medium sets its own standards for fitness beyond the biological.

Cooperation within Tribal Groups

Cooperation is itself an inherited traditional skill. It is a distinctive feature of humanity. From hunting to agriculture, from building cities to fighting wars, from science to technology, humans cooperate on a unique scale. Today we have over seven billions humans who cooperate globally in industrial production and distribution, in research and communications, in regulation and governance. Indeed, it has become common-place to claim that cooperation is the main evolutionary strategy of hominins. How could such an evolutionary strategy work?

One could think of evolution in terms of the capacities of individuals within a species. In that case, fitness is achieved by improving each phenotype. Rabbits jump fast and so avoid predators. Butterflies flit around and so avoid birds. Similarly, one could think of hominin evolutionary success as due to individuals developing more strength, running speed, brain power, or other characteristics which are then inherited by subsequent generations.

An alternative possibility for species success, however, is for individual organisms to work together with other members of the species, rather than to improve the capacities of individuals. How could an individual hominin take down a mammoth? It would need enormous strength and skill. However, if a dozen hominins coordinated their activities by chasing, trapping and surrounding a mammoth, the strength of individuals would be less important. This is species fitness by cooperation. (Recall that by “cooperation” I just mean working together. The word must be stripped of its everyday connotation of voluntariness.)

Cooperation in some form can be found throughout nature. Bees and other social insects can be said to cooperate (in my extended use of the term.) Wolves and other pack animals work and live together in groups. Mammals take care of their young; that is already a form of biological cooperation. Early Hominins, however, appear to have extended cooperation beyond genetic kinship to form small tribal groups based on culture and tradition rather than just biology. Such groups lived, foraged and raised offspring together and thereby enhanced their survival.

When such tribal cooperation started cannot be pinned down to a definite date. The evolution of cooperation was a slow gradual development over some millions of years. While there may be earlier species of the genus *Homo*, such as *Homo habilis*, the best evidence for cooperation is for *Homo erectus*. *Erectus* originated about 2.3 to 2 million years ago, probably during a period of drying climate, which would explain the adaptation to a savannah environment, to the hunting of big game and to the need for tools, fire and other technologies. This suggests that group cooperation dates back at least two million years.¹⁰

Cooperation took many forms. For examples, consider the following:

— Hunting: Many of the hunting techniques on which humans came to depend could only be carried out conjointly. Hunting large animals, for instance, required organized hunting

parties. Here is some evidence for hunting of antelopes during the Early Stone Age as early as two million years ago.¹¹

— Protection: Group members could alert others to predators and defend each other against them.

— Technology: Making stone tools was a specialized skill that few individuals could possess, so individuals had to depend on others for their tools.¹²

— Fire: As humans learned the advantages of cooking food, they depended on the use of fire. Since starting a controlled fire is difficult, consistent fire use—which might go back as far as 1.6 million years ago¹³—requires that campfires had to be kept burning. Individuals who tended the camp fire had to rely on the spoils of hunters and gatherers.

— Birthing: While other primate females can give birth alone, as human brain size increased, birthing through a narrow birth canal became more difficult and the resultant change in the orientation of the infant made it useful for hominin mothers to get help from midwives.¹⁴

— Childcare: Hominin mothers lacked the fur that chimp babies cling to, so hunting and gathering become more difficult for them and they came to rely more on kinfolk to either supply food or to care communally for the babies.

— Protecting children: Since human infants need years to mature, parents had to rely on the community to prevent infanticide, a problem in other species.

— Education: Parents depended on others to teach their children specialized skills. Creating some flint tools, for instance, may have requiring firing the flint to specific temperatures, neither too hot nor too cold.

These examples illustrate the kind of cooperation that was central to the very successful evolutionary strategy of early hominins. Outsourcing various needs to the group meant that individuals didn't have to rely entirely on their own strengths.

Such outsourcing to the group can be thought of as self-domestication. Domesticated animals don't need as much individual strength or intelligence as their wild counterparts, for they are looked after by humans, who ensure they get enough food, mates, and protection from predators. As a result they are often weaker, slower and less intelligent than their wild

counterparts. Cows, for example, are slower and less aggressive than their wild cousins. As an evolutionary strategy, getting themselves domesticated by humans has been a huge success for cows. Like pigs, dogs, and cats, cows didn't become so numerous by increasing their individual capacities, but by getting domesticated. Analogously, within a tribal group individuals no longer have to fend for themselves or evolve the capacities to do so. The group offers protection from predators, assistance in foraging, help in child-rearing, and so on. Tribal cooperation can be thought of as human self-domestication.¹⁵

Cooperation itself needs skills. It is not so much that cooperation reduces the need for individual skills, as that it modifies which skills are required. While individuals may no longer need the capacity to escape or ward off sabre-toothed tigers by themselves, they need the social skills to be able to get along with the other members of the tribe. They must learn not only what their own tribal roles are, but also the roles of others in the group, otherwise they cannot work together. (Some have referred to this development as “survival of the kindest.”) Keeping track of other individuals, their roles and relationships is a task that grows exponentially with the number of members in the group.

Dunbar claims that the evolution of human cooperation is tightly associated with an increase in brain size. According to his “social brain hypothesis,” the larger brain volume is needed, not primarily for using tools, but for maintaining cooperative social relations. The task of keeping track of relations between individuals in a group requires more neural resources as the size of the group increases. Chimpanzees, with a brain size of 500 ml live in groups of about 50. *Homo erectus*, with twice that brain size, could relate, he hypothesizes, with 100 people. Early humans lived in tribes of about 150 and have a brain volume of about 1500 ml. Tool use changed relatively little and slowly from two million to 10,000 years ago; it is the increase in the size of the social group that created the selection pressure for the increase in brain capacity.¹⁶

Objection: Competition

The formation of small tribal groups of cooperators changes the evolutionary landscape. In purely biological evolution, it is the set of capacities that are transmitted genetically from generation to generation that define a species. Individual organisms live and die; it is the species that is the unit that evolves over time, that survives and replicates or

becomes extinct. Hominins are significantly different in that they rely also on life skills that are transmitted by culture. Such transmission occurs largely within each tribal group, so over time each group develops its own distinctive tradition of life—its culture. Some traditional ways of life are more successful than others, that is, some groups survive and replicate better than others. The most cooperative tribes—where cooperation means between generations (that is, tradition) as well as between contemporaries—are the ones most likely to prosper. So in life that is based on culture, each tribal group plays a role analogous to a biological species. In evolutionary terms, then, the group becomes the unit of evolution that survives or becomes extinct.

This approach to group evolution has been challenged by some proponents of game theory. They argue that cooperation within an evolutionary group is impossible in principle. The argument is that, since evolution is based on competition, the fittest strain of organisms is the one that will win the competition in the long run. Even if a cooperative group of organisms were to be formed, individuals within it who took advantage of group resources without contributing—free-riders—would have a reproductive advantage, so over a period of generations non-cooperators would predominate, and the cooperative group would break down. Indeed, it could never get started. The argument is not based on an ideology of human selfishness: its claim is that any individual biological organism with a genetic inheritance that allows it to control more resources than its peers will be the most successful replicator in the long run. According to this theory, groups can therefore never be the units of evolution: individual competition will always trump group cooperation.

Empirical phenomena show that this theoretical challenge must be wrong, even on a purely biological level. Slime mould, a type of amoeba, is made up of individual organisms in competition for resources. Yet under certain conditions of scarcity, when resources are not available for all the individuals to reproduce, the organisms coalesce into a structure in which most of the individuals support the production of spores by only one. That is, individuals in the structure sacrifice their own reproductive potential to the welfare of others. Even in biology, individual competition by free-riders does not always win.

But we don't need to look to exotic phenomena like slime moulds. Multicellular organisms involve individual cells subordinating their own potential for reproduction to the

needs of the whole organism. When some cells fail to do so, we have the growth of cancerous cells who, as it were, “selfishly pursue their own destiny.” If the game theory position were correct, there could be no multicellular organisms.

On the human level, all societies involve some similar sacrifices of individual goals for the good of the whole. Medieval scholars were often celibate, renouncing their own reproduction to promote research and education for the benefit of the community. Warriors may sacrifice not only reproduction, but also their own survival, to protect their society.

So there is plenty of empirical evidence that cooperation between individual organism, including hominins, actually occurs. On the other hand, the theoretical argument that competition by the fittest individuals must always undermine cooperation is plausible. How, in the face of these theoretical arguments, can we understand how cooperation came about? I will examine, and reject, a couple of “contract” theories, before offering my own solution.

Contract Theories

Thomas Hobbes, writing in the middle of the seventeenth century, offered a social contract theory for the origin of society. He is famous for claiming that, prior to the formation of political society, people lived in a state of nature, individually competing with each other, often violently, for resources. In this “state of war,” the life of man was solitary, poor, nasty, brutish, and short. Since it was to their rational advantage to overcome this state of war, people decided to make a contract to cede power to a sovereign who would establish a social order, a state, within which competition could be restrained and social cooperation enforced. Hobbes' primary concern is to offer a rational justification for the state and its power, so he is only incidentally proposing an historical account of how it came about. Nevertheless, the image of individual humans forming a rational contract to cooperate has been deeply ingrained in Western political thought ever since.

Tomasello, a twentieth century thinker, offers a more plausible account than Hobbes. Building on the intellectual abilities of chimps, he theorizes that early hominins learned first how to operate in pairs. A pair of hominins might initiate a joint project, to hunt a gazelle for instance. Such a joint project required communication and in this early pre-linguistic situation Tomasello hypothesizes that they relied on gesture and pantomime.

Tomasello's next evolutionary stage is group-based. This is the stage of *Homo sapiens*. Unlike the *ad hoc* pairing for projects on the basis of individual ability, the collective stage involves setting conventional, transpersonal, cultural norms for “how things should be done” by everyone in the community. These norms originate in the group, group members feel obliged to obey them, and people can assume that everyone else in the group will also follow them, even those they have not previously met or not yet worked with in joint projects.

This is the collaborative creature we are imagining here, and for the most clarity we may focus on its culmination in hominins of about 400,000 years ago: the common ancestor of Neanderthals and modern humans, the ever mysterious *Homo heidelbergensis*. Paleoanthropological evidence suggests that this was the first hominin to engage systematically in the collaborative hunting of large game, using weapons that almost certainly would not enable a single individual to be successful on its own, and sometimes bringing prey back to a home base (Stiner et al., 2009).¹⁷

Tomasello maintains that at this collective stage cooperation is achieved by group identification. Each individual identifies with the group and recognizes other members as “one of us.” Those who follow “our” way of life are members of “our” community; they share our projects, our beliefs, our values, and our practices. Each of us can tell which humans are those from whom we can expect cooperation, even if we have never before embarked on a joint project with them. Appearances and practices identify them as “one of us” and so we can rely on them to join a hunting party, to follow standard mating or marriage rituals, or to cook food the same way as “we” do.

An individual who does not do things our way is not “one of us,” they cannot be counted on, and so will be excluded from collective ventures. Individuals enforce these norms on others in an agent-neutral way (by gossip, for instance.) Guilt, shame, and concern with reputation show that each person enforces the norms on themselves – “normative self-monitoring.” This in-group/out-group psychology, claims Tomasello, is unique to the *Homo sapiens* and is not to be found among other apes.

Tomasello's hypothesis about human evolution has much going for it, and my own position borrows significantly from his ideas. Between the competitive behaviour of the apes (and presumably of our common ancestors) and full-blown human cooperation, there must be intermediate stages. His early human stage, however, assumes that individual hominins pair off in twos for various tasks. Since chimps and bonobos live in groups, the speculation that

the immediate stage is based on parties of two is unlikely. Empirical evidence for this assumption is hard to find, and as a conjecture, I don't think it is very plausible.

Apart from empirical evidence, however, I am more concerned with a conceptual issue. Tomasello seems to imagine an Early Human as rationally planning to organize a hunting project with a second individual. It is as if the two are self-conscious, rational individuals, lacking only developed language. The image is of modern selves with something like locked-in syndrome. Since my project is to trace the origin of selfhood, it is crucial not to smuggle into this history a concept of the self that only comes into being at the end of the story.

On this point, Tomasello is following Hobbes. The Hobbesian notion of cooperation by contract presupposes that, in the state of nature, there are selves already formed who can intellectually evaluate the advantages of cooperation and voluntarily enter into a contract for which they are responsible. Such contracting would surely not be possible without a developed language. But it is the initial formation of such a rational, linguistic self that is at issue, so this approach is circular. Responsible, rational, autonomous, contracting selves are a construction of society, I will claim, not a prerequisite for it.

Group Identity

We need an account of the origin of human groups that is not circular, that does not anthropomorphically (and anachronistically) project our contemporary notion of selfhood back into hominin evolution. We must not presuppose what we are setting out to explain. I want to propose a way in which early hominins groups get formed on the basis of pre-linguistic, pre-rational and pre-intellectual processes. The account must also explain why individuals pursuing their own fitness—free-riders—do not inevitably undermine the formation of cooperative groups. This is the positive task to which I now turn.

Rather than assuming that early hominins were like us and relied on reason and consciousness, my position is that early cultural patterning is best understood as structuring bodily behaviour and emotion. The concept of a thinking mind obeying rational norms that are communicated to it by explicit language risks anachronistically importing more recent developments into the evolution that only later led to selfhood. Early group identity, I

hypothesize, was based on emotional bonding, not on intellectual or pragmatic reasoning by autonomous, self-conscious individuals. Tribes were held together by group identities that set norms for cultural behaviour, norms that were learned in childhood and reinforced in various ways in adulthood.

What would such pre-rational group identity look like? Let me use the analogy of a contemporary team sport, such as soccer or hockey. A team scores and the fans stand and cheer. The fans act spontaneously, not deliberately. They are not thinking of some obligation of group membership, rationally deciding that cheering is the right thing to do or following some rule laid down that dictates how they should act. They are just going along with the others because “this is who we are—we are Jay supporters;” they are expressing their solidarity with their team. What it is to be a team fan is, among other things, to dress like one, to wave banners, to stand and cheer. That is, they conform to the expectations and traditions of the group. Imagine a person in the group who doesn't stand: the fans expect her to stand and cheer, they are perplexed and distressed that she does not. She provokes puzzlement, a sense of impropriety, possibly anger, maybe a desire to exclude her—she'd better sit some place else! Organizers sometimes place fans of opposing teams in different sections to avoid such potential violence. Her behaviour, however, is unlikely to provoke a claim that she has violated some explicit, rational rule: none of these reactions are based on autonomous reasoning. They may not even involve language. Team spirit is an emotional phenomenon, a series of synchronized bodily behaviours, not a contractual, intellectual matter.

This analogy has its limits. While sports fans may check their autonomy at the entrance and act exclusively as part of the group, they do have another life outside of the game. If things get out of hand, other norms of conduct, criminal law for instance, may cut in and subdue the enthusiasm. Instead, we need to imagine an early human mode of group bonding that is exclusively this kind of life-way, without outside laws, without autonomy of any kind. My conjecture is that such behavioural conformity can largely account for paleolithic group identity without smuggling in a notion of a rational self, or of a Hobbesian solitary individual.

Some primates, such as orangutans, live largely solitary lives. But it is unlikely that early hominins were solitary creatures. The closest primates to the *Homo* genus are

chimpanzees and bonobos, who are social species. So it is more likely that hominins lived in groups from the very beginning—even before the beginning—of the genus. Imagining Hobbesian individuals rationally establishing groups, or Tomasello's individuals forming pairs for mutually beneficial projects, is not a promising way of looking at the origin of hominin groups. A better picture is that of primate bands held together by instinct evolving into hominin communities held together by group identity. A normative system that is biological morphs into a normative system that is also cultural.

Cultural individuals do not *join* groups; each is in a group from the beginning. An infant is already within a culture as it becomes an individual. Indeed, its becoming an individual, and an individual of a certain kind, is itself a cultural activity. The baby does not outsource its food supply to the mother's breast. It does not outsource the patterning of its cognition, its emotions or its hormones to the community. It never had these things originally and independently in the first place.

Children are reared within a group from their beginning. Babies are bonded to their mothers from birth, or even earlier. While the bond is initially biological, the mother's care and affection reinforce the relationship and carry it into adulthood. The child's biological body develops within the tribe's culture which patterns this body. The child does not enter the tribe from the outside, as it were, and have to explicitly learn what the cultural norms are and deliberately decide to obey them. Rather, the child is always already within a culture that patterns its maturation. The way the baby moves, how it dances, how it relates to others, its submission to its parents or to a tribal leader, how it learns to chip stone tools or build a fire are all patterned by the culture. The child's emotions and hormones, affections and rages, enthusiasm for the hunt, fear of snakes, and so on are all infused with the culture's meanings. The child does not take on its group identity as if it were a pre-existing self secondarily entering a framework of team spirit. Its becoming an individual, and an individual with a particular identity, is a cultural activity. Tribal individuals are cultural from their very beginning.

Note how this conceptual model has changed the question. Groups, whether family groups or non-kin groups, pre-exist the individual who is born and matures within this context. We no longer have to ask how or why an individual might join a group as from the outside. Our question must be how individuals' behaviour gets patterned by the culture and how each

hominin comes to learn, from the inside, the traditional norms of the group. That is, how do they develop their group identity?

The most likely means are imitation, emotional contagion, music and dance. Let us look at each of these.

Imitation

Imitation is the most obvious way that historical patterns of behaviour can be passed on. Unlike explicit instruction, it doesn't require developed language. Non-human primates already use imitation. In some areas of West Africa, for instance, chimpanzees use rocks to break nuts open. Others in the area imitate them and so form a local cultural pattern, while chimps further away do not practice the same behaviour.

In exploring imitation, we need to distinguish two modes of copying behaviour. When chimps observe others using a new technique, they *emulate* the relevant, effective aspect of the technique and implement it in their own, possibly superior, manner. Humans, on the other hand, seem to *imitate* the exact movements that they observe in a slavish, ritualistic way. In contemporary studies, after observing a model performing a task, human children are more likely than chimps to copy not just the causally relevant behaviours, but other irrelevant movements by the model—"over-imitation." They copy additional features, even when they have been told and understand that these are irrelevant. For example, they may use exactly the same finger as the model, even though they realize that any finger would achieve the result equally well. Surprisingly, five-year olds are even more likely to "over-copy" in this "high fidelity" way than three-year olds.¹⁸

If early hominins had this same tendency to over-imitate, it might help to explain their greater capacity than other primates to pass on traditional culture. Tool-making, for instance, is not passed on as an abstract idea, an intellectual inheritance: it is a highly ritualistic series of bodily movements. Hominins do not just copy the behaviours they see; they learn the ritual norms that are involved. They learn that there is a right and a wrong way of doing things. Since children imitate their parents and others around them in this way, over the course of generations everyone in the group adopts the same standard for these behaviours. This "way things are to be done" then becomes common to all members of the group. In other groups,

individuals may imitate alternative ways of doing things, so over time distinct cultural groups evolve.

Exact ritualistic imitation of bodily movement, in contrast to emulating—and perhaps adapting—a technique, encourages stability. The Early Paleolithic Acheulean tradition of stone tool-making, often referred to as “hand axes,” was maintained with little change from 1.7 million years ago until about 300,000 years ago.¹⁹ Manufacturing these tools was a complicated, multi-step process, yet they were made in much the same way and with the same materials for well over a million years.²⁰ In a world where explicit instruction could not be relied upon, ritualistic imitation of the bodily movements involved in the tool-making is the most likely explanation for this stability over about 50,000 generations.

Emotional contagion

This ritualistic transmission of cultural tradition has an emotional and neurological basis. Neural networks that are activated for particular behaviours are also partly activated by seeing those behaviours in others—the phenomenon of mirror neurons. When a primate raises an arm, onlookers have some tendency to raise their arms. If I fold my arms, the person I'm speaking to will often folds theirs too. This is not a deliberate or rational action: neither party may notice their actions. Rather, there is a kind of bodily contagion of behaviour. Such contagion may explain how group members, especially children, inherit norms of behaviour.

Inheritance by contagion is not confined to pragmatic activity, but includes emotional responses. When one individual expresses anger, an observer “understands” that feeling because their own equivalent neural processes partly fire in synchrony. An individual's grimace of pain elicits an empathic experience of pain in those who see it. A mother's startle at the sight of a snake infects her baby's reaction, who learns by such events the right way to respond to snakes. A child may come to feel safe and comfortable in a cave because her family group or tribe exhibits these emotions when they gather there. The norms for such emotional responses—not just for overt behaviours—are learned from those around them and contribute to the sense of belonging to the group.

It is tempting to say that group identity is an emotional phenomenon, and indeed I will often phrase it that way. “Emotion,” however, is a very imprecise notion. Especially in a self-

conscious, intellectualist culture such as our own, the word “emotion” is frequently used loosely in a negative way to refer to any non-rational, non-deliberative behaviour. In that sense, saying that group identity is emotional tells us little else than that it is not a matter of explicit thought and decision. If I am to say something more positive, I need to clarify what I mean by emotion.

“Emotion” sometimes refers to a fleeting feeling, as when the media refers to someone who weeps as “becoming emotional.” But when I say that group identity is emotional, I am using the term in a different sense. It is not a fleeting event, but an enduring disposition to act and feel in specific ways. Just as a fan's emotional identification with their team lasts for years, though it may only be expressed episodically during games, so the identification of a tribe member with their tribe is a life-long state, though it manifests itself in the distinctive ways they feel loving towards the tribe's babies, feel joyful over a successful foraging expedition, or feel grief over a tribe member's death.

Some emotions are primary and genetic, others are secondary and learned. Non-human primates, like other animals, respond to dangerous situations by becoming fearful or reacting angrily. Human groups go beyond such genetic reactions by inculcating secondary (sometimes called “pro-social”) emotions: birth leads to rejoicing; grieving follows death; defending one's tribe, even to sacrificing one's own life, is glorious; a team's scoring provokes enthusiasm. Children are trained to feel in culturally appropriate ways. As adults, they should feel respect for their elders, they should be wary of strangers, they should maintain solidarity with the group, and so on.

A social emotion may have a biological foundation. The initial basis for empathy with another's joy or pain may be the firing of mirror neurons, but this biological response can be emphasized or downplayed by a culture. Oxytocin—the “love” hormone—may be the biological basis that binds sexual couples, and nursing mothers to infants. It may also contribute to group identity and to the ingroup-outgroup dichotomy so distinctive of humans.²¹ The tendency of unrelated adults within a group to find babies cute is probably a cultural elaboration of this biological process and helps to protect the children of the group. Tribes at war with another group have no such inclination to protect the other group's babies. Developing the emotion of love towards the children of a group is one of the tasks of group

cooperation: non-family members must learn that they should protect their tribe's children. It is the sense that they are *our* children that makes it possible for parents to rely on others in the tribe to provide food, fire, shelter, and protection for their children. The tribe's nurturing of its offspring is a cultural elaboration of the biological bonds of parenthood.

Some social emotions, such as hostility to a foreign tribe, may be shared simultaneously by many members of the society. However, even secondary emotions expressed by a tribe member who happens to be alone are not purely individual phenomena: in so far as they are socially learned, they have a communal basis. A tribe member who feels protective towards a crying baby of another member is expressing an emotion that is social in the sense that it is what the culture expects, even though it may not be shared by anyone else at that particular moment.

Social emotions express group identity. One feels pride in the achievements of the group, or even of other individuals in the group. Consider the pride or disappointment of contemporary sports fans who have themselves made no contribution to their team's activities. It is as if the fans felt they "belonged" to the team. One can feel shame over the actions of family members, or of one's country's activities, even if, as an individual, one had no control over the events.

Before the arrival of social contracts, autonomous selves and explicit laws, social emotions likely contributed to the unity of a group, provided we understand emotions not as fleeting reactions, but in the precise sense of long-term dispositions of bodily behaviour and feelings.

Music and Dance

Apart from imitation and emotional bonding, Dunbar proposes other means of enhancing cooperation and group identity. He claims that, among non-human primates, grooming releases endorphins, so those who regularly groom each other form an emotional bond and are more likely to cooperate with each other in other activities.²² Keeping track of such grooming relations, however, is time-consuming and the time needed increases exponentially with the size of the group. According to his "Social Brain" hypothesis mentioned above, hominin brains need to be larger to manage more extensive social relations. Brains, however, are energy-intensive, so as they get bigger more time has to be spent on food

gathering. Hence, as group size increases the time available for grooming is reduced rather than increased. Despite more efficient means of nutrition, such as moving from a vegetarian to an omnivore diet and inventing cooking, there would not be enough hours in the day for both food-gathering and the necessary grooming. Even in *Homo erectus*, with an intermediate brain size, Dunbar calculates the time-budgeting involved and concludes that means of social bonding other than grooming would have had to be found if such larger groups were to be held together.

Dunbar suggests that communal music and dancing developed early as one way for bonding to occur in larger groups, perhaps as early as *Homo erectus*.²³ Direct evidence for this speculation is hard to find. Indirect evidence, however, comes from Dunbar's psychological investigations which show that making music together releases endorphins in modern humans and that the endorphin release enhances social cooperation.²⁴ Other studies find that pre-linguistic infants pay rapt attention to songs, regardless of the linguistic content or the identity of the person singing.²⁵ There is evidence that contemporary adults who actively make music together and move in synchrony are more likely to cooperate on other tasks. Music is almost invariably associated with clan gathering or national celebrations. Sports teams and religious groups typically rely on music to incite a sense of unity. Some contemporary hunting and gathering cultures engage in rituals that involve drumming and dancing with strong communal connotations. While all this evidence is based on modern humans, the universal presence of music among humans offers some grounds for thinking that music may have contributed to group identity from a very early period.

But what kind of music? There is no paleontological evidence for pitch instruments earlier than 40,000 years ago. However, the ability to move in time to a rhythm, such as a drum-beat, – that is, to dance – is universal among contemporary humans. Among primates, moving to a rhythm is unique to humans, so it may have evolved early enough to contribute to pre-linguistic group identity.²⁶

The question of whether musicality in this restricted sense is of genetic or cultural origin is not of great importance for our purposes. As long as the trait is reliably inherited, then early human groups that moved together to a rhythm, that danced together, could have

developed the strong emotional identification needed to overcome individual competitiveness and the temptation to free-load.

Religion

Religion may also have played a role in group cohesion. Religion creates a strong sense of belonging to a community, of helping coreligionists, and of defending them from outside attacks. Even today, when, in the ultimate form of cooperating with their group, soldiers are willing to sacrifice their lives, they often do so “for God and country.”

As always, however, we need to avoid parachuting the contemporary experience of religion into its origin. The rise of world religions and theological claims about a single transcendent god in whom individuals voluntarily express faith is associated with explicit language and even writing, so they are likely to be of relatively recent origin, probably after the advent of agriculture. Attributing such a phenomenon to early hominins would be anachronistic. We need a very different image of religion if we are to speculate on its role among early hominins.

According to Gauchet, earlier humans experienced the world as having an intrinsic sacred order within which all, whether human or caribou, whether chief or follower, must unquestioningly live. Humans are but one part of an order of nature which has its own structure and values. He claims that the idea of the sacred being projected beyond the world into a transcendent realm is a later, probably neolithic, development that desacralizes the everyday world.²⁷

Contemporary hunter-gatherer societies often take an attitude of sacredness towards the natural world, a sense of living in a symbiotic relationship with those parts of nature on which they depend. Tribes that hunt caribou, for instance, treat the caribou as sacred; they must be revered if the hunt is to be successful. In early “religion” certain people, places, times, species, rivers, trees, or mountains have special cultural meanings—some may be taboo or sacred—and one should behave towards them in special, culturally defined ways.

Shamanistic rituals are such sacred times that can involve some special adepts—shamans—wearing masks and costumes by which they become a kind of incarnation of an animal, perhaps a bear or a wolf. Using music, dance, drugs or other techniques, shamans

may enter a trance in which they merge with the animal, or with the spirit of the animals. During a trance, a shaman just *is* that animal. He or she may hallucinate the spirit of the caribou and receive revelations about where they are waiting to be hunted.

In some cases, apart from the shaman, other tribe members who participate in such rituals may identify themselves with an animal or other symbol, their “totem.” The totem comes to symbolize the unity of the tribal group, so that in identifying with the totem, they strengthen their identify with the tribe. In such rituals, group identity is expressed not by direct reference to the group in the abstract but by each individual identifying with a symbol that represents the group. Even today, many sports teams and military units pick an animal name or a mascot to reinforce group unity.

There are, of course, rituals today that involve groups gathering with music, dance, and drugs in which individuals “lose” themselves and submerge their individual identity into that of the group: some rock concerts, for instance. Military groups today use uniforms, drilling, singing anthems together, and so on to suppress individual autonomy and inspire soldiers to die—and to kill—for their group. Intellectual lectures about regulations won't hack it!

While such experiences may give us some insight into the role of ritual in early group identity, we must be aware of the inadequacy of our current experiences to capture what may have been happening among early humans. My speculation is that early “religious” rituals assisted in the cohesion of cooperating groups but that, unlike the contemporary examples, early hominin cooperators were not yet selves who had to relinquish their selfhood in such rituals. Selfhood, I claim, comes only later in evolutionary history.

Group rituals in which individuals synchronize their dance movements with others may be another instance of over-imitation, and may in turn reinforce this phenomenon. Whether we should call such rituals “religious” or not is arbitrary. The point is that, when one suggests that religion might have been a factor in holding early hominins together in groups, the suggestion depends on a notion of “religion” that is very different from our contemporary idea of religion as based on doctrines, biblical language and voluntary faith.

To summarize: my conjecture is that early human societies did not come about by individuals, or Tomasello's pairs, entering into some kind of rational agreements or contracts

between self-interested individuals. Instead, small family groups of primates evolved into larger hominin tribes that overcame the divisiveness of individual competitive fitness by relying on imitation, music, dance, totemic symbols, and religion-like bodily rituals. These developments created traditions of strong group identities that facilitated working together.

This account of early music/dance, and of religion/ritual is highly speculative. It is a “Just So” story, not a report on paleontological investigations. Empirical evidence to confirm it is meagre. Even the categories are hypothetical: while in today's world we distinguish rhythm from melody, music from dance, religion from ritual or trance, it is unlikely that these were distinct phenomena in early hominin evolution. More likely there were traditional group activities, different in different cultures, that preserved and reinforced individuals' identity with the tribe. Only in contemporary thought do we conceptually separate the later development of such phenomena into different categories of behaviour. What I am imagining are repeated communal events with drumming, animal imagery, decorative clothing, individuals moving and making sounds in synchrony, trancelike states, and so on. Labelling these activities as music, dance or religious ritual is anachronistic: they are distinct only in retrospect.

The aim of this discussion has been to explore how early hominins, by developing cultural traditions, could have constructed group identities even without relying on language and rationality. I want to provide an alternative image of cooperative groups that does not depend on prior autonomous individuals deliberately contracting to work together for a conscious goal. That would be to presuppose the kind of selfhood we are trying to explain. It would illegitimately project the current stage of human existence back into the evolutionary conditions that led to it.

The Context of Group Norms

The formation of hominin groups involves the emergence of a new kind of evolutionary unity. Just as multicellular organisms require that component cells play the roles needed to maintain the organism as a unit, so within tribal groups individual hominins have roles that they need to fulfill if the group is to operate and evolve as a unit.

The analogy has many limitations, but it does illustrate the relationship of an evolutionary unity to its components. I have claimed that in early hominin history it is group identity that brings about the cohesion needed for a tribe to face its evolutionary environment

with one unified destiny. For a tribal group to be a unity of evolutionary selection, however, individual tribal members must work together for the common good. This does not mean that each individual has to perform the same role. One of the advantages of a cooperative group is division of labour: within the group individuals take on roles as tool-maker, fire-tender, parent, shaman, season-tracker, and so on. Skills can be honed and accumulated over generations because tribal members can rely on others to fulfill the specialized tasks required for community life. That is the point of “self-domestication:” no individual needs to master all the skills needed for survival.

The culture as a whole not only defines the roles of individual members, it also determines what will count as succeeding or failing in carrying them out. Analogously to the way the food detector derives its functionality from the amoeba as a whole, the heart from the body, and the antifreeze from the codfish, it is the cultural context that sets down what behaviour is appropriate for the fire-tender, tool-maker and so on. If, while others are on a hunt, the fire-tender lets the fire go out, that is a failure of their role; it is inappropriate within this social structure. Just as the amoeba-system is the context within which there are norms that define biological functions, so a cultural group is the context within which there are norms for tribal roles, norms that define which behaviours count as successes or which are failures.

Individuals within each culture must learn the specific techniques of their crafts—how to build a fire, make a tool, gather roots or nuts, and so on. But, on a second level—that is, a “deutero” level—they must have the general capacity to be bound by norms. From infancy, they develop the habit of conforming their bodily behaviour and emotional responses to align with the patterns of those around them. They learn that some behaviours are appropriate for them and others not.

Early cultural norms should not be thought of as explicit rules that individuals grasp intellectually. We need to avoid any suggestion that tribe members are already selves equipped with an inner reflective life who introject norms exterior to them. This would be the mistake of importing later concepts into our history of origins. In this early group setting, it is not yet a matter of autonomous persons choosing to follow, or to disobey, rules. What we have is unreflective conformity, not freedom or choice. Individuals do not learn, for instance, an explicit rule that the tribe's children are to be protected. Rather, one is attracted to the

tribe's children; one learns to see them as cute; one loves them and feels like protecting them. Similarly, it is not that incest is prohibited by some rule; rather, one's sexuality has been patterned in such a way that parents are sexually unattractive. Eating the flesh of a fellow tribe member may be taboo, so one finds it disgusting. These norms get built into one's behaviours, hormonal responses, and feelings.

These norms constitute group identity and are maintained by the community. Just as cancerous cells must be eliminated from an organism, an individual who doesn't, or can't, follow the appropriate norms is not identified as a group member: she is not "one of us." She cannot be counted on to tend the fire, to share the dangers of the hunt, or to protect the group's children. If emotional expressions of disgust or shunning are not enough to enforce the norms, the group may leave her out of their projects and get along without her. In the final analysis, others will not work together with freeloaders; that is, those who do not follow the norms will be punished, exiled from the cooperative group, or even executed. The group will therefore end up with members who, for the most part, follow the norms. If the norms are no longer being learned or enforced, then in fact these individuals no longer constitute a cooperative group. This is a logical point: if most individuals violate the norms most of the time, then they are not working together. While communities vary in how they handle freeloaders, every group has to find some strategy for enforcing its norms and maintaining cooperation. A group that fails to do so is no longer a group that faces selective pressure as one unified entity.

Normative status

Following group norms is not an add-on, imposed on or voluntarily accepted by an individual who is already a group member. What the individual *is* is a biological organism whose behaviour—including brain, hormonal and, emotional behaviour—conforms to the group's norms. Being a member of the tribe is not just being a biological offspring of a group member; it is not just being in a particular place or time; it is not just being associated with others in the vicinity. It is about the very nature of the individual hominin. Membership is the state of being bound by group norms. Individuals who are not bound in this way do not qualify as members of the tribe. Group identity determines what it is to be a tribe member in the first place.

Being a tribe member is a contextual status: without the tribe, the biological organism would not qualify as an early human. Even if we could artificially, in a lab, grow an isolated organism with hominin DNA, it would not be a human individual. On the other hand, if in some science fiction scenario we had an artificial intelligence or robot that could live up to tribal norms, it would be a member of the tribe, that is, a human, even though it would not be a biological organism.

Unlike our contemporary medical scenario of artificial hearts, no alternative mechanism—such as an Artificial Intelligence—was historically available to evolution, so the primate biological organism was the only option for an entity that could become a tribe member. This historical fact, however, must not mislead us into thinking that humanity is defined by the biological organism. Being a human is more than the body: it is a normative status that depends on the context of the group.

This normative analysis parallels my functional account of biological organisms, such as my hypothetical amoeba or the homeostatic cell. It is not its physical mechanism that qualifies an amoeba as an amoeba: it is its role in the amoeba system that defined what it is. Similarly, what qualifies an individual as a member of a tribe is the set of roles they are supposed to play in the group context. Recall my analogy with money. It is not the metal that qualifies a coin as money: only its role in the exchange system defines the coin as a dollar. A tribe member, of course, must have a biological body, but it is not their body that defines their group identity. It is because they are bound by the group's norms that they are who they are. If they were not bound by these norms, then they would no more be a member of the tribe than a disk of metal transported to another planet would be a dollar. A tribe member is a normative entity that is what it is only in the context of an evolved hominin cooperative group. (The norms maketh the man!)

This is not just a practical point. It is about the very nature of the individual hominin. It is not just that an individual's ability to survive and reproduce happens to depend on the tribe; their status as a human being has been outsourced to the group. Their being is contextual: without the tribe, the biological organism would not qualify as a human.

Cultural World/Umwelt

A group's cultural tradition not only defines the nature of tribe members, it patterns the world they live in. It categorizes the repertoire of actions available to them and gives meaning to the things they perceive.

First, consider actions. All animals, even amoebas as we've seen, perform actions, in the sense of behaviours which are functional for the organism in its Umwelt. For purely biological organisms, the set of actions available is determined by the genetic inheritance of the species. Hominin groups, in addition, inherit culturally defined repertoires of possible actions from their traditions. One can make a stone axe, dance in a ritual, tend the campfire, join a team to hunt gazelles and so on. Such actions are not available to organisms with purely genetic inheritance: the very possibility of performing such actions depends on the tradition of the group into which a tribe member has been encultured.

It is the traditional norms, inherited from ancestors—possibly from the gods in mythological time—that define actions. The cultural norms lay down how an individual has to behave in order to qualify as making a stone axe, for instance. One can chip at a stone randomly, but such behaviour does not count as the action of making a stone axe. It is the cultural norm that unifies a series of bodily movements into one single action and categorizes this unity as of a particular kind.

Secondly, traditionally defined actions bring with them categorized entities in the Umwelt. That is, in patterning action, the culture also pragmatically categorizes a world of things common to members of the group and which are perceived and acted upon by individuals.

The kinds of things in the Umwelt of an amoeba—such as food-for-amoebas in my example—depend solely on its species, that is, on its genetic inheritance. The genetic inheritance of hominins also contributes to the meaning of things in their group's world, but the tribe's traditional actions define new kinds of things. Consider a new technological thing—the stone axe in my illustration. A particularly shaped piece of stone is not an axe-in-itself. It is not the stone material or the shape that makes it a tool. Its significance as an axe requires the context of individuals with the skills to use it and established norms for what it should to be used for.

Such significance should not be interpreted in an intellectualist manner. That would be anachronistic. The thing is an axe not because it is thought of, understood, or represented by individuals in a particular way: it is how the axe gets used, or rather, how it *should* get used, that constitutes its “axeness.” There is no need to presuppose symbolic thought such as the word “axe” or some sign that refers to an axe. It is the role that it plays in the hominins' world that qualifies it as an axe.

Many culturally defined entities are social things. Already among non-human primates, individuals perceive others as dominant or subordinate, but in early human cooperating groups, individuals learn to recognize the many different roles that individuals play in the group. That is, members of the tribe get categorized into fire-tender, tool-maker, butcher, or baker, and so on. As part of their upbringing, individuals learn to perceive things such as camp fires, ritual dances, food-gathering parties, hunt-leaders, mates, and shamans. Crucially, on the basis of appearance, habits, decorations, tattoos, or even walking styles, other hominins are perceived as either members of the cooperating group or as foreigners with strange customs that belong to other groups. When strangers are encountered they will immediately be perceived as “them” rather than us. That's part of what it is to have a group identity.

It is less obvious, but also true, that the meanings of natural things depend on the culture. The things of nature that people perceive and interact with at this evolutionary stage are not defined by universal, scientific concepts, but are components of the group's Umwelt. Animals are not classified by Linnaeus' genera and species, but on a pragmatic basis: those that we eat, those we hunt in Spring or those that we avoid because they are dangerous. Places are not points on a map or defined by geographic parameters, but are perceived as safe caves in which we shelter, rivers in which we fish, areas where we over-winter or where we live in summer. Explicit language is not needed for individuals to perceive things in these ways: tribe members learn from their group's practices what kinds of things there are and how to recognize them. These are not things defined anachronistically by scientific theories. They are things-for-the culture: Umwelt-things.

Things, whether social or natural, are not value-neutral like scientific objects; they are normatively defined. As pragmatic entities, they are related to action and so imbued with

meaning for agents. Nature is made up of scary forests that should provoke alertness, juicy berries-to-be-picked, dangerous swamps-to-be circumvented, prey animals-to-be-caught and so on. We should not dismiss such values as “mere” emotions as if early humans live in a scientific objective world to which they incidentally have subjective reactions. It is the things themselves, as defined by their pragmatic significance, that are value-laden. Cultured individuals live in normative situations in which their reactions are appropriate—or inappropriate—to the things in their world. A cave is not a geological structure: it is a place where one should feel safe. If one gets scared by a safe cave, one is responding inappropriately; one has violated one of the tribal norms. Hunt-leaders are real things within the cultural situation, and are to be obeyed. Being submissive toward a leader is not only appropriate; the norm is an integral part of the individual having the role of leader. If there were no such norm, the biological entity would not be a leader. Within the tribal context, a leader has the value “to-be-obeyed,” and has the role of leader only in so far as others follow. Umwelt-things, how they are perceived, the perceiver's hormonal and cognitive responses, and the actions that are supposed to follow are not subjective add-ons; the norms are part of the very essence of such entities.

This is not individualistic. It is not the individual that determines the significance of objects and sets the norms for how one should respond to them. It is not up to individuals to determine what a thing is or is not. I am not defending a subjective relativism. The norms for objectivity are communal. Individual amoeba do not determine what is food-for-amoebas: it is the species that establishes that meaning. Similarly, that a campfire is supposed to be kept burning is not due to the feelings of the individual fire-tender. It is the communal norms that determine the nature of a campfire, and the particular ways that members of the tribe should respond to it. If an individual doesn't perceive a campfire that way, they are not, in this respect, a member of the tribe. To a chimp, a wildfire is not a thing that it is bound to tend. There is no such thing as a campfire for a chimp. It is only in the context of a hominin community and its group norms that there are campfires that should be tended and that there are roles for fire-tenders bound to tend them.

The things in a culture's world are correlated with the tribe member's capacities for perception and action. Indeed, these are two sides of the same coin. An individual's actions and perceptions are about things as categorized by the group's tradition: what it is to a thing

in the tribe's Umwelt is for it to be perceivable by group members and to be the target of possible actions. A hominin group is an emergent entity that establishes the context within which, on the one hand, members of the tribe learn the norms for how to perceive and to act, and, on the other hand, the things in the tribe's world get defined by their roles for the group. Individuals and their Umwelt are systemically correlated: each is what it is only in terms of the other. In this sense, each early hominin group forms a system that sets the context for its components.

The group's present context, however, is also its historical context. The group's roles, actions, and Umwelt-things are inherited from ancestors. The way tribe members *should* live is the way that has been handed down from ancestors—actual, mythical or divine. Any tribe that does not mostly conserve the traditional way of life reduces its chances of survival. Just as the species' evolutionary history defines the functional structure of a biological organism's life-way, so the hominin group's evolutionary experience defines the norms for how the contemporary group members should live. How a group actually behaves is a factual matter, but the norms that define how the members *should* dance or *should* make axes are set by the tradition. The context that determines the normative definitions of actions and things is ultimately a temporal context – the context of the ancestors' cumulative tradition.

Conclusion

In this chapter, I have conjectured that between the purely biological stage of evolution based on genetics, exemplified by my toy amoeba, and the arrival of modern selves, there is an intermediate stage based on learned, traditional ways of survival, that is, on culture. Such traditional lifeways are preserved and transmitted by a variety of small hominin cooperative groups. Within each group individuals conform to tribal norms. Each cultural group forms a system with individuals whose traditional perceptual, actional and emotional capacities are correlated with the things, social and natural—the Umwelt—that these capacities are about. Each group and its world form an emergent evolutionary system analogous to the system of the amoeba and its Umwelt discussed in chapters 1 and 2.

In the case of the amoeba scenario, what makes something an organism is not just its material construction, but its functional, biological role in the system. In the case of early hominins, what makes a primate organism a member of a tribe, that is, an early human, is its role in the tribe. What differentiates a hominin from other primates is not primarily some intrinsic characteristic such as their larger brains—though of course their brains must be complex enough for them to carry out their tribal roles. Rather, what defines a hominin individual is a relational attribute: their being bound by the norms of the group.

The point of this chapter has been to explore how such early humans could have come into being. I have invoked emotion, music, religion, and other ways of creating group identity to offer a story that makes it plausible that cooperating groups could have survived prior to the evolution of rational, free, contracting selves. Even without relying on language, such a group could set appropriate standards of behaviour that individuals, for the most part, followed. At this stage, it is not that individuals freely or consciously follow an obligation to obey the standards of the group. What we have is conformism: individuals imitated the historically developed ways of life into which they were born. Those who didn't or couldn't conform to a group identity were not “human.”

In our journey from purely material, causal processes towards selfhood, we have now moved from systems based on biological functions that differentiate between what is normal and what is abnormal, to the emergence of social systems based on learning and tradition. Group norms have taken us beyond biological functionality and cultural identity has transcended species membership. The emergent system, nevertheless, retains the essential features of any system: the overall traditional pattern defines the behaviours within it and the standards of appropriateness for them; it constitutes individuals within each culture and determines their group identity; and it defines and gives meaning to the things of their world, natural and institutional, among which these individuals live.²⁸

The structure of the new emergent cultural system is Janus-faced. It looks back to the more primitive, biological system from which it evolves, but also looks forward to the narrative system within which selfhood can come into being. In this gradualist theory, however, there is yet another step needed before autonomous selves can be possible: the evolution of language and its norms.

Chapter 4

Language

Language is essential for selfhood. As I will argue in the next chapter, selves depend on narrative and without language there can be no narrative. But what exactly is language, and how could it have evolved in human prehistory?

Defining Language

Language is, first and foremost, a mode of communication. Communication is a transfer of information, but not all transfers qualify as communication. Light from one ancient galaxy will have reached some other galaxy billions of years ago, but while these photons carried information, I do not count such purely accidental, causal transfers as communication. Perhaps we might label them “signals.” I will reserve the term “communication” for information transfers that are functional in the sense in which I have defined function in chapter 1: a process has a biological function in so far as its existence is due to its having conferred evolutionary fitness on the ancestors of the current organism. It is in this technical, functional sense of communication that I claim that language is a mode of communication.

Consider, for example, the claim by Gorzelak and colleagues that some trees “communicate.” When attacked by some pathogens, such a tree stimulates processes in fungi that travel through the soil and get picked up by other trees in the vicinity. This gives the other trees advance warning of the pathogen and results in their immune system preparing a defence in advance of becoming infected. Is calling this communication simply a metaphoric mode of expression, a kind of anthropomorphic projection onto trees, or should we consider it “real” communication? As I am defining the term, this is an empirical issue: we have to investigate whether in fact the current existence of this mechanism of information transfer has evolved because it increased the fitness of ancestral trees. Since the indications from Gorzelak is that this is the case, then this transfer of information is indeed functional, and so the process qualifies as “communication” in the way I am defining it.²⁹

“Language,” as I will use the term, is one kind of communication; it is unwise to use the term for all forms of communication. (Gorzelak, wisely, does not describe the communication of trees as language.) Calling all communication language obscures one of the most distinctive features of human communication: it is governed by cultural norms. The roar of a lion is a mode of communication – it may communicate territorial claims, for instance – but the functional norms for the roar's success or failure are purely biological. Language may build on biological grunts or cries, but linguistic behaviours are actions defined by the group. Like all cultural actions, linguistic acts are appropriate or inappropriate in accordance with the culture's norms. Linguistic communication need not be vocal; indeed it is possible that early language was primarily gestural. Overemphasis on the features unique to contemporary language, such as writing or syntax, must not blind us to the evolutionary continuity of language with other cultural achievements. Language is a cultural phenomenon: it is a mode of behaviour learned by individuals as an element in the group's ancestral tradition.

Language is important for the evolution of human cooperation. It increases cooperation within the language group in at least three ways. First it enriches communication between current group members. Secondly, language facilitates the transmission of information across generations. Thirdly, each distinctive language differentiates its speakers from foreigners and reinforces group identity.

Language, seen in retrospect, may look like a dramatic break from emotional communication in that language expresses meaning. But this would be an oversimplification. Emotional communication already transfers information about the world, about things, individuals, and actions. Language does, however, enrich the information transfer and makes it more precise. Planning a gathering expedition, deciding who will cooperate with whom and where today's foraging will take place, assigning some to searching, others to digging, still others to transporting the spoils, deciding who will commit to keeping the fire burning, and so on, can all be managed much more efficiently by language rather than by communication limited to purely emotional expressions.

Language also enhances group cooperation across generations. Once one generation has discovered that seals arrive on ice floes in the Spring, this information can be

etched into tribal memory by stories or chants so that traditional knowledge can guide the hunting of future generations. Language allows much higher fidelity transmission of ancestral survival skills than imitation, emotional contagion or other techniques.

Language also contributes to the distinctiveness of the group. Even today, linguistic styles signify social class; cultural subgroups create argots to differentiate themselves from the cultural mainstream; and children use distinctive linguistic styles to define themselves in opposition to parents. Local regions elaborate their own dialects not simply because the language of isolated groups drifts randomly away from the mother language, but also because distinctive ways of speaking enhance the sense of belonging to the local group. We can presume that in early linguistic evolution, each language is not only a product of the group's culture, but also contributed to the group's cultural identity.

Language, however, does not only enhance cooperation, it is itself a cooperative enterprise. The speaker does whatever they can to encourage the hearer's correct interpretation of the speaker's utterance. The hearer, in turn, takes the speaker's sounds to be carrying information of value to the hearer, and in interpreting them, assumes that the speaker is trying to pass on information valuable to the hearer. Such communication is more effective if both speaker and hearer follow common norms for the exchange, that is, if they speak the same language.

A group that has common norms for the exchange of information is likely to be more successful in group projects, to be better able to accumulate traditional wisdom, and to build tighter group identity; so for these reasons, the development of language would have increased group cooperation and therefore the evolutionary fitness of the group.

Emotion and Music

How could language, in this sense, have evolved? Language as we experience it today, especially in academic, scientific or philosophical contexts, serves to communicate truths about the world: it expresses verbal propositions that could be objectively true or false. Projecting this image of language onto early hominins could be misleading. It is more likely that language began as an elaboration of bodily expressions of emotion, and attempts to influence the emotions of others. Just as a mother may coo at a baby to sooth it and lead it from fear to joy, so early language expressed emotion and triggered emotion in others. One

individual's anger might elicit fear in a targeted individual, or provoke mirrored anger in companions. Initially, such communication was primarily biological, but as it became patterned by cultural traditions it gradually developed into “language” in the sense in which I've defined the term above. In so far as it became governed by group norms, gestural expression—“body language”—already qualified as language.

Music and dance may have been influential in imposing group norms on both body language and vocal language. Morley points out that making sounds is just as much a bodily activity as gesture or dance.

Manual gesture, body language (corporeal gesture), facial expression (facial gesture), and vocalization (vocal gesture) should all be considered as elements of a single system of gestural expression and perception of emotional state. ... As well as being able to observe such cues in others, we can empathically experience something of their emotional state in mirroring them with our own bodies.³⁰

He claims that in a number of contemporary hunter-gatherer societies, group rituals involving rhythm and dance also rely on vocalization. Although such contemporary groups, of course, have language, their singing does not always depend on words, but often involves vocables, that is, word-like sounds that are without semantic meaning. (Gaelic liltng, or jazz scat singing, are examples of such communication by vocables). In early human evolution, Morley claims, those most skilled at this undifferentiated communicative ability would have had a selective advantage in cooperative, pair-bonding and parent-infant relationships. Only later would this ability differentiate into music, dance, and language: “This system of vocal and kinesthetic communication of emotion constituted the foundation for vocal communication out of which later emerged culturally-shaped melodic, rhythmic musical behaviors and semantic, lexical linguistic behaviors capabilities.”³¹

While the specialized linguistic structures of vocabulary and grammar allow the transmission of enormous quantities of meaning, in normal speech these occur in association with variations in total contour, rhythmicity, and intensity—the “music of speech”—which themselves transmit highly significant information. Such variations of expression —“prosody”—seem to have a far older evolutionary heritage than the actual meaning of the words and may be related to capacities possessed by our primate relatives. The cadences of linguistic expressions are probably a continuation of emotional communication.

Darwin, in the nineteenth century, already claimed that articulate speech was derived from the emotional expression of music.

All these facts with respect to music and impassioned speech become intelligible to a certain extent, if we may assume that musical tones and rhythm were used by our half-human ancestors, during the season of courtship, when animals of all kinds are excited not only by love, but by the strong passions of jealousy, rivalry, and triumph. From the deeply-laid principle of inherited associations, musical tones in this case would be likely to call up vaguely and indefinitely the strong emotions of a long-past age. As we have every reason to suppose that articulate speech is one of the latest, as it certainly is the highest, of the arts acquired by man, and as the instinctive power of producing musical notes and rhythms is developed low down in the animal series, it would be altogether opposed to the principle of evolution, if we were to admit that man's musical capacity has been developed from the tones used in impassioned speech. We must suppose that the rhythms and cadences of oratory are derived from previously developed musical powers.³²

As groups became larger and more sophisticated in their techniques, emotional expressions were too amorphous for organizing cooperative activities. Once behaviours became governed by cultural norms, communication about them had to evolve in parallel to be capable of referring to such behaviours. If a team member failed to defend another member from a dangerous animal they were hunting—as was their role—it is not enough for the endangered member to simply express anger or shun the perpetrator in later hunts. Simply yelling inarticulately at the offender lacked the precision needed to either express frustration that the action violated the group's hunting norms or to influence them to change their behaviour in the future. The expression of emotional reactions needed to be gradually supplemented by more precise, culturally normative communication that was capable of referring to things and events in the group's world.

Linguistic Meaning

With the arrival of culturally-shaped communication comes “linguistic meaning.” Pre-linguistic communication transfers information by means of biological and emotional reactions. The lion's roar induces terror; a shriek of pain produces empathy (probably via mirror neurons); cooing to a baby soothes it; threatening gestures lead to fear; a startled yell alerts the hearer to a snake; and so on. Language, however, communicates by means of its “linguistic meaning.” “There is a snake in the grass” warns the hearer only if they understand the meaning of the expression. If they speak a different language, the sounds are

meaningless to them. Unlike biological communication, the sounds themselves (or bodily movements in the case of gestural language, such as signing) are not the significant element. Analysis of the frequencies of the air vibrations produced by vocal cords would not tell us anything about a snake. In the case of language, unless we grasp the linguistic meaning conveyed by the sounds, we don't gain any information about the world. We don't know what the sounds are about. And, as we will see, without linguistic meaning there can be no selfhood.

What is this additional feature, "linguistic meaning," that makes sounds into language?³³ I will discuss three theories of linguistic meaning: Adamic, Mentalist and Contextual.

Adamic theory

First, one might think that words are meaningful in themselves. To a young child perhaps it is unquestionable that "dog" just means dog—the pet they are playing with. There is a Biblical story that God showed Adam each animal and whatever Adam called it, that became its name. Another Biblical story offers an extra wrinkle: when humans built the tower of Babel to challenge God, God made the builders speak different languages so that the resulting confusion led to the building being abandoned. In the seventeenth century, faced with the multiplicity of languages, some people—such as Leibniz—tried to find the original Adamic language to discover the Real Names of things, that is, the words that intrinsically refer to their objects.

This position assumes that each word has a meaning in and of itself: it can refer to an object in isolation from other words, from the context of a language, and from how a society uses the word. This account is implausible. The same sound can carry a different meaning in different languages. The Spanish sound "Sí," the French sound "Si," and the English "Sea," are the same sounds, but within their languages they have different meanings; they are different words. Without the larger context of a language a sound doesn't have the particular meaning it has; indeed, it doesn't have *any* meaning and doesn't even qualify as a word. The sound might be made by the wind in the trees.

Mentalism

A second theory of meaning rejects the notion that words have meaning in themselves. Instead, it holds that meaning is an idea that is mentally present in a person's mind. Since minds are private, in the sense that you don't have access to my thoughts, communication consists of me "expressing" my ideas in sounds so that you can "interpret" these words and so form similar ideas in your mind. The sounds of language in themselves are meaningless; they are just physical vibrations in the air between us that serve as vehicles. Only an accompanying mental event makes a linguistic expression meaningful. Sounds made by a parrot, or a computer, are meaningless gibberish since they are not accompanied by a mental idea (unless you attribute minds to parrots or computers.)

This second, mentalist, theory of linguistic meaning was held by Descartes and has become more or less the default theory of contemporary common sense. It has, however, many problems. Since it assumes minds are private, there is no way to know whether the idea you attach to a linguistic expression is the same as mine. For instance, when I say the word "red" its meaning is my mental idea of redness, but if you were colourblind the same word could have the meaning of your mental idea, which might be greyness. We might never find out that the meaning of the word is different for the two of us. In principle, all your accompanying ideas might be different from mine.

In *Alice in Wonderland*, Humpty Dumpty says, "There's glory for you." Alice can only find out that by "glory" Humpty Dumpty is expressing his idea "a knock-down argument" when Humpty Dumpty tells her that's what he means. But how can he tell her unless she already knows what he means by "knock-down argument?" You cannot *tell* me what you mean by an expression except by using other expressions that I must already understand. So we are faced with a vicious regress: unless I have access to your private idea that is the meaning of at least some of your expressions, you cannot tell me what your words mean.³⁴

In any case, removing intrinsic meaning from words and attributing it to mental ideas instead only postpones the problem. How do ideas have meaning? How does your "elephant" idea refer to an elephant? Maybe my "elephant" idea refers to a mouse. Moving the notion of meaning-in-itself from external words to internal ideas only delays and disguises the need for an account of meaning. (Besides, as we will see in the next chapter, the very notion of an internal private mind is itself dubious.) Even if, in a more recent version of the theory, one

holds that ideas are cognitive brain structures, we are owed an explanation for how one brain structure means mouse, while a different one means elephant.

As always, in offering an evolutionary account there is the recurring danger of projecting our current status as selves back into their prehistoric origins in a circular manner that explains nothing. Even if contemporary selves did experience words or ideas as having linguistic meaning in themselves, to assume that early hominins had similar private experiences that were at the origin of language is to presuppose the notion of linguistic selfhood and so is anachronistic.

Contextual meaning

The third, “contextual,” theory of linguistic meaning rejects the notion of meaning-in-itself, either as a property of words or of mental ideas. It rejects the appeal to a private mind and holds that meanings are public. The theory claims that linguistic meaning, in continuity with cultural meaning, originates in the context of cultural norms. The meaning of a linguistic expression is the set of inferences that the group holds the speaker to. When a speaker utters sounds that make up a linguistic action in accordance with the group's norms, the community expects the speaker to be committed to further actions, linguistic and otherwise. It is this set of normative expectations in the context of the group's language that is the “linguistic meaning” of the expression.

This is a more promising theory, but we need to unpack this abstract position by imagining how such contextual meaning could have evolved in some early hominin scenarios.

A tribe setting out on a hunt might use gesture and pantomime, such as pointing, to indicate that one member should stay and tend the campfire. Such gestures have a meaning: they are about the fire-tender, the fire and their relationship. If the form of such gestures has become routine and traditional within the group, then this already qualifies as “language”—albeit body language—according to how I defined the term above. Sounds might occasionally be made at the same time and if these sounds become habitual and get handed down to later generations, then we would have traditional “verbal language.” Other tribes may hand down different habits, and so we have the start of culturally distinct languages. What such sounds are about, that is, their “linguistic meaning,” depends on the traditional norms of each culture.

As another thought-experiment, imagine an apprentice trying to imitate the manufacture of a stone tool. The teacher might use emotional expressions of approval or disapproval, possibly biologically based. Approving a particular choice of material, or correcting a chipping angle would be more effective if different sounds were used for one or the other. If the same sounds were repeatedly used by the teacher, it is possible that the apprentice would use the same differential sounds when they, a generation later, became a teacher. Thus, a traditional set of sounds with stable meanings could be inherited by the group.

Once sounds for fire-tending or axe manufacture become traditional, there are norms for their correct use. Groups in which different teachers use different sounds for correcting behaviour lose the advantage of multiple sources of group tradition. The selective advantage of groups with a “common” language, would encourage the use of one traditional language within each group.

Like other traditional actions, communicative actions are governed by the normative expectations of a culture. Just as a hunt leader is supposed to act in a particular way, so a speaker is supposed to follow the group's linguistic norms. The cultural norm may be that someone who makes the sounds “Wolf coming” should also utter sounds such as “a small mammal is coming,” “run into the cave,” “It has no tusks,” and many other sounds. These are the “inferences” expected of anyone who utters the sounds “Wolf coming.” The expected inferences are not only linguistic: the speaker should express fear, run into their hut, or grab a weapon. Other group members who follow the communicative norms would expect to see a wolf appear. If a mammoth appears instead, then others would consider the speaker to have violated the group's norms. According to these norms, when a mammoth arrives, the sounds uttered by the speaker should be “Mammoth coming.”³⁵

Different sounds can have the same linguistic meaning. No two vocal sounds are ever physically exactly the same. Everyone has their own accent. Linguistic communication is not a matter of the physical sound, but of what the sound means. The set of sounds that the community treats as the same in practice, what it considers a “repetition” of previous sounds, are those that commit the speaker to the same implications. This is a normative point. Whether or not a speaker in fact lives up to the group's expectations, the communal norms for

what they *should* do or say next is what constitutes the linguistic meaning of the sounds. For instance, in the case of the sounds “There is a wolf coming,” “A wolf is arriving,” “A wolf is approaching,” the implications to which the speaker is committed are the same. That's what it is for these different sounds to have the same meaning.

Truth and Lies

Linguistic meaning is not the same as truth. A group may well have an expectation that some expressions—assertions—should be true, not false. Such a norm, however, is secondary to the norms that establish meaning. The sound of wind in the tree tops, or a reflexive cough, is not true, but neither is it false. Only sounds that already have a linguistic meaning can it be either true or false.

I am not making the implausible relativistic claim that cultural norms determine truth: it is how things are in the world that determines what is true or false. My position is that the communal norms determine what meaning sounds have. Only for meaningful sounds does the issue of truth or falsehood arise.

Much has been said about lying and about the evolutionary need for techniques to spot deception, but lying is possible only if both parties share the same community of meaning. I can only lie to you about where the gazelle is to be found if the term “gazelle” has the same meaning for both of us. If I were to try to deceive you about the *meaning* of the word gazelle, then all communication, including deceptive communication, would simply break down.

Mistakes and misunderstandings, of course, are possible. Others may hold a speaker committed to positions that he did not intend. Such error is not confined to linguistic actions. If the correct cultural action is to bow to the chief, a child could mistakenly look them in the eye. The action would be an affront, despite the fact that it is respect that the child intends. The cultural meaning of their action is not determined by their intention. Linguistic actions are similar. In trying to truthfully tell fellow hunters where the gazelle is, a speaker may mistakenly use the expression “near the river” when the gazelle is near the mountain. A mentalist might insist that the *real* meaning of the utterance is “near the mountain,” since that's what the speaker has “in mind.” But the meaning of the phrase is not theirs to determine. The meaning of the expression is determined by the community, not by the speaker's intention. Unlike

Humpty Dumpty, the speaker cannot make words mean whatever they want. Rather, they are just mistaken; they miscommunicate.

“Linguistic meaning” is a normative phenomenon: it is the set of expectations for group utterances. In learning the tribe's language, any tribe member has to master the network of expectations associated with each uttered sound. Given the group's norms, the speaker in the example above must mistakenly have been using the sounds “Wolf coming” with the linguistic meaning “Mammoth coming.” Even if the speaker had the idea of a mammoth in their mind, the sounds they uttered, “Wolf coming,” have a meaning in the group's language that is independent of their mental idea. Neither the sounds in themselves, nor any private mental experience of the speaker, determines linguistic meaning. It is *not* that the community holds speakers committed to certain inferences because of the meaning that words or mental ideas already have; it is the set of linguistic (and other) actions that the speaker should be committed to that gives sounds their linguistic meanings in the first place. It is the context of the group's norms that constitutes meaning.³⁶

If a tribal member consistently fails to follow the communicative norms of the group, then they do not understand its language; they are talking unintelligible gibberish, cannot be relied upon for further information and will have to be excluded from any project coordinated by language. Such sanctions enforce the linguistic norms that constitute a human, cultural language.

This contextual account of linguistic meaning dispenses with any appeal to “meaning-in-itself,” whether applied to sounds or to mental ideas. It explains why “Wolf coming” refers to a wolf, and “Mammoth coming” is about a mammoth. What a word refers to is a matter of use: the tribe's norms require that “Wolf” is the appropriate sound when there is a wolf. The sound itself has no intrinsic relationship to the named animal, nor is any accompanying mental idea at the origin of the “aboutness.” Meaning comes from the linguistic community's use, or, more precisely, from its norms for correct use.

Syntax

The distinctiveness of my approach to the evolution of language can be clarified by contrasting it with Noam Chomsky's theory, which is probably the most popular current approach to the nature of language. Since the 1950s Chomsky has insisted on a sharp

distinction between “semantics”—what I’ve been calling linguistic meaning—and “syntax,” that is, the grammar of sentences. “The cat caught the mouse” and “The mouse was caught by the cat” are semantically identical—they express the same meaning—but their syntax, their grammatical structure, is different. “The dog chased the rat” and “The wolf caught the deer” have the same syntax, but carry different meanings, that is, they have different semantic content. Chomsky holds that the essential and distinguishing characteristic of language is that it is syntactical.³⁷

It is humans' syntactical capacity that allows us to speak and understand an infinity of expressions. Once we've grasped “The dog chased the cat” we can immediately interpret “The wolf caught the deer” or “The elephant killed the tiger” or any number of other substitutions, provided we know what “elephant,” “tiger,” “deer,” and so on, mean. Syntax is infinitely generative of new sentences, sentences that anyone who has mastered the syntax can understand.

Syntax also gives us recursion. We understand “The dog chased (the cat that caught the mouse (that ate the cheese.)) Animals cannot do this; they lack recursion. Indeed, they lack all syntax, according to Chomsky.

Chomsky claims that syntax is innate in all human brains. His evidence for this position includes the claim that syntax is universal in all contemporary languages, and that it is impossible for children to obtain enough information to learn it—the “poverty of the stimulus” argument. As recently as 2014, he, and some of his followers, claim that there was a genetic mutation that enables these features and which occurred only once in human evolutionary history.

According to the[ir] ‘Strong Minimalist Thesis,’ the key distinguishing feature of language (and what evolutionary theory must explain) is hierarchical syntactic structure. The faculty of language is likely to have emerged quite recently in evolutionary terms, some 70,000–100,000 years ago, and does not seem to have undergone modification since then.³⁸

Chomsky devised his theory at a time—the 1950s and 60s—when the paradigm for computers was based largely on the manipulation of discrete symbols, and his conception of the brain's cognitive functions was modelled on this paradigm. The subsequent discovery of the learning capacities of neural networks cast doubt on this model of brain processing. In a further blow to the theory, current Artificial Intelligence speech recognition technology has

advanced rapidly, but only after it dropped the rule-based, syntactic approach in favour of statistical probability correlations. Given these newer advances, the poverty of the stimulus argument is weakened: if AI programs can manage without innate syntax, it is possible that human brains can too, so the assumption that language is universally supported by one, unique rule-governed cognitive process of manipulating discrete symbols is no longer theoretically necessary.³⁹

Apart from such theoretical problems, empirical evidence challenging Chomsky's approach has been accumulating. Evidence—contested by Chomskians—that the songs of some birds rely on syntactical structures, and the discovery that some monkeys communicate with cries that have inflectional endings implies that syntax is not the exclusive privilege of human language.⁴⁰ Other linguists claim to have found languages, most notably in the Amazon, that lack features such as recursion. Some studies of how children actually master language suggest that the poverty of the stimulus argument is uncertain. These findings put the universality of innate syntax into question.

My main disagreement with the syntax-first school, however, is their definition of “language.” Bolhuis, Chomsky, et al. insist that

In our view, for the purposes of scientific understanding, language should be understood as a particular computational cognitive system, implemented neurally, that cannot be equated with an excessively expansive notion of language as communication.⁴¹

They acknowledge that “The language faculty is often equated with ‘communication’—a trait that is shared by all animal species and possibly also by plants.”⁴² Their syntactic view of language is an explicit rejection of this position.

This syntactical understanding of language, however, has two problems. First, it is circular. Their claim that treating language as communication is “expansive” makes sense only because they have already stipulated that “language” is to be defined as syntactical.

Secondly, the position Bolhuis and Chomsky defend tells us little about *why* language evolved. It suggests that its appearance is due to a sudden, random genetic mutation. My claim is that language is functional: it evolved *because* of its communicative function. Bolhuis and Chomsky deny this: “Externalized language may be used for communication, but that particular function is largely irrelevant in this context. Thus, the origin of the language faculty

does not generally seem to be informed by considerations of the evolution of communication.”⁴³

I think this position of Bolhuis and Chomsky is mistaken. My view of language as culturally normative communication offers a more plausible evolutionary account.

Evolutionary explanation

Since the evidence is that the hominin strategy of cooperative groups is at least a million years old, perhaps even two million, some form of communication based on communal norms likely existed from this very early period. Syntax may well be a recent evolutionary development, as Chomsky maintains. My claim is that, even if this is true, language, in the way I am defining it as communication governed by cultural norms, has been around for much longer than syntax.

It would, however, have been a non-syntactical language. What would such a language look like? Each different expression would have to be governed by its own norms independently of the norms of other expressions. Individuals would have to learn what commitments each particular expression involved. Knowing the meaning of one piece of language would be no help in understanding the meaning of any other.

Such a language would not be parseable, that is, the component sounds of an expression would have no linguistic function or meaning in themselves. This may be the case in learned animal communication. Dogs, for example, must be trained separately for each specific command. Even if a dog responds correctly to “Your bone is in the garden” and also obeys the command “Bring me my slippers,” it may be unable to respond to “Bring me your bone.” That is, it doesn’t grasp the syntactical structure, “Bring me xxxx,” in which various nouns can be substituted for xxxx. Similarly, an early hominin without syntax might learn how to respond correctly to the sounds, “Bring-a-spear,” but would have to learn from scratch what to do with the sounds, “Bring-an-axe.” Mastering one wouldn’t help in mastering the other. Each expression would have a linguistic meaning, but it would be a meaning only of the whole unit. The component parts would be meaningless and unsubstitutable.

Syntax has many advantages. Nevertheless, evolving the cognitive capacities to master syntax is costly. Nowak presents a mathematical model to argue that once the number

of speakers and number of objects to be spoken about exceed certain thresholds, syntactical language becomes more efficient than non-syntactical language. If Nowak is right, then the Chomskian insistence that syntax is the essence of language can be replaced by a simple pragmatic approach: at some point in hominin evolution, communities whose language came to rely on syntax gained a fitness advantage. This would explain why contemporary languages have syntax.⁴⁴

Nowak's approach makes it clear that it is simplistic to think of evolutionary explanation as based solely on chance genetic mutations. A mutation contributes to evolution only if there is selective pressure to preserve it. If an individual had developed the genetic changes that enabled the cognitive faculty for syntax three million years ago, in a context in which human cooperation was not yet significant, there would have been little selective pressure for its replication, so it would likely have died out. Only when the hominin way of life had already become cooperative and already relied on culturally normative communication—"language"—would mutations for syntax have a selective advantage. It is not syntax that explains the origin of language; it is pre-existing language that explains the arrival of syntax.

For comparison, consider the genetic changes that led to some Europeans producing lactase, which gave them the ability to digest milk. These changes replicated only because of the agricultural way of life within which they occurred. For all we know, the ability to digest milk may have appeared a number of times in various individuals over millions of years, but, without the agricultural environment 5000 years ago, the ability was not selected for, and so went extinct.

Syntax need not be an all-or-nothing characteristic. That some monkeys add inflections to their calls, or that some song birds recognize some syntactical errors, does not imply that they have full access to infinite substitutability or to recursion. The adoption of syntax probably happened in a gradual, stepwise process, like most evolutionary changes. We don't have to think of it as a giant step explainable by a sudden, single, chance genetic mutation.

Talking to Oneself

Individuals who have mastered their group's linguistic norms, syntactical or not, may learn to talk to themselves. Initially this might involve talking out loud but might develop into

whispering to oneself, and then later into conversing silently without vocalizing. (For comparison, it was once common to read texts out loud and only quite late in the medieval period did silent reading develop.) In whispering, and even in silent talking to oneself, there are vocal cord movements that are internal to the body. Higher level linguistic brain processes may occur even without incipient vocal cord involvement.

One must be careful, however, in referring to this development as “internalization.” Such talking to oneself is of course internal to the body and the brain. If one is still a captive of mentalism, however, one might misunderstand it as internal to a mind. The danger would then be to revert to the notion that linguistic meaning originates in the mind, Cartesian fashion. But just like talking out loud, the linguistic meaning of whispering or even of higher level neural processing is still determined by the group context. One must not confuse normatively defined actions with their underlying causal mechanisms. All behaviour, linguistic or otherwise, depends on body and brain processes. One cannot perform a ritual dance without brain activity, but that doesn't make the dance “internal,” as if it were all in the mind. The performance qualifies as a ritual dance because it follows cultural norms. Similarly, the linguistic meaning of language activities, whether loud or silent, is still determined by the communal norms.

The notion of “internalization” can be misleading in another way. Some think of norms as starting off “external” and later being internalized. This way of thinking suggests that individuals initially follow social norms because of external pressure, but later come to follow them voluntarily. Such a suggestion presupposes that individuals start off with their own internal, mental life based on their own values and goals, and then later subordinate themselves to their group's norms, linguistic or otherwise. This is a mentalist interpretation that projects contemporary selfhood back into infants, and into early hominins. My claim is that, from infancy on, in coming to identify with a group, individuals develop habits of obeying social norms. Following group norms is “second nature.” It is not that infants initially talk to themselves (perhaps internally) in some baby-language and later have their community's language imposed on them. Only in learning to follow the group's linguistic norms do children gain the capacity to wield language in the first place. Just as following group norms in general makes individuals into tribe members, following linguistic norms makes individuals into speakers. It is in the nature of speakers that they follow linguistic norms. Suggesting that

norms become “internalized” just confuses the issue. They were never “external:” they were always learned habits.

Instead of “internalization,” it would be better to think in terms of rehearsal. The skill of governing one's speaking by communal norms involves anticipating what commitments others will hold one to. Linguistic skill is complex; to be successful, one must “speak to oneself” and listen to one's potential utterances as if one were the recipient: will the listener be able to grasp the meaning of what one says? Will others ask for the reasons for one's assertions? The best—the fittest—speakers are those who judge their potential utterances by cultural norms in order to grasp in advance what their meaning is before releasing their utterances out loud to others. In this, language is no different from many other actions. Before spearing a gazelle, a hunter may rehearse the results of aiming at the neck rather than the head. They can then apply their group's traditional hunting method and get it “right.” Governance by linguistic norms applies to utterances in a similar way, whether the utterances are directed towards others, spoken aloud to oneself alone, or only rehearsed silently. Such rehearsals of what one might say—one's “thoughts”—continue to depend for their meaning on the linguistic norms of the community.

Neural Processes

None of this, of course, implies that the brain is not involved in language. Without the necessary brain capacities, following communal norms is impossible. This does not mean, however, that language is internal to the brain. Bolhuis and Chomsky claim that “language should be understood as a particular computational cognitive system, implemented neurally.” As primarily a mode of communicating meaning, which is how I see language, cognitive mechanisms *implement* language. They are not themselves language. Language is a communicative *function*, so like all functions, it is normatively defined. As long as individuals can succeed in living up to their community's linguistic norms, it doesn't matter which cognitive and brain mechanisms allow them to do so, though, of course, there have to be *some* mechanisms.

Placing language solely in the brain is a temptation for anyone, like Chomsky, who describes himself as a kind of “Cartesian.” Even if one gives up Descartes' position that the meaning of an utterance is the mental idea accompanying it, one could remain a Cartesian

“materialist” by locating the meaning of utterances within the brain. There are, of course, cognitive states in the brain implemented neurally, but without the context of a linguistic community, such states could have no linguistic meaning. By rigidly separating syntax from semantics, and defining language exclusively in terms of the former, Chomsky side-steps the issue of semantics, of how language carries meaning. Only by ignoring the communicative function of language could it be defined as a cognitive system.

Cognitive brain states and mechanisms are not intrinsically referential: they are not of themselves about things in the world. Material causal brain processes in themselves are just as incapable of being meaningful as are Cartesian ideas. What makes it possible for linguistic behaviour to refer to things in the world are the communal norms of usage. A brain state cannot, by itself, be *about* an elephant. A brain process could result in the sounds, “There’s an elephant,” but these sounds are about an elephant only because that is how the utterance should be used in the community’s language. Which brain processes cause the sounds is irrelevant to their linguistic meaning. Communication, like any function, could in principle be multiply realized, that is, could be implemented in different ways. An alien could master human linguistic communication with a totally different brain structure than *Homo sapiens*. Even among humans, left and right dominant brains probably use different neural areas, and stroke-recovery patients may relearn language using new parts of the brain. It is even conceivable that every human speaker implements their linguistic capacities in different ways. The internal brain processes that produce sounds no more account for the linguistic meaning of these utterances than the chemical makeup of the amoeba explains its functional food detector.

Assertions are actions performed by a speaker as a unified entity. An action, such as throwing a spear, is performed by an individual, not by their muscles or neurons, though those implement the action. Neurons and muscles are involved in throwing and in speaking, but they are not themselves the agents who perform the actions. “It was my muscles that did it” cannot excuse my killing another tribal member with a spear. Actions are defined by the cultural tradition of which the agent is a member. Linguistic actions are no different: an assertion in a language commits the speaker as a unity to the implications laid down by the linguistic tradition. Commitments cannot be attributed to vocal cords or brain states. A linguistic assertion is not an internal brain process: only the speaker as a whole, following the

norms of their linguistic community, can perform the action of uttering a meaningful linguistic expression.

Linguistic Umwelt

As language gradually develops, the world-for-hominins becomes ever richer. Nowak claims—as I mentioned above—that in the course of evolution the number of things to be spoken about increases. The increase is one of the factors that he claims increases the relative fitness of syntactical language. But it is not simply that more things lead to changes in language. This is not a one-way influence: a richer and more precise language in its turn make more things available to speak about.

For example, without words for labelling the situation, an action such as “waiting until the Spring Equinox before setting out to hunt seals” could not exist. Without words, a group might in fact only go hunting for seals in the Spring, but without the precision and future reference that language enables, such an action could neither be referred to nor undertaken. Language makes possible such sophisticated actions. Without language, actions like these could not exist in the group's repertoire.

Language is communal, so the things it refers to are common to all members of the linguistic community. It is the normativity of language that makes objectivity possible. An individual who says sincerely that a king is a commoner—that's how they feel about it—is not “subjectively right.” They are objectively wrong. The rightness or wrongness of their assertions are not theirs to determine. A speaker's utterances are able to refer to the things in the community's world only because they are a competent speaker of the communal language, that is, only because they are bound by its linguistic norms. Otherwise, their sounds are neither true nor false; they are gibberish. To share a common language is to live in a common cultural world that is objective in the sense that the individual's “subjective” beliefs should be subordinated to how things actually are in the culture's Umwelt.

The world that culturally endowed humans live in is, as I claimed in the last chapter, made up of actions and things defined by the culture's tradition, not by our contemporary scientific categories. Hominins who are equipped with language can talk about, perceive, and interact with a much larger domain of things. This is not just a psychological point: respecting a king is not a mental, subjective activity. A king is a real entity in the world; the action of

bowing before him is a real behaviour. It is not that the king is a pre-existing object that language secondarily gets to name. It is only within the linguistically defined world that there can be kings. Recall: food-for-amoebas is not an idea in an amoeba's mind but a real object in the world-for-amoebas system. Over and above biological and cultural things, a language-endowed human lives in a world partly constituted by their community's language. As we move from genetically determined organs to traditional learned languages, the worlds that these structures refer to, what they are about, evolve accordingly.

Conclusion

Language sets communication into a learned, cultural context. A lion's roar is effective or ineffective with respect to some biological function within the lion's context, perhaps mating or territorial defence. This is a species function, determined by the evolutionary history of the species. A linguistic utterance, "This is the king, to be respected" has a function—its linguistic meaning—within the context of a group's language. The advent of language transforms communication from a biological function into a cooperative enterprise governed by the traditions of the group with which the individual identifies. The implications that a speaker should be committed to constitute the linguistic meaning of the utterances they make.

Against appeals to mental ideas or to syntax, I claim it is social normativity that makes sounds into language. What words mean—and that they have any linguistic meaning at all—is not up to the individual, but is determined by the group's linguistic norms. Just as culture is a way of life transmitted by learning, so languages with their linguistic meanings depend on traditional norms. All extant human groups have language: it is a universal feature of all humans that has contributed to evolutionary success. Since human communication is based on tradition rather than genetics, however, each group has developed its own distinctive language based on cultural norms unique to each linguistic community. Which specific linguistic norms a group has developed is largely a matter of chance historical events, cultural "mutations" as it were.

The evolutionary development from purely genetic inheritance of ways of life to an inheritance based on learning is a radical change that distinguishes the hominin lineage from other biological life. Language, on the other hand, is less a radical change than a continuity, a gradual application of the cooperative strategy to communication. Nevertheless, this

application opens the way for more sophisticated and precise communication and makes possible a new kind of lived world.

Most importantly, for my story, the advent of linguistic meaning adds one of the most important preconditions for selfhood. Language prepares the way for narrative reference to one's individual history and for self-monitoring reflection. It makes it possible for individuals to take responsibility for their beliefs and actions. These, as we will see in the next chapter, are essential elements of selfhood.

Part II: Selfhood

Chapter 5

Self-identity

After this evolutionary preamble into its origins in Part I, we are now in a position to directly address the nature of selfhood. My position is that, like cells and organisms, selves are new, emergent unities. They evolve from the cultural individuals I have discussed in chapter 3. Like cultural individuals, selves are normative entities that are contextually defined by a social system. In addition, each self serves as a context that gives significance to its perceptions and actions. A self endures through time, that is, it remains self-identical throughout its personal history. It is the same self that makes commitments and assertions and takes responsibility for them in the future. Being responsible is the core of each self's freedom. Explaining this position is the goal of the next three chapters.

First, a comment on terminology. "Self" is a slippery term. People use the word in different ways.

One could, as an extreme example, use "self" in the amoeba system to distinguish the unified organism from its surrounding world of food and predators. Such a use, however, is so broad that it could not be helpful for picking out what is unique about human existence.

It might be more plausible to use "self" for individuals within the cultural system I described in chapter 3. Individuals in early human groups learned to take on different roles: hunting, gathering, tending the fire and so on. This individuality of roles could be described with the language of selfhood, but I have avoided doing so in order to bring out the distinctness of later evolutionary stages.

Some authors talk of a later stage when cultural individuals monitor their own mental states to track their own incipient reactions in order to reject in advance those that are

dangerous. An individual who becomes aware of their anger towards a chief can inhibit insulting behaviour and so avoid being clobbered. Such private monitoring of one's own states could be termed "self-awareness" and such individuals could be called "selves."

But I prefer not to do so. An organism's monitoring of its own states is very likely a prerequisite for what I want to call selfhood, but this is still not a self in the way I want to use the term. A further requirement, as we will see, is that a self be an enduring entity that preserves its identity over time, indeed over a lifetime. One could label such an entity a "person," but this term has legal connotations—"corporate persons," for example—that may distract us, so I will instead use the term "self" for this persistent entity. The nature of this entity, its mode of identity, and how it could come into being through evolutionary processes are the central themes of this current chapter. By the end of this analysis, the phenomenon I want to designate as "selfhood" should become clear.

The Problem of Self-identity

Let me start with the common conception of selfhood as autonomy. This concept stems from Enlightenment thinkers of the seventeenth and eighteenth centuries in Europe. It is based on the liberal idea that each person should be free to decide for themselves how they should live: how they should act; what they believe; who they should relate to; what politics and religion they espouse; and so on. The principle is that one is, or should be, free to live one's own life: to "be oneself."

Autonomy is an ideal. In practice, of course, there are economic factors, political powers and social forces that constrain individual autonomy. As Rousseau puts it, "Man is born free, and everywhere he is in chains."⁴⁵ But what exactly is this freedom, this autonomy?

It may be easier to pin down autonomy by considering, examples of non-autonomy. A slave in chains is not autonomous. Since they can see and feel the chains, they have the advantage of being aware of their lack of freedom. An evil neurosurgeon, however, could hypothetically implant a device in someone's brain that gave them control over their victim's action, without the victim being aware of this control. A popular understanding of hypnosis suggests that a hypnotist can make someone believe falsehoods that they experience as their own beliefs. A date-rape drug can make a person feel that they are consenting to sex when

that is not actually what they want. A man who drove some kilometres and killed his mother-in-law was acquitted on the grounds that he was sleep-walking and so not responsible for his actions. Someone with Dissociative Identity Disorder (formerly called Multiple Personality Disorder) may perform behaviours when one of their “alters” is active, behaviours of which the primary personality is totally unaware. Robert Louis Stevenson's 1886 story, Dr. Jekyll and Mr. Hyde, though fictional, is a classical description of this phenomenon. These are all obvious illustrations of non-autonomy.

Less obvious and more problematic is brainwashing of captive soldiers who may later disown their actions. War propaganda is somewhat similar: Nazi prison guards believed that they were performing actions of their own free-will. Some think—though this is controversial—that religious cults warp their members' beliefs so that they can no longer act autonomously. Commercial advertising may be less extreme, but the point of most advertising is to manipulate people into buying things the corporation wants them to buy, things that people, left to their own resources, might not buy.

In all these cases of non-autonomy, individuals come to hold beliefs or to perform actions which are to a greater or lesser extent not their *own*. Waking from sleep-walking, I might repudiate the recent behaviours and claim that they were not *mine*. In contrast, to say I am autonomous is to say that my actions are mine, that they are my own.

What is this *mineness* or *ownness* (I will use these terms synonymously)?

One might be tempted to see the organic body as the basis for ownness, but this cannot be correct. Although behaviours while sleep-walking, hypnotized or drugged may be exceptional, their existence make it clear that attributing actions to a self is not the same as discovering which body causes the behaviour.

Dualists hold that since the body cannot be held responsible for actions, there is a soul, a different “thing” beyond the body, to which we attribute the responsibility for action. Some dualists, Plato for example, discuss the possibility that after death a person's soul gets reincarnated into a new body but at that point everything that happened in the previous life is forgotten. Nevertheless, the myth of reincarnation has been used to account for one's destiny: the good or bad actions of the previous life explain one's present character and situation. It seems unreasonable, however, that I—today's self—should be held responsible for previous

actions I know nothing about just because they are attributed to the same soul. The sleep-walker, or the person with Dissociative Identity Disorder, rejects ownership of actions of which they have no recollection. Could someone with a reincarnated soul not be equally justified in claiming of the actions with which that soul was associated in a previous life, “They were not mine.”?

Attributing action to a soul is therefore no more successful at solving the problem of ownness than attributing them to a body.

A similar difficult arises with sensation. Imagine I am suffering from pain. Now imagine that, intense as it might be, the pain lasts only for an instant, and that I immediately forget it. As an example, consider a car crash that is very painful for an instant before the victim becomes unconscious and when they recover from the concussion they have no memory of the event. Even if a momentary pain reoccurs every minute, as long as I completely forget it, I may be happy and joyful throughout the rest of the time. I myself would not be a chronic sufferer because the pain is not mine. Since each instant comes and goes, unconnected to the next one, it doesn't enter into my ongoing experience. The fact that each instantaneous pain occurs in the same body or in the same soul doesn't make the pain *mine*. It is not an experience in the life of my self.

Some hold that experience is nothing but a series of discrete instants, of unconnected points in time. Let me call this idea of life “Punctualism.” Descartes holds that time is punctual in this sense and that every instant is powerless to give rise to the next instant. It may appear that I continue to exist through time, but this appearance, he claims, is due to the miraculous intervention of God who recreates the world every moment in such a way that it gives the (false) appearance of endurance. Even if we reject Descartes' theology, we must face the question of how a self endures through time. How is it that, unlike the sleepwalker or the person with Dissociative Identity Disorder, I am the same self over time? How is it that my actions and sensations today are mine, that they belong to the same self as I was yesterday?

The concept of autonomy by itself cannot account for this persistence of the same self over time. Even if at every instant a punctual agent were completely free of any constraining factors—that is, autonomous—there would still be no continuity, no enduring self. Each momentary action might be autonomous and unconstrained, but the previous and

subsequent actions, while still being autonomous, would be no more mine than the actions that happened in a previous incarnation, or that were due to an alter in the case of Dissociative Identity Disorder. So autonomy alone cannot solve the problem of personal identity, of mineness.

Inadequate Solutions

Memory

If the difficulty with the punctual understanding of autonomy is that the past is forgotten, is it possible that the solution to the problem of the self's unity and persistence over time is memory?

John Locke, in the seventeenth century, explored this possibility. He devised a scenario that has dominated discussion of the topic ever since: the Prince and Cobbler switch. Imagine that a prince wakes up one morning with the memories of a cobbler, while his own memories have been erased. Who would he be?

If self identity is based on the body or on a soul then the prince is still the prince, but endowed with the cobbler's memories. Asking him who he is, however, will get the answer, "I'm the cobbler," since what he remembers is the cobbler's life and falling asleep on the street the night before. Telling him that he is "really" a prince is hardly likely to convince him.

Even if the prince's body still incarnates the prince's soul, it would seem wrong to reward the person who wakes up in the palace for the prince's previous actions that he no longer remembers, or to punish him for the prince's previous misdeeds. From this forensic, social, or legal, point of view, what we have is the cobbler's self in the prince's body. The conclusion of this Lockean argument is that neither the body nor a soul can account for selfhood. It is the continuity of the stream of memories that identifies the self.

In the Lockean scheme, memory unifies the self: two states are states of the same self just in the case when one state is the memory of the other. But what about forgetting? If I forget some event in the past, does that mean this event ceases to be part of my life? Would this leave a gap in selfhood? A possible response is to modify the theory to allow overlapping chains of memory. For example, a soldier, victorious in battle, remembers being flogged as a boy, but has forgotten the whipping by the time he retires as a general. (The example is

Thomas Reid's.) Nevertheless, as long as the general remembers the victory and the soldier remembered the whipping, the boy and the general are still the same self. In this theory, it is these overlapping memories that account for the identity of the self over time.

This memory theory is inadequate, however, for it involves a kind of circularity. Rather than accounting for the unity of the self, this notion of memory presupposes selfhood. A memory is only *my* memory if it refers to my earlier experience. If one of your memories were to be transplanted into me—think of hypnosis or brain protein transplants—it would not be a memory of one of my experiences, that is, it would not be *my* memory. Transplanting a series of experiences from different people into my mind could not account for my self-identity. For a memory to contribute to my identity it must already be mine. Only if there is one enduring self, the same one that had the original experience and that now has the recollection of it, is the newer experience a *memory* at all. An experience could only be a memory of a previous experience if both of them are mine; if both are owned by the same self. Rather than a stream of memories being the basis for self identity, without the identity of the self a mental event could not qualify as a memory. Memories are not stand-alone entities waiting to be connected to one another: they must be *owned* to be memories in the first place. Far from accounting for mineness, the memory theory presupposes it.

Narrative

The narrative theory of selfhood offers an account of the unity and identity of the self that attempts to overcome some of the problems of the memory theory. Instead of a stream of stand-alone memories, a narrative sets a context within which memories are related to each other.

There are various versions of the narrative theory of the self. In general, the theory claims that what makes events and actions mine is that their temporal structure is like that of a story. A weak version of the theory proposes that pre-existing selves take on a narrative structure under certain cultural and historical conditions. This weak version of the theory presupposes that in certain societies a narrative form is superimposed on already existing selves. A stronger version of the theory holds that narrativity is intrinsic to selfhood, that it creates selves in the first place. It is not just that people tell stories about various characters and that they may in some cultures tell stories about themselves. The stronger claim is that

what makes a self distinctively different from a biological organism is the role it plays as a character within a story. The invention of story-telling is what led individual *Homo sapiens* to think of their own existence as a narrative. It is the story that accounts for the identity of a self over time. In the strong version, I do not take my pre-existing experiences and order them into a narrative: rather they are only *my* experiences in the context of a narrative structure. What makes experiences *mine* is that they are events in the same story.

Narrative is a linguistic structure; it is made up of events described in language. Hence, according to the narrative theory, selves are intrinsically linguistic entities. Dennett proposes that the linguistic module of the brain produces a set of sounds which, when interpreted within a language, produces a story that involves a protagonist and that protagonist is the self. Conan Doyle created stories in which there is a fictional protagonist, Sherlock Holmes. Analogously, the linguistic module of the brain creates stories in which the self features as a protagonist. In this account, the self is analogous, not to Conan Doyle, but to Sherlock Holmes. A self is a virtual entity; it exists as a protagonist in a story. As Dennett puts it, "Our tales are spun, but for the most part we don't spin them: they spin us."⁴⁶

Not all linguistic structures are stories. A simple chronicle of events makes up a story only if it holds together. At the very least the events are laid out in a temporal sequence, but that is not enough. There must be some progression in which later events build on, or are explained by earlier ones. That it rains and, at a later time, a bridge collapses are part of the one narrative if the bridge collapses *because* of the rain. Similarly, narratives about humans relate actions and beliefs to earlier experiences in the protagonist's life that make their current situation understandable.

Consider, as an example, a narrative explanation of Mary attending a human rights demonstration. One could say that Mary desires a world in which human rights are respected, and she believes the demonstration announced in a notice will help. A narrative account could go beyond this current belief-desire explanation and place her action in Mary's personal history. It might tell us that Mary—this same person—has in the past spoken with victims of torture, and that she has a vision of a future world based on justice. This story helps us understand *why* this notice influences Mary to act, while others may be unmoved by it. Unlike a simple chronicle of events, the narrative allows us to understand the significance the action

has for Mary. Attending the demonstration is not just an event: it is an episode in Mary's life-story. The integration of her action into a story about how she has come to be the person she currently is—how she has come to have the values and future aspirations she has—is central to the account being a narrative.

It is these temporal interrelationships that make up the unity of each protagonist. For a narrative to be a coherent story it must posit selves with persistent identity. A story is coherent only if an action and the experiences that led to it are attributed to the same person. The identity of selves over time is one of the central features that makes up the unity of a story, that makes it a story rather than a chronicle of unrelated events.

Narrative theories of selfhood may place the self's own narrative into a larger timeframe. One can think of an individual's story as beginning early, even before birth, when they were just a twinkle in their mother's eye, when their parents had a dream about the kind of life their hoped-for offspring would live. Each self is already a protagonist in a story before they master language, before they develop the skills to understand or narrate a story themselves. Parents and grandparents offer stories of their own histories. These stories set up models for their children, models for how they should live and for the roles they are destined to play, that is, the norms their lives should obey. "Your father was a fisher, as was your grandfather before him, and when you grow up, you will be a fisher too."

Family stories like these merge into mythical one. Myths are communally authored narratives inherited from the tradition of ancestors. Typically, they recount the origin of a group and are important in group identity. Romulus and Remus founded Rome and gave it its destiny. God calls Abram, renames him Abraham, and gives tribal identity to his seed forever. Māui fishes Aotearoa out of the sea and establishes Māori identity. Such myths set up a framework within which individuals find their roles within the culture. They serve to justify the social order and account for the origin of the group's language, technology, and way of life. Viewed this way, the narrative that establishes one's identity over time, one's ownness, is but a chapter in the ongoing epic that expresses one's community's group identity.

Strengths and difficulties of the narrative account

The narrative theory of selfhood has a number of strengths.

First, the narrative theory is an improvement on the overlapping memories account of selfhood. Instead of a series of memories that just happen to overlap, the narrative theory presents life events as integrated into the context of a story. This provides a more plausible basis for a self who persists over time.

Secondly, while the memory theory is individualistic and could allow for an isolated individual to have a stream of overlapping memories without reference to the social context, narrativity can offer an account of how an individual life story could develop within the context of a community. Social and parental stories can set the stage for each person to become the author of their own life. Models for selfhood are available to the developing child. Indeed, the very notion of a narrative self is bequeathed from the community, from the stories that are told, the expectations of parents, and the conception of life within the culture.

Thirdly, the narrative theory highlights the crucial importance of language for selfhood. The very notion of narration essentially involves language. A story, whether told to others or to oneself, is a linguistic entity. The memory approach relies on juxtaposition of memories or of overlapping memories. Narrativity relies, more plausibly, on language to integrate the stream of memories into a self.

Finally, the theory addresses the issue of how an experience could be mine. Events are mine, components of my ongoing life, in so far as they enter into my narrative, the narrative that makes me me. Events within a narrative take their meaning, identity and individuality from the story. If John gets married, that event qualifies as a marriage in so far as John courted his partner in the past, is now fulfilling some culturally mandated rituals, and anticipates a future life together with the person he is marrying. Without this context, the episode is not a marriage. Indeed, it is not even an episode, with boundaries and definition. The memory approach gets little traction here. Simply remembering the ritual events or the experiences that preceded them cannot themselves make up a marriage. Without the overarching context of the narrative, the memory of courtship, the current ritual behaviours or the anticipations of the future would not be those episodes; they would not have these meanings; they would not be *these* experiences. John's selfhood is not just a series of overlapping memories. It is his life story that gives these remembered episodes their meaning

in the first place. And, according to narrative theory, it is his life story that accounts for the identity of the self—"John"—to which these episodes are attributed.

The narrative theory therefore has major advantages over accounts of selfhood that rely on the continuity of body, soul, or overlapping memories to explain the persistence of the same self over time and the mineness of its experiences.

Nevertheless, despite these strengths, the narrative theory of selfhood, at least in this simple form, raises a number of questions. What exactly is the status of the narrative? While there are occasions when a self-narrative is recited aloud, in the case of a public confession for instance, or on a TV tell-it-all show, do not those who keep their narratives to themselves still qualify as selves? Some may publish autobiographies or record their lives in private journals, but the majority of people do not. Do the rest recite their stories to themselves silently? If one is not a good author, might one's story be poorly integrated and fragmentary? In that case does one have a fragmented self? What if someone didn't understand that their life was a story?

Could an individual live out a story of which they knew nothing? One interpretation of Freud, for instance, is that neurosis is due to an emotional event that one cannot speak about, even to oneself. Psychoanalysis, Freud's "talking cure," verbalizes the event and enables it to be integrated into one's conscious story. The unconscious narrative gives meaning to behaviours in a manner analogous to the way conscious narrative does. That meaning, however, is unavailable to the conscious self. If this account were correct, would it mean that, before the cure there were two narratives, and so two selves? One doesn't have to be a true believer in psychoanalysis to hold that parts of a person's life story could be unconscious. How could episodes of a story that one cannot express in language, that one doesn't know one has, be one's own? In what sense could they be mine?

Worries about the conscious status of narrative, however, are but the tip of the iceberg. The narrative theory of selfhood, at least in the simple version I have outlined, offers an intellectualist and individualistic misconception of the self. It conceives of selfhood as a kind of knowledge: I am who I am only because I know who I am. The true story of who I am, the story that makes me who I am, is assumed to be ultimately private. I may occasionally

express parts of my story to others (though not necessarily truthfully) but it is the story that I tell myself in my private inner life that is the real me. I am who I know myself to be.

This understanding of the narrative self is inadequate. If all there is to selfhood is self-knowledge, how could one ever be ignorant of oneself? Yet we've all met people who are deluded about themselves. Could I not be one of them? The Socratic admonition to "Know oneself" makes no sense unless we assume that there is more to the self than what one knows. This intellectualist account is a remnant of Cartesianism. It assumes that one is transparent to oneself, that one is a *cogito* in Descartes' way of putting it.

The narrative approach is an advance on Locke's idea that the identity of the self over time can be explained by memory and I will incorporate significant features of the approach—such as the importance of language—into my own account. Narrativity alone, however, is inadequate. It largely neglects bodily habit, evolutionary pre-conditions and the role of the community in defining each self. Above all, it fails to focus on the important element of normativity.

My own approach to the nature of selfhood—to which I will now turn—attempts to rectify these flaws.

Chapter 6

Normative Selfhood

Memory and narrative are major contributors to my project of understanding the self as a new, emergent entity. They take us beyond the idea that the self is an object, whether a physical object such as the body or the brain, or a spiritual object such as an immaterial soul or mind. Narrativity brings out the essential roles of language and personal history in the unification of the self over time. But we're not there yet! The crucial element that is still missing is normativity.

We have already seen, in chapter 4 on language, that linguistic meaning requires that speakers adopt the norms of their linguistic community and commit themselves to the implications of their assertions, and that listeners hold speakers responsible for these commitments. My claim is that selfhood has a similar structure. It is in the community holding individuals responsible for all their actions, not just linguistic ones, and in individuals accordingly holding themselves responsible for them, that the self comes to be as a normative entity.

The norms to which a self is committed set the context for feelings, perceptions, actions, and beliefs. It is this normative context of the self-system that makes up the being of a self and explains its identity over time, that is, the ownness/mineness of action and of thought. When, in the course of its evolution, a society sets norms for selves, including the ideal that individuals set norms for themselves, we have the arrival of selfhood.

During childhood, selves learn how to bind themselves to personal norms and develop habits of commitment that allow others to rely on them. Over their individual history each self develops a character that enables them to bind themselves instinctively and emotionally to the norms they commit themselves to. This capacity for commitment establishes a reputation that allows others to trust them. It is the continuity of their character

over time that establishes their self-identity. In so far as their norms and actions flow from their character, a self is authentic and free.

Unpacking this dense cluster of concepts is the role of this chapter. I will start with the concepts of binding and character (a kind of second-order binding), then discuss freedom, authenticity, and reputation.

Binding

Let me start with a very simple example of a normative temporal relation—promising. If I promise you yesterday that I will give you \$20 today, then I have created a new norm for myself. As a result of my commitment yesterday, today I am bound to giving you \$20. This binding changes my world. The action of giving you \$20 now has a particular meaning for me: it is the fulfillment of the new norm. If I don't give you \$20 then I have failed to follow my norm. The action of giving someone \$20, could be an act of charity or generosity, or maybe the result of extortion. To qualify as the fulfillment of yesterday's promise, however, handing over the money today must refer back to yesterday's commitment. What is the nature of this referring?

It is not a causal relationship between yesterday's event and today's, in the mechanistic sense. The physical facts of causing the sounds and of passing over \$20, the factual sequence of events, cannot account for the nature of the actions. A young child who has not yet mastered the skill of promising (or a computer) might utter the sounds "I promise" one day and hand over \$20 the next day. This would not be a case of promising.

Neither is it simply that the two events are in the same narrative. One could tell a story that mentions yesterday's and today's events, but it is not the narrative sequence of events that makes the episode into a promise. The referring is normative. Only if by yesterday's utterance I bound myself to the norm that I should pay the \$20 tomorrow is today's action the fulfillment of a promise. Yesterday's promising binds me in the sense that I placed myself under the obligation to pay the money today. Only within a normative context do the two actions qualify as the making and fulfillment of a promise.

It is not just the actions that are defined by the temporal relationship. Perception is similarly normative. My perception of who you are today—the promisee—depends on

yesterday's binding. My emotional state—the feeling of obligation—would be different if I hadn't made the promise. Note that these actions, perceptions, and emotions are unique to me. Others, not having made yesterday's promise, are not bound by the commitment to pay \$20. These things in my world have the meanings they do only within the unique context of the norm I have committed myself to.

Let me generalize this structure to the overall life of a self. Trading and economic activities, marriages and other contracts, teaching and apprenticeship relationships, politics and friendships all involve actions that one is obliged to perform because of commitments one has made earlier. These personal norms contribute to the construction of the world a self inhabits, its *Umwelt*. For example, if I contracted last year to write a book, taking time today to write counts as a success; neglecting to write is a failure. Because I committed myself in the past to fighting for justice, I am now bound by the norms of justice, so I interpret the sight of poverty as a challenge and I judge the responses I take today by whether or not they promote equality. Past experiences are not just passive events: they actively change the structure of the self and the meaning that future events will have. The nature of the self as an entity, and of the world it lives in, are determined by the norms it has bound itself to in its past.

Character

How could such an entity come to be? Why would anyone trust me to fulfill my commitments? Even more crucially, how could I trust myself to live up to my own norms? The ability to bind oneself is not just given from the get-go. Nor can I decide at this instant that I'm going to be the kind of individual that commits myself to my own norms. That would be circular: I cannot commit myself to being any particular kind of individual unless I am already capable of commitment. Rather, the capacity to bind myself to my norms comes about gradually over time as I develop my "character."

The term "character" has many uses. I will be using it exclusively to refer to the capacity of an individual to live up to their commitments. Some people, for instance, break their promises often; they have an unreliable character. Those with a reliable character are the ones that others can count on. Unlike the genetically based capacity of amoebas to follow norms, however, the human capacity for binding to personal norms is learned over a lifetime rather than being hard-wired from birth.

One learns to be a self. In opposition to the idea that God implants a soul by a miraculous creation, or Descartes' notion that the mind is always already fully formed—"I think therefore I am"—my position is that selfhood develops gradually. Any human individual starts life with a genetic endowment made up of biological skills accumulated over billions of years of evolution. Growing up within a community, they also absorb their culture's traditional norms to which they largely conform. Selfhood, however, goes beyond such conformity. In some, evolutionarily recent, societies, one of the social norms is that individuals should set personal norms for themselves. That is, such societies expect individuals to become selves by learning how to bind themselves to their personal norms.

Personal norms take selves beyond conformity. Traditional cultural norms originate in the evolutionary history of the tribe. The norms that make up the character of a self, however, originate in the personal history of an individual. A tennis player, for example, could commit themselves to practice every day. The rules of tennis are social norms of the Tennis Association to which all players must conform if what they are doing is to qualify as tennis. Committing to daily practice, however, is a personal norm, not a case of conformity to a social rule. According to this personal norm, they *should* practice every day. If they do, that's a success; if they don't, that counts as a failure. This "should" is not imposed by any cultural tradition, but comes from their commitment to practice regularly, an event in their personal history. The norm binds only the individual who adopts that norm. It is by the adoption of one's own norms—the development of one's own character—that individuals move beyond conformity to selfhood.

Character is itself normative. It involves adopting the norm that one should reliably live up to one's norms. Character involves being bound to such second level norms—"deutero norms." An example may help to explain what I mean by "deutero." Some food producers commit themselves to the norm that their products be organic. Because customers may doubt the corporation's commitment, other third party organizations are called on to certify that what the producer labels "organic" really is organic. Such certification is a deutero claim: it is a second-order claim that the producer's first-order claim can be trusted. Analogously, deutero norms of character certify first-order commitments. For instance, if I promise to pay \$20, the promisee can trust me because I have a trustworthy character, that is, because I have

committed myself to the deuterio norm of living up to my promises. Having a good character involves deuterio commitments to carry out first-order commitments.

In principle, there can be third or even higher order norms. Someone who asserts, “The cat is on the mat” must be committed to the linguistic norms that make the assertion meaningful, even if it is a lie. A self could also adopt the deuterio level norm of making this assertion only when it is true. A higher level norm would be to commit to telling the truth in all cases, that is, to having an honest character. Committing to being a good person, that is, to honesty, trustworthiness, reliability and so on, is higher level still.

“Character,” then, as I am using the term, refers to the set of norms of various levels to which a self is committed. Since one's norms grow out of one's history, they may be a disordered jumble, possibly even with conflicts. There is no guarantee that a self is fully coherent. Life may throw up situations where one is torn in more than one direction. Sometimes it is only in such quandaries that one realizes that one must modify one's commitments.

Character must have a certain stability. The point of commitment to norms is that they bind in the future. If by a promise I commit to paying you \$20, but when the time comes to pay you I deny the obligation, then I have failed to bind myself by the promise, that is, it was not a real promise. For an adult to be trustworthy, they must have a stable character. Someone who appears trustworthy today but not tomorrow is simply not trustworthy. A character without stability is not a character in the sense that I am using the term. Just as genetic inheritance must be stable for an organism to be a member of its species and an ever-changing cultural tradition is simply not a tradition, so character must be stable for one to be the same self over time.

Stable, however, doesn't mean eternally static. While evolution requires modifications—mutations—they must be infrequent if successful skills are to be preserved by the species. Since character is learned rather than genetic, it is subject to change over a lifetime rather than over generations. One's character can be developed by new learning throughout one's life. For an individual to remain self-identical over time, however, character changes cannot be too frequent or too drastic. How one has learned to live must be largely conserved, or such learning is pointless.

Such changes, however, are cumulative, that is, they fit into the context of the self. Many influences impinge on a self, but they do not simply push it around: how events change a self depends on how the self interprets them. External events are only opportunities, not causes. In cultural evolution, as we saw, technology was cumulative: roots could be used as food only after fire had already been mastered. In the life of a self, it is the context of the self's present character that determines how outside influences are interpreted and how new norms are integrated into its future character. That's what it is for character to be cumulative. It is only by later commitments being adopted within the context of earlier ones that a self preserves its identity over time.

For example, the friend of an adolescent girl might propose that she embrace a mystical life based on drugs. Since the girl feels that the norms of such a life are incompatible with her prior commitment to try out for the Olympics, she rejects the proposed values and refuses to incorporate them into her life's norms. Another girl, dedicated to religious values rather than athletic ones, might interpret the proposal positively and adopt the value of mysticism into her developing character. How her friend's proposal influences the girl's future character depends on the character she has already adopted.

A self, therefore, has some responsibility for its own character. As a child, biological inheritance and parental influences play the major role in character formation, but as a self matures it gradually takes on more responsibility for its own character. The norms it has already adopted enable it, more and more, to commit itself to a life based on its own values.

Such cumulative development is not just a matter of memory. Consider again Reid's example of a boy who is whipped for disobedience. In adulthood, he becomes a soldier because he has an aggressive character that values strength and courage. He doesn't just remember his whipping in an episodic recall of the past event. The past has formed his character and influenced his norms. Later in life, as a general, he values warfare and victory. Whether he has a memory of being an aggressive soldier, or has forgotten the whipping, is not the central point in who he is. In the memory theory of identity, it is the overlapping chain of episodic memories that accounts for the continuing self-identity of the boy, soldier, and general. My claim is different. The whipping is an event in the developing character of the boy that, in conjunction with many other events, leads to the soldier adopting aggressive norms

which in turn leads to the victory-obsessed character of the general. One's present character is shaped by one's personal history, whether or not the events can be recalled in memory.

There is more than one concept of memory. Locke's memory theory of self identity conceives of memory as a series of conscious, individual experiences – as a kind of intellectual knowledge. More recently, psychologists have distinguished “episodic” memory from other notions of memory. Skill memory, or know-how, for instance, is not episodic and need not be conscious. One can remember how to ride a bicycle without being conscious of how one is doing it, or being able to recall the occasion when that skill was learned. Although language learned in childhood is remembered in adulthood, there may be no recollection of the episodes when this skill was learned. While one's character depends on the past, it is more like a skill rather than a recollection of discrete episodes.

Character formation should not be misunderstood as an intellectual process. It is not that children think about their character and decide consciously how they want to develop it. That would be circular: we are trying to explain how such self-consciousness comes into being. Rather, a child brought up in a family that values honesty, feels guilty when he lies. He becomes disgusted with himself and feels (rather than thinks) that he wants to be different. As the child learns to avoid such bad feelings, honesty becomes habitual. The child has acquired a skill for dealing with his world. As Aristotle would put it, one becomes virtuous by practising virtue. Guilt and shame, rejoicing in being aggressive or being compassionate, are emotional, not intellectual states. Virtue (or vice) becomes habitual, embodied in one's visceral reactions to situations.

It is in these habits that that make up the self as an entity that endures over time. The temporal nature of the self means that the self at the current moment is not abandoned alone in the instant with only its immediate awareness and intellectual judgment to depend on. It already sees the world in a way that depends on its character and is motivated by habit to respond in a manner consonant with its norms. It can rely upon its “second nature,” – on what it has become. The warrior in his aggressiveness sees another person as an “enemy,” a perception that provokes the action of killing. Such perception and action are embedded in his character as a warrior. A self with a pacifist character sees others as fellow humans towards which he feels compassion and instinctively acts to help them.

These perceptions and responses to others are not primarily present intellectual judgments.⁴⁷ One does not have to decide anew every moment which norms one will be bound by or recommit oneself every instant. As an historical entity whose past commitments endure into the present, a self being bound to their norms is not a momentary phenomenon, but an ongoing habitual one. One's ingrained habits set the context for how one sees and evaluates events in the world and for how one should respond.

Habit should not be understood in an intellectualist way. As I am using the term, habit refers to the set of norms a self has acquired over its history. While one can reflect on a habit and make it an object of knowledge, usually one does not. Indeed, reflecting on the habit of bicycle riding or skating may lead to falling! Habit refers to how the past defines and evaluates the current moment, how the world—the Umwelt—is perceived, and how one is instinctively inclined to act. One's habits are the norms that make up one's character.

We can think of character as playing a role analogous to tradition. A culture inherits traditional norms that facilitate the group's survival. A self relies on the skills it has developed over its lifetime when facing its future. Choosing the appropriate response in the present could be seen as a kind of cooperation between past experiences and the current situation. For someone with a trustworthy character, for instance, only rarely would they have to explicitly or intellectually decide whether they should keep a particular promise or not. That response is, as it were, pre-decided.

Character development, in other words, is structural. It is not just the learning of a collection of facts or the memory of purely passive information, trivia style. Past experiences reprogram the self. What has been learned shows up in the structuring of the self's relationship to the world. The history of the self consists of it interpreting episodes, incorporating them into its set of values, and thereby modifying its own habitual nature so that it interprets future events in a characteristic way and acts accordingly. The self is an enduring system of contextualizing norms: experiences influence the self by modifying the context within which events and actions are evaluated.

We are familiar with this mode of understanding. The amoeba's functional reaction to food-for-amoebas is a crystallization of its species history. It is on the basis of its ancestral traditions that each culture sets its norms for the tribal individuals within it. In these cases, the

past inhabits the present not as a set of facts, but as a mode of structuring the current situation. Similarly, the historicity of the self shows up not so much in its memory of previous episodes, or even in the narrative it tells itself, but in the formation of the norms that make up its character and that maintain its self-identity over time.

So to answer my earlier question: I can trust myself to live up to my norms because of my character. “I” am not a punctual entity, an autonomous self, a pure “cogito” new and isolated in each instant. I could not trust a momentary intellect to be guided by my norms; but I am not a pure moment of consciousness. I am an ensemble of habits that endures from the past into the present and on to the future. My present moment is the tip of an iceberg that is carried along by the underlying mass of accumulated habits. What I can trust is the entity that endures through time—my self.

Freedom

My account of the self as being bound by the norms of its historical character offers us a positive way to understand what it is for a self to be free.

There is a temptation to define freedom as the opposite of determinism. When European thought in the seventeenth century devised the fundamental principle of the mechanistic worldview that all events have a cause governed by a universal law, there was a quandary about what to do about the mind. If the mind were part of the causal, deterministic world, then values like truth and reason would be illusory. In response, some insisted that the mind was “free” in the sense that it was beyond the reach of deterministic laws. This notion of freedom buys into the dichotomy of bodily causality and immaterial, spiritual mind.

By now, it should be clear that this is a dichotomy that I reject. A cell's homeostasis creates new norms but does not violate any physical laws. An amoeba is defined by biological norms, not by its mechanistic processes. Cultural norms cannot be understood by mechanism. A self may depend on causal mechanisms, such as those in the brain, but it is defined by its normative status. So we need a quite different approach to freedom.

Freedom is sometimes portrayed in a negative way as freedom *from* some kind of constraint: a slave become free when they escape from their shackles; the people become free when they throw off the yoke of an oppressor. This negative concept of freedom is central

to the Romantic idea that selfhood and society are antagonistic to each other. In the European Romantic era of the mid-nineteenth century, to be oneself meant to reject social norms, to act spontaneously, to go it alone, or to be a heroic genius fighting against the pressures of society. In music, art, and literature the Romantics held that the true destiny of the self was to escape from society and pursue its own course at whatever cost. William Wordsworth rejected industrial society and glorified the self's relationship to nature. Henry David Thoreau actually went to live alone in the forest to get away from society. Beethoven eschewed classical forms of music of his society and converted music to a heroic expression of his own individual and heroic emotions. Isaiah Berlin describes the Romantic idea of freedom this way:

The painter, the poet, the composer do not hold up a mirror to nature, however ideal, but invent; they do not imitate (the doctrine of mimesis), but create not merely the means but the goals that they pursue; these goals represent the self-expression of the artist's own unique, inner vision, to set aside which in response to the demands of some "external" voice—church, state, public opinion, family friends, arbiters of taste—is an act of betrayal of what alone justifies their existence for those who are in any sense creative.⁴⁸

While I agree that to be a self is to be a free being, we need to ask what freedom is. The Romantic ideal that freedom is pure, unbounded spontaneity runs up against the problem of mineness. A biological reflex, like a knee-jerk, is not bound by any social or personal norms, but while it may appear spontaneous in the sense that it is not bound by society, it doesn't qualify as a free action because it is not mine. The self has no responsibility for the movement. Spontaneity in this sense is a poor model for free action.

The phenomena of inspiration and possession are also poor models of freedom. If I am *inspired*, in the original sense of being possessed by a spirit—good or evil—then the actions performed cannot be attributed to me. In ancient Greek Dionysian cults, adepts were possessed by Dionysus, threw off their social roles and engaged in wild, "spontaneous" activities. While they may have felt liberated from the oppressions of society, they didn't thereby become themselves: they became Dionysus. A more recent phenomenon is the wild activity of soccer fans when cheering a goal. Their enthusiastic state is not so much an expression of their personal characters as a surrender of their individual norms to the crowd's. Their actions belong more to the collective than to the individuals. (Etymologically, to be "enthused" originally meant to be inspired or possessed by a god. In Greek, *en* + *theos*.) If we

take literally the claim of some writers that it is their Muse that composed their poetry, then the poetry is not theirs; it does not flow from their selfhood. Such phenomena are not examples of the self's freedom. Reflexes, mystical possession by a god, or merging with the crowd don't count as one's own actions in the first place, much less as expressions of freedom.

In his novel *Crime and Punishment*, Dostoevsky's explores the ambiguities of the Romantic notion of freedom. His protagonist, Raskolnikov, commits a murder while in a distraught, feverish state, Raskolnikov tries to interpret the murder as an expression of his absolute—Romantic—freedom and as a rejection of society's control over him. In the aftermath he can no longer sustain this interpretation and his guilt leads to a disintegration of his personality.

Since I have, over a lifetime, bound myself to respecting the rights of others, then today I am not free to be a murderer. Murdering someone would not be an act of freedom for it would not be consonant with my character. It is not a freedom that I would want, for it would violate my personal integrity. The way I have developed my character norms is such that I can trust myself not to murder others. Murder is no longer in my repertoire. It would not be, for me, an act of freedom. Freedom is not a matter of rejecting social norms; it is a matter of being oneself, of acting in character.

The Romantic era's claim that free selves and society are antagonistic is particularly ironic in that as an "era," as a social movement, Romanticism came about in a specific social and historical environment. The idea that the self is opposed to society is itself dependent on a particular social structure. Romantics seem to take selfhood for granted, as a given, as if it fell from heaven. Perhaps they think of selves as springing fully formed from Nature or from a creative action of God that miraculously endows all people with freedom. Once Darwin published *On The Origin of Species* in 1859 the Romantic understanding of the self was ultimately fated to disappear, though anti-evolutionists have tried to keep it on life-support for over a century.

After Darwin, selfhood can no longer be taken as an eternal given: eons of biological and cultural evolution had to occur for selves to come about. While there may be occasional conflicts between personal and social norms, society and selfhood cannot be fundamentally antagonistic. Indeed, selves depend on society for their very existence. It is only in a

particular kind of society, one that holds individuals responsible for their actions, assertions, and thoughts, that there can be selves at all. Without the appropriate social norms, selves would not have evolved.

Rather than thinking of freedom negatively—as freedom from society—we should think of it positively as being oneself. For me to be free in a positive sense is for my actions, feelings, and thoughts to be bound by the personal norms that I have embraced in my life history.

It may seem paradoxical that binding is what enables freedom. Yet we have already seen in chapter 4 that only by being bound by the norms of a linguistic community is it possible for individuals to make meaningful assertions. Someone who loses the ability to follow norms, loses the ability to make agreements with others. Only when an individual learns how to commit themselves, do they gain the freedom to make contracts.

Freedom is not an endowed right bestowed once and for all at birth as the nineteenth-century Romantics claimed; it is an accomplishment that is achieved gradually as one becomes oneself. Where they claim that freedom involves the rejection of norms, my claim is that the ability to bind oneself to norms is required for an individual to be capable of making commitments. Being bound by one's past decisions is not determinism; it is freedom. Only by following one's personal norms can one maintain one's identity over time. My action could not be free unless it is my own. My character determines who I am. I wouldn't be me if I didn't act this way, if I didn't see things this way. As Luther put it, "Here I stand, I can do no other." I become free by developing my character, that is, by progressively binding myself to my own norms. To be free is to be oneself.

A self's temporally developed character takes it beyond the conformity of tribal members in an early hominin culture. The very nature of tribal members is to conform to traditional norms. As I claimed in chapter 3, being bound to these cultural norms is what defines one as a hominin individual rather than a purely biological organism. For a self, in contrast, conformity means acting and thinking in accord with norms that are not one's own. What makes an individual a self is getting beyond conformity and following norms consonant with the character they have developed through their own life-experience.

Selfhood is a kind of ideal. Those who strive to be selves do not simply leave cultural conformity behind. Often in their daily lives, contemporary humans simply conform to the norms of their culture. They allow themselves to be directed by others rather than directing themselves. In other words, much of the time they act like traditional tribal individuals. For a parallel, recall that the behaviour of early hominins is often governed by biological norms: cultural norms build on biological norms rather than just replacing them. Similarly for selfhood: individuals who are selves do not cease to be members of a culture and much of the time they simply conform to their society's expectations rather than following their own personal norms. Selves cannot float free of their society once and for all. If they could, there would be no need for the sage's admonition to "be oneself."

Being oneself doesn't require novelty. When a self does commit to a personal norm, they may just adopt pre-existing cultural norms as their own. In a militaristic culture, for instance, a young man might take it for granted that going to war and killing others is his patriotic duty. That's what all his friends are doing. That is simply conformity. On the other hand, it is also possible that he might actually commit himself to some cause and go to war as a way of living up this personal norm. This is no longer conformity, but acting in character. A new personal norm can be original and created by oneself—or not—but what makes it one's own is not its uniqueness or creativity; it is that it flows from one's character, from one's own historical commitments.

Reputation

A self, however, cannot create its character in isolation. A social environment in which reputation counts is essential for the existence of a self. Note that in defending this claim, I use "reputation" in a particular sense. While the word "reputation" can be used to refer to what people say about someone, that is, to gossip, as I use the term it means the established social expectation that an individual will accept responsibility for their commitments.

Again, consider promising. Promising is a social affair. It is my reputation in the community that gives me the power to make promises. If I try to use sounds or gestures to make a promise but the individual I'm addressing fails to take them as promises, then my attempt at promising fails. Maybe the other individual doesn't hear me or speaks a different

language. Perhaps they think I am actually a bot or an automated voice. Unless the other acknowledges that I am a self with a reliable character, then nothing I say can bind me.

Others trust me because of my established reputation. If I consistently break my promises, I will lose my reputation, that is, I will no longer be trusted by others. Without this reputation, it will be impossible for me to make a promise at all. My utterance will not bind me. I may insist as often or earnestly as I want that I promise to repay a \$20 loan—I may even promise that this time I will keep my promise!—but my words won't count as promises. In losing my reputation, I'll have lost my ability to make a promise.

The requirement for reputation applies not just to promising, but to all commitments. All contracts depend on the contracting parties recognizing each other's capacity to make an agreement. One cannot, for instance, make a contract with a baby. It is not only that young children have not yet built the character needed for them to reliably bind themselves to their own norms; they also lack the social entitlement to make contracts. They gain this entitlement only as they mature and others come to recognize them as reliable; only then can their word or signature bind them into a contract.

An adult can lose their entitlement. No matter how much they might insist, the word of someone drugged or drunk doesn't count. A sleepwalker doesn't have the necessary social relationship with others; no one will trust them, no matter what they say. An individual suffering from severe dementia has lost their reputation; any document they sign is invalid. Without an established reputation for reliability one cannot agree to any commitment.

This structure is analogous to my earlier claim about language: someone who regularly rejects the implications of their utterances will be deemed not to be making assertions. They will no longer be recognized as speakers, as members of the linguistic community. They will be dismissed as talking gibberish.

Similarly, being a self is a communal relationship, not an isolated attribute of an individual organism. Being a self, that is, an organism capable of making commitments and binding itself to norms, requires communal reputation as well as individual character. One can no more be a self by thinking of oneself as a self, than one can bind oneself to a promise on a desert island. One cannot establish one's reputation without the cooperation of others. No

matter how consistent an individual's behaviour might be over time, they cannot have a reputation unless others recognize them as a self.

Such treatment is not momentary. Like character, reputation has a temporal structure. Today's reputation is a kind of crystallization of what others have learned about my character from my previous behaviour. They have seen my fidelity to the norms I have adopted. If all goes well, my reputation reflects my character. What I should do according to the norms I'm committed to and what people expect me to do are two sides of the same coin. Just as my character is an expression of my lifetime experience, so my reputation reflects my social history.

An individual's self-identity over time therefore is not up to that individual alone. There must be a community that holds each individual responsible for their commitments and actions. One's reputation—in my sense of an established social expectation that one will accept responsibility for one's past commitments—is a requirement for being a self. Selfhood is a normative status: only an entity capable of binding itself to norms can be a self. But without social reputation, without others' recognizing one as bound, one cannot have this status. One cannot be a self by oneself.

Summary

A self, then, should not be thought of as an isolated entity, complete in itself. It is part of a larger system. When a culture sets a norm that individuals should be governed by their own norms, we have the arrival of a self-system. The interrelated complex of character, reputation, commitment, and freedom establishes the system within which selves and the worlds they live in can come to be. Genetic evolution set up the amoeba-system within which amoebas and their Umwelt were possible. In contrast, it is a learned cultural environment that establishes the self-system. The components of this system fit together in correlated ways. Individuals should develop their own characters and their community should attribute to them reputations that match their characters. Selves should unify themselves over their lives, while society should hold them responsible for the past commitments they have made.

It is only within the context of such a self-system that selves and the worlds they inhabit are possible. In evolutionary terms, the self is a new emergent entity. It is not an independent object-in-itself, a soul, a body, or a cognitive neural structure in the brain. It is a

normative, not a factual, unity. A self is a temporal unity of personal norms that it has signed onto over its narrative history of commitments. This history manifests itself in the present as the self's character, that is, as the habits of perception and action by which it lives in its world, in its Umwelt. Personal identity—the “mineness” of actions, perceptions, and experiences—refers to the relationship between the self's present character and the past commitments that led to it. A self is free in so far as its actions and beliefs flow from its character and identity. Commitments are possible only in the context of a community that holds selves responsible, that is, a community that sets the stage by establishing a norm that individuals should set their own norms. Without reputation—the recognition by the community that an individual takes responsibility for their commitments—an individual cannot become a self. A self is not an object; it is the status of being bound to personal norms.

Chapter 7

Interiority and Belief

But I hear an objection!

The account of selfhood you're giving is ludicrous. It violates common sense. It misses the central point of being a self! The most fundamental feature of the self is surely that I am a being that experiences things in the present moment! I am an inner stream of consciousness, the series of ongoing interior experiences that are private to me. Even if I then place my immediate experience into an historical narrative, or decide to act in accord with social or personal norms, surely such things are secondary. First and foremost, my selfhood is characterized by self-consciousness, by my awareness of my own thoughts and feelings. It is my own private interiority that makes me a self. Could anything be more obvious?

Well, yes! It is even more obvious that there is a world of things in themselves: stones and flowers, money and police, the flat ground of the earth, the rising sun. These things seem self-subsistent and independent of humans, as if they are waiting there already for me to see, to think about, to use, to walk on. Scientific advances, however, have undermined these common-sense, obvious things. Stones are conglomerations of atoms separated by empty space and appear as unified solid objects only because of the causal interactions of these atoms with the neural process of the human perceptual apparatus. Money is a social construct and would cease to exist if society vanished. The earth is not flat, but round. The sun doesn't rise; the earth turns. The common sense obviousness of the everyday world dissolves when examined more closely.

The obviousness of a private stream of consciousness that I am immediately aware of—the “myth of interiority”—fares no better under scrutiny. Descartes, in the seventeenth century, claimed that each person has a spiritual mind, an “inner” space with representations of things in the “outside” world. As a pure interiority, all the mind can know are the ideas present inside it. We have no direct access to the things in the exterior world, he claimed. Since his time, so many have bought into his metaphor of “inside the head” (or “in the mind”)

as opposed to the “outside world” that this way of speaking is often taken for granted without analysis. It has come to seem like common sense that what I actually see are really images on some inner screen surveyed by an “inner eye,” images or representations of “external” things that are themselves inaccessible to me.⁴⁹

When we think critically about this myth, however, its obviousness evaporates. There is no inner eye, and if there were, there would be no inner light to let it see anything—unless there were a crack in the skull! By what process could an inner eye perceive an image on an inner screen? Would the inner eye have to have an even more inner image of the image it looks at? And so on to infinity?

If we suspend the power of this myth, we can describe what is actually going on. When I look at a table, I am not looking at some interior image. It is the table itself I see—the table made of wood that I touch and bump up against. It is the same table that you can see, not some private representation of it inaccessible to anyone but myself. Far from being an obvious certainty, the notion that a self is an interior mental space of images is a highly theoretic postulate that collapses under scrutiny.

I've already argued in chapter 1 that an amoeba detects and eats food-for-amoebas, food that is in its world, not some internal image of the food. Of course, to detect food, the amoeba must possess food detectors, but it is not these detectors that the amoeba detects. It detects food in the world. To see a table, I must have a retina, a visual cortex and cognitive capacities interior to my brain, but it is not these processes that I see. They are mechanisms that enable me to see the table that is in the world before us. While there are cognitive brain processes that are interior to my skull and to my brain, I am unaware of them. Far from them being components of a “stream of consciousness,” I have no consciousness of them whatsoever. I know of their existence only indirectly from the research of neuroscientists. What I perceive are not inner processes, but things in my world.

What More?

The temptation to think of the self as a spiritual object—a soul—or to conceive of the mind as an “inner space” arises when one realizes that the self is more than the body. The temptation is particularly strong when the body and brain are explained in a mechanistic way.

From a causal point of view, every brain process is governed by deterministic laws. Within these processes there is no room for freedom, responsibility, or meaning. From this perspective, the existence of selves seems impossible. Yet as selves, we know we exist! My claim is that it is because people neglect the role of contexts that they are tempted to solve this conflict by inventing souls or interior mental spaces.

Imagine someone who realizes that a chess piece, say a rook, is more than the wood (marble, plastic, digital bytes) it is made of. What more? Could there be a chess soul in each rook that accounts for it being a rook and not just a piece of wood? Could there be a private interior process that makes the wood into a rook? No one would be taken in by such theories. The chess piece is a rook because of how it is used in the game of chess. The rook is indeed more than a piece of wood, but the “more” is outside the piece: it is the context of the game that confers on the wood its status as a rook. My claim is that what makes a self more than the body is its communal context, not a mythical interiority.

Consider (again) the heart. There is more to the heart than the factual, causal mechanics of pumping blood. What more? Vitalists in the nineteenth century believed that within the heart there was a “vital force” that accounted for its “life.” My account claims that the “more” is its function. A heart not only pumps blood as a matter of fact; it is *supposed* to pump blood. That norm comes from its role in the context of the body as a whole, and originates in the evolutionary history of the organism. It is the neglect of these contexts that leads to the myth of a spiritual vital force to explain the life of the heart.

Forgetting the context also explains the misunderstanding that parallel mental ideas are what give words their meanings. As we have seen, the meaning of a linguistic expression derives from the norms that the linguistic community enforces. The theory that it is ideas in the mind that give assertions their meaning may be obvious “common-sense,” but its obviousness is due to forgetfulness of the communal linguistic context.

Those who are captive of the myth of interiority, tend to dismiss others who challenge the myth as “behaviourists.” If the only answer to the “what more?” question that a dualist can conceive of is a private, mental space, then they think that those who reject this myth must be saying that there is nothing more than physical movements. “Either you believe in spiritual minds or there is nothing but physical processes.” The position I am defending, however, is

not behaviourism. There *is* something more than the physical, but it is not a series of mental objects in a spiritual space. The self cannot be understood either as the behaviour of a physical object, nor as a non-physical interior space.

My position is that the self is indeed more than the body. The “more,” however, is not a stream of consciousness in a private interior space. What differentiates a self from a physical body is its normative context that originates in its evolutionary, social and personal history.

Belief

Animal beliefs

The dualist notion of an interior space is used by Descartes and others to justify human exceptionalism. Since non-human animals do not have spiritual minds, they are just machines and can be treated as such. By the twentieth century, many became uneasy with this exceptionalism and, to be more respectful of nature, came to hold that animals have minds. In recent decades, some have gone a step further and maintain that some animals have a “Theory of Mind,” that is, they attribute inner mental states to other animals. Premack and Woodruff define the “Theory of Mind” this way:

An individual has a theory of mind if he imputes mental states to himself and others. A system of inferences of this kind is properly viewed as a theory because such states are not directly observable, and the system can be used to make predictions about the behavior of others.⁵⁰

The Theory of Mind approach has been very contentious for a number of reasons. My worry, however, is that it perpetuates the myth of interiority and, in trying to be respectful to other animals, extends a bad idea even further.

A better approach is to analyze the phenomena, human and animal, in terms of beliefs.

The word “belief” is used in many ways. Sometimes it just means trust, as in the phrase “believe in:” “I believe in you! You can do it!” Closely related is the religious sense: “I believe in God.” I will not be using the term in these ways, but only in the sense of “believe that.” For example, “I believe that it is snowing;” “I believe that there is a lion close by;” “I

believe that the moon is made of green cheese.” A “belief that” refers to how things are. It is a judgment about the world. Unlike a “belief in,” a “belief that” can be true or false.

Animals have beliefs that guide their actions. I have already claimed that organisms are normative entities defined by their roles within a system. Lions and gazelles, to take non-human examples, are not defined by the atoms of which they are made, but by their interrelated roles in the ecological system. Part of the role that qualifies an organism as a gazelle is that it should recognize lions, fear them and flee from them. The state of recognizing a lion is the belief that there is a lion. While this state depends on retinal, neural and other bodily processes, these processes are not in themselves about the lion. Their meaning—what they are about—comes from the roles they are supposed to play in the gazelle's relationship to its world.

Any belief is about the state of the world the believer inhabits. The gazelle does not have a belief about a collection of atoms. Nor can it have a belief about food-for-amoebas, shamans, dollars, or creditors. It cannot have a belief about chess pieces, not even a false belief about them. None of these are entities in its world. What it recognizes is an entity in its Umwelt, an entity defined by its pragmatic value: the lion is a predator-to-be-avoided. This lion-directed state could be triggered in error, in which case the belief is false. Yet even when mistaken, the state has a meaning: it is about a lion.

Every predator must be able to predict where the prey will go next. A gazelle that believes a lion is tracking it can just flee from the lion and try to outrun it. The lion can predict where the gazelle is going and pursue it accordingly. If a gazelle runs behind a thicket, the lion can predict when and where it will emerge and can head in that direction. The gazelle, however, can mislead the lion by changing direction while behind the thicket. Such a deceptive tactic requires that the gazelle can predict where the lion will, falsely, expect the gazelle to emerge. That is, the gazelle must be able to distinguish where it actually is from where the lion believes it is. In this (totally hypothetical) scenario, the gazelle not only has a belief about where the lion is, but also has a belief about what the lion believes. The gazelle attributes a belief to the lion.

The gazelle does not have to adopt a Theory of Mind to mislead the lion. It does not have to believe that the lion has a private interior space. What it attributes to the lion is a

relationship to its world, an attribution based on the actions that lions in the gazelle-lion ecosystem are expected to perform. The gazelle does not just perceive the lion's actual behaviour; in recognizing it as a lion, the gazelle sees it as a normative entity, in particular, one governed by the norm that it should predict where the gazelle will exit the thicket. The belief the gazelle attributes to the lion is not an idea in the lion's interior mind: it is the normative state of the lion that it should head towards the predicted end of the thicket to intercept the gazelle.

In my hypothetical scenario, the gazelle has a belief that the lion has a belief. What the gazelle does not have is a belief that it itself has a belief.

Self-attribution

A self, however, unlike a gazelle, does have beliefs about its own beliefs. How could this come about?

Let me suggest the following account. Like a gazelle, humans can believe that another human has a belief. Suppose someone observes a pattern in my speech, emotional reactions and behaviour in which I downgrade the value of those with a particular skin colour. The observer might then point out to me that I believe people of that colour are inferior to me. The observer is not using an introscope to observe an item in my private mind: his remark is an interpretation of my visible behaviour. I may never have noticed that I harbour such a belief; indeed I may initially deny vehemently and sincerely that I have the belief. By pointing out the pattern of my behaviour, however, the observer might convince me. That is, what is initially an attribution of belief by another can develop into a self-attribution.

Self-attribution of belief probably begins early in life. A young child might try to put her glove on the wrong hand. Her father might say, "You seem to believe gloves are all the same." The child may accept her father's attribution and say, "Yes, I believe all gloves are the same." The child then has adopted a belief that she herself has a belief that all gloves are the same. Through exchanges like this, a child can come to the more general belief that she herself has beliefs. She may also learn from such events the skill of identifying her beliefs by observing her own behaviour.

Beliefs that are about one's own beliefs—as opposed to being about entities in the world—are deuterio beliefs. My proposal is that such deuterio beliefs are initially derivative. First, others attribute beliefs to a self; secondarily, in accepting what others say, one learns to attribute beliefs to oneself. I call such second level beliefs, “reflections.”

Historically, this development almost certainly depended on language. While attributing beliefs to another individual could occur without language—even gazelles can do it—it is difficult to see how, without the use of language, one could convince another person that they have a belief. Recall, however, that by language I mean culturally normative communication. Since language has developed gradually over millennia, the cultural habit of self-attributing beliefs about one's own beliefs has also come about gradually. Reflection is not all or nothing.

There is no selection pressure on gazelles to attribute beliefs to themselves. In the human social context, however, those who recognize that they themselves have beliefs are often in a better position to negotiate their dealings with others. In societies in which reflection is prized there is selection pressure for developing and preserving reflection.

Some beliefs are enduring: I've believed for many decades that the earth is round, but seldom think about it. Other beliefs occur in the moment: I believe this is a computer screen that I see before me. Once I believe that I have beliefs, I can reflect on the beliefs that change from moment to moment and come to believe—correctly—that I have a stream of beliefs. This is a third level belief: a belief about the beliefs I believe I have. While one could call this stream of beliefs “self-consciousness,” there is a danger that those enthralled by the myth of interiority may misunderstand this stream of beliefs as a stream of events in a mental inner space. Indeed, this misunderstanding may be the origin of the notion that there is a stream of consciousness—a stream of private experiences—in the first place

Rather than being locked away in a private interior space inaccessible to others, one's beliefs, like one's personal norms, are accessible to others. When an author writes a biography of someone, she is not prying into a private stream of consciousness. She is writing about what the person believed, what he lived for, what his values were, how he lived up to—or failed to live up to—the personal norms he espoused. These are the elements that made up his life. While the biographer may need to do some detective work to discover them, they

are not intrinsically inaccessible. Indeed, one's beliefs, like one's character, are often more visible to others than to oneself. Racism, as mentioned above, is an example of beliefs that one may fail to believe that one has. Reflection is not easy, otherwise it would not be a struggle to "Know Thyself."

This doesn't mean that secrecy is impossible. One can lock passwords in a safe. One can burn one's love letters. Likewise, one can pretend to make a contract while secretly intending to break it. Secrecy in this everyday sense, however, is entirely different from the notion of a private interior space accessible only to one's own "inner eye." Deception requires deliberate efforts to control one's speech and actions so that others will have difficulty determining what one really believes. If a self were a private interior space no effort would be needed to keep its contents secret.

World/Umwelt

Beliefs are about the world one lives in. The world that one perceives and acts on is made up of entities and situations that are meaningful. When I see a table it is defined by its possible uses for me: a support for my glass of beer or my book. I do not see wood fibres or molecules. A car is a unified entity that I could own. Ownership makes no sense to a gazelle; it lives in a different world. I see a friend's welcoming wave, not an anatomically defined moving limb. What any organism perceives are entities in its Umwelt. Beliefs that result from such perceptions are judgments about entities in the believer's world.

Humans, like gazelles, can have beliefs about entities defined by biological norms. As language gradually evolves, however, the world humans inhabit is made up more and more of entities that are defined by their linguistic context.

I argued in chapter 4 that the linguistic meaning of an assertion comes from the norms of the linguistic community. If I say, "It is snowing," the assertion is about white stuff falling from the sky because of the communal norm that this is how these sounds are used in English. Even if I had wanted or intended to claim that there are fish in the sea, the assertion I made is still about falling white stuff. I am free to make whatever sounds I want to, but I am not free to determine the meaning of these sounds. As assertions, what they mean is the prerogative of the community. It is the community that determines what an assertion commits

me to—such as the implication that I should say that what is falling is white or that I should perform the action of putting boots on if I go out.

That the meaning of an assertion is communal does not depend on how loud it is. I can shout an assertion so that everyone can hear it, or I can whisper it quietly, *sotto voce*. Indeed, if I don't make any sound at all, but just exercise my vocal cords, or even if it is only the relevant circuit in my brain's language module that gets activated, the meaning depends on communal norms. Silent beliefs about linguistically defined entities are still assertions, and therefore they have the same communally determined meaning as the equivalent assertion uttered out loud. Holding a belief sets norms for what else one should believe—its implications—and for how one should act. If I believe it is snowing, then I should believe that what is falling is white, whether I say it to others or only to myself. If I believe that what is falling is black, then I have violated the communal norm for the meaning of "It is snowing." It is what the belief claims about the world that identifies it as this belief, rather than another. This aboutness is due to social usage even when the assertion is silent.

Because the implications of beliefs are communal, others can typically attribute beliefs based on actions. If others see Mary putting on her snow boots, they assume she believes it is snowing. They may be wrong—perhaps Mary is just trying them on for size. When the attribution is right, as it often is, then the belief others attribute to Mary has the same meaning that Mary's assertion would have if she said it out loud. In practice, it would not be unusual for people to remark, "Mary must be saying to herself that it is snowing." Beliefs about entities in a linguistically defined world are governed by the same communal norms as assertions about them.

Defenders of interiority may dispute this by pointing out that in the case of lying there is a difference between what one says publicly and what one really believes. They analyze the difference in terms of two facts: what I utter in the exterior world is one fact; the presence of an inner private idea is another fact. It is this second fact that constitutes the real meaning of my belief. When I lie, they say, these two facts do not correspond. In my alternative analysis, which rejects an interior basis for meaning, the difference is normative: what I actually say—the lie—is one thing; what I should say is another. It is what I *should* say that is my real belief. We don't need to hypothesize a private mental space to explain lies.

Those who treat beliefs as inner objects in the mind—as facts—are unable to explain how beliefs are about the world. A fact is not *about* anything; it just *is*. Aboutness requires normativity. My claim is that the meaning of silent assertions—of beliefs—is just as dependent on the communal norms for how assertions should be used as is the case for assertions proclaimed aloud. A self lives in a self-system which forms the context within which entities in its world as well as beliefs about them are meaningful. A belief is not a fact in a self's mental space; it is a component of the self's relationship to its world.

The world a self inhabits is a primarily social world, that is, the entities in it are communally defined. The meanings of one's beliefs are therefore not under one's exclusive control. They are not transparent: one may not always fully understand what one believes.

A child might use the word “elephant” as just a name for any large animal. Adults understand that what it *really* means includes the presence of a trunk. So if the child believes there is an elephant in the zoo, it doesn't quite know what it believes. A belief is not an object that carries its meaning on its surface. Like an assertion, a belief is a cooperative entity whose specific meaning depends on others.

It is not just children who rely on others to determine what an assertion is about. For instance, in 2020 many adults came to believe, on the basis of media reports, that the Russian protestor, Alexei Navalny was poisoned by novichok. Despite holding the belief, however, some people had no idea what novichok was. They relied upon others, perhaps journalists, to determine the meaning of the word “novichok.” Probably the majority of people with the belief didn't know that novichok is a nerve agent, specifically a cholinesterase inhibitor. One could say that they didn't know what they believed, or that they had no idea what the word referred to. This approach, however, treats the meaning of a belief as a purely individual matter. If, as I am maintaining, what a belief refers to is a cooperative matter for the community, then the claim that most people didn't know what they believed is misleading. All linguistic meanings are out of the hands of individual selves. Relying on neuroscientists to know “for us” what novichok means is perfectly legitimate. We're all in this together.

The norms that govern the meanings of our assertions and beliefs, therefore, typically are not the private property of individual selves. As a result, our beliefs are often initially vague and as we learn more we can make them more precise. As a child, or an adult, learns

more, they bind themselves to new norms for the use of their assertions and beliefs. In the case of novichok, the norms are established by the more specialized community of neurophysiologists. To come to understand that novichok is a cholinesterase inhibitor is to bind oneself to the norms of that community. It is not just legitimate to rely on others in the community to set norms for the use of beliefs: it is essential. As a communal enterprise, the meaning of beliefs is distributed. I can only assert or believe, “Novichok is a nerve agent,” if my linguistic community has set norms for the meaning of this phrase.

Second order beliefs—beliefs about beliefs—are no different in this respect. The meaning of my reflective belief, “I have the belief that Novichok is a nerve agent,” depends on my community having a norm for when, “So and so has a belief,” is meaningful. Similarly, the belief, “I have a stream of beliefs,” is possible only within a community that has norms for self-attribution of beliefs.

Those who accept the “myth of interiority” think that one can simply, by oneself alone, introspectively observe that one has a stream of consciousness. I am claiming that this is a mistake. One can, however, come to believe that one has a stream of beliefs, but such a conclusion cannot depend on a solitary self. It is a cooperative position that would make no sense unless one relies on others for the meaning of such a higher order belief.

I am not saying that such a reflective belief is imposed on one by others. The belief is one's own. What I am claiming is that a precondition for any individual self coming to this belief is a linguistic community that sets norms for what it means in the first place.

Experience

But what about experience? Isn't my experience uniquely my own independently of others? Not quite! Experience is meaningful: what it means depends on one's beliefs. Some people—“traditional empiricists”—hold that what we experience are “sense-data”—patches of visual colour or simple shapes—and it is these raw sensations that cause beliefs. This position is confused. There are indeed causal processes that enable experience and these involve neural firings and brain processes, but we cannot experience *them*. What we experience are things in our world – snow, apples, \$20 bills, friends and so on.

Say there is a green apple before me. The light from the apple stimulates the green-responding cones on my retina, and neural signals get transmitted to my visual cortex and higher brain areas. Imagine that something goes wrong along the way, so that I come to believe there is a red rose in front of me. Maybe I've taken some hallucinogenic drug; or perhaps there is some miswiring in my brain; I might have been hypnotized. Whatever the cause, if I come to believe that what I'm seeing is a red rose, then it is a red rose that I experience. A neuroscientist objectively investigating the event might insist, "What you are experiencing is a green apple." I can respond, "I'm the one who knows what I'm experiencing, thank you very much, and it's a red rose." Who is right?

The conflict arises because the phrase, "what is experienced" is ambiguous. It could be used to refer to the objective cause of my experience—a green apple—or it could refer to what the experience means for me, the experiencer—a red rose. In this second sense, what I experience is what I believe I experience. It is the second sense that is important for the life of a self: it is experience in this second sense that is remembered by a self, that enters into its narrative and could contribute to its habits and its character.

Even when things go "right," however, what is experienced depends on the current makeup of the self. A self is not a blank slate on which a stream of raw, meaningless sense-data impinge. A self is structured by its historically developed habits and character, by its cultural and personal norms. What it experiences is conditioned by all these factors. Within the self-world system, the kinds of entities that a self can experience are those that match its perceptual and cognitive capacities, including language. Experience is not the start of the story; it is the latest stage in a self's relationship to its world.

Consider the following scenario. I go to the airport to pick up a friend. She is an entity in my world. I know what she looks like: identifying her is a skill I have already learned. I am primed to spot my friend. My friend-recognizer processes are on high alert. They may even mistakenly get triggered by strangers. When she herself eventually arrives, my friend fits into a meaning-frame already prepared for her. I do not experience a mass of sense-data and thereby synthesize an entity called a friend, as empiricists theorize. Rather, the meaning of what I perceive is pre-established by my framing of the situation even before any sense-data hit my eyes. The belief that there is a friend in front of me is prepared, ready to go. It fits into

my life, into my narrative. Once I come to believe that it is my friend I am seeing, then it is her that I experience.

What I am arguing is that experience is meaningful; its meaning is what it is about in the world. More precisely, its meaning is what the experiencer *believes* it refers to in the world. It is in this sense that experience depends on belief. It is what one believes that determines what an experience means. In the absence of belief, an experience would be meaningless, that is, it would not be an experience.

One might object that an illusion is an experience without belief. People experience the sun rising even though they do not believe that the sun is rising; they believe the earth is turning. Hence, there can be experiences that are not beliefs.

I think this objection is mistaken. Consider the Fata Morgana illusion. Under some atmospheric inversions, a ship appears—sometimes upside down—floating in the air near the horizon. In the past, some sailors took such a “ghost ship” as a harbinger of doom—the Flying Dutchman. Note that they could not take it as an evil omen unless they first believed that there was a ghost ship in the air. Later sailors, more schooled in science, who believe that a ship hanging in the air would violate the laws of physics, are faced with conflicting beliefs: they must either reject the belief that there is a ship in the air as false or revamp their beliefs in physical science. So the objection is a misinterpretation: illusions like these are not conflicts between beliefs and experiences. They are conflicts between beliefs. That there is a ghost ship in the air is not a beliefless experience. If the light coming from the horizon did not lead to a belief in a floating ship, there would be no conflict in the first place and hence no illusion.

Illusions show that not all beliefs are equal: some are less credible than others. For the most part, new beliefs that conflict with one's set of accumulated, well established beliefs are marked as illusory or false – though one must remain open to changing one's mind.

Note that a belief marked as false does not get expunged from one's memory. A striking event such as a Fata Morgana illusion is likely to remain in one's narrative and one may well tell others over the years about the belief one formed, especially because it was false. Past beliefs that one currently holds to be false may well have contributed to one's historical narrative. The struggle to reconcile past beliefs that conflict with one's current beliefs is often central to the development of one's character.

Not all conflicts of belief, however, are noticed. To notice that two beliefs are in conflict is to form a deuterio level belief about them. It requires that one already attributes the two beliefs to oneself and then examines them for coherence. Coherence is the norm for a unified self. Faced with a recognized conflict, one should evaluate at least one of these beliefs as false. Such an evaluation, however, requires that both beliefs be mine. Hypnosis, for instance, is a dissociated state. A hypnotic belief is not compared with the ensemble of my beliefs to see if it fits with the other beliefs I have embraced. That's why it is not *mine*.

So is a conflicting belief not mine? Mineness is normative, not factual: my beliefs *should* not conflict with each other. It is one thing for two people to have differing beliefs; quite another for a single self to harbour such conflicts since they challenge the unity of the self. Everyone has some conflicting beliefs, so the unity of the self is always a work in progress. If there are too many incoherent beliefs then self-identity is in jeopardy. In the extreme we might have a case of Dissociative Identity Disorder (Multiple Personality Disorder).

Coming to believe that one's beliefs are in conflict is a higher level belief. This deuterio level of beliefs is what I've been calling reflection.

Reflection

Reflection is the forming of beliefs about one's own beliefs. Initially, in childhood, others attribute various beliefs to me on the basis of their observations of what I say and do. If I accept their attributions, then I come to believe that I have those beliefs. Later, I develop the skill of observing my own behaviour and attributing beliefs to myself. I may observe not only my overt behaviours and assertions, but even my silent assertions or my inclinations to actions. On this basis, I can learn to attribute beliefs to myself.

Reflection is not introspection, if by that term one means looking at objects in an inner space transparent to oneself. Reflection is observing one's assertions and actions and thereby attributing beliefs to oneself, much as others would attribute them. Unlike the notion of passively seeing an inner idea in one's mind, reflection requires effort. It is a kind of detective work: one works back from clues to the belief of which these clues are implications. Like detective work, reflection is far from certain. Beliefs are not transparent to the self. One can be surprised to discover what one really believes. Often other people can grasp what we believe better than we can do ourselves.

Reflection is not an easy skill to master, so some people are better at it than others. One can fail to believe one has a belief—a racist belief, for instance—even though others can see that one holds it. One can mistake what one believes: someone might believe that they believe their father is admirable, but under psychoanalysis realize that actually they believe he is despicable. Reflection is a life-long endeavour, not something achievable once and for all.

The most important belief for being a self is the general, second-order, belief that one has beliefs. A gazelle believes many things about its world, about lions and thickets, but it doesn't believe it has beliefs. Unlike a gazelle, a self does have the higher order belief that it itself has beliefs, indeed a stream of such beliefs. The capacity for such self-reflection is an essential feature of selfhood.

There is a temptation to call this higher order belief “self-consciousness,” but this would be a dangerous move since it can lead to a misinterpretation. Some people think of beliefs as factual objects in a non-physical space, rather than as normative states. They misinterpret the stream of beliefs as a private “stream of consciousness” and see it as independent of any linguistic or other communal context. It is just there all by itself. This error is at the origin of the myth of interiority that my account of beliefs is refuting. The better way of putting it is to say that selves have a stream of beliefs, in principle visible to others, and a second order reflective belief that they have these beliefs.

The capacity for reflection is not always activated. A self, in its daily life, is primarily concerned with the world around it. Most of the time, one focuses on worldly projects, making and fulfilling commitments, communicating with and relating to others and so on. In attending to the calls of the world, a self is preoccupied by its pragmatic perceptions and actions. The habitual nature of these activities means that they can be performed without reflection. While its character, the norms it has adopted, and the commitments it is facing get their meanings from their context, especially their temporal context, a self tends to take this situation for granted. For the most part, a self lives in the present. Its norms and beliefs are obvious and unquestioned. One seldom thinks about how a \$20 bill has monetary value, what the norms of

one's language are, or why one is committed to one's norms. Contexts, historical or contemporary, remain largely unnoticed.

This obliviousness of context promotes the assumption that entities are self-contained: the bill is simply, in itself, \$20; the meaning of a linguistic assertion is just built into it; a belief is a fact in a mental space. Without reference to its context the self too gets mistaken for an object. Perhaps it is thought of as physical, as part of the brain. Those who consider the physical option to be implausible may ask, "What more is there to a self than the body?" If they think only in terms of objects rather than contexts, they may conclude that the self must be a spiritual object. It is the ignoring of the historical and communal context that leads to the misconception of the self as a mystically self-conscious mind, that is, to the myth of private interiority.

It is this natural, everyday assumption that a self must be some kind of object-in-itself that my account is rejecting. The self, I claim, is a normatively defined entity. Only within the self-system—the historical and contemporary contexts that confer meaning on its components—can there be selves and the worlds they inhabit.

Chapter 8

Are Selves Real?

What I am saying, however, is that the self is an illusion. The sense of being an ego, an I, a thinker of thoughts in addition to the thoughts. An experienter in addition to the experience. (Sam Harris)⁵¹

Is the self real, or is it an illusion? To answer this question we need not only an account of what a self is, but also an account of what it is for anything to be real. So far this book has focused on the nature of the self as a normative entity unified over time. What we require now is an understanding of the nature of reality.

There are many ways to understand what we mean by “real.” If someone is infatuated with me, I might well wonder if their love is real. When someone gives me a \$20 bill I might be suspicious that it is counterfeit, that is not a real \$20 bill. If I feel hot, could that be just a subjective impression or is it a real fever? Does reality mean the same thing to an amoeba, a tick, a pigeon or a human self?

First I will discuss the nature of reality which I have been presupposing. Unlike the mechanistic notion of things-in-themselves, the reality of entities, I claim, depends on context. What I mean by *context* will therefore be explained next. Since systems form the context of entities, the following section will try to pin down the nature of a *system*. These analyses prepare us for an explicit definition of one of the most central concept in the book: *normativity*. Finally, I will reflect on the *status* of my whole discussion: it does not claim to be absolute truth from a God's eye view: it is a historically situated account by a self that is formed by evolved culture and language.

Reality

“Reality” is not a term with a self-evident, fixed meaning. When people question the reality of selves, however, they tend to bring to the question a preconception, a model of what it is to be real: to be real is to be an object, a thing-in-itself. Materialists, for instance, call something real if it is a piece of matter, something solid. The image is of hard billiard balls that

bump into each other. These motions are governed by universal causal laws: one ball impacts on a second and causes it to move. This is the image of reality held by those I have been labelling mechanists. It is often presupposed by those who declare the self to be an illusion.

This notion of reality arose in the seventeenth century and continues in some popular understandings of science even today. The notion, however, has many problems. Actual billiard balls made of marble appear solid. By the eighteenth century, however, it became clear that such balls are far from solid. They are made up of molecules (and later, atoms) with space between them. Only the atoms are “a-tomic:” the only hardness and solidity is at the microscopic level. The solid materiality of the billiard ball is an illusory appearance to human perception.

In the twentieth century we learned that atoms themselves are not solid objects but are made up of subatomic particles: protons and electrons. Then we discovered that even these are just appearances of muons, quarks, and other sub-nuclear components. Finally, (that is, for the moment), quantum theory tells us that matter is “really” made up of mathematical probability wave functions. These are as far away from hard solid chunks as one could imagine. The wave function collapses into objects only in the presence of an “observer” – though it remains unclear whether the “observer” is itself only a probability wave. Apart from the disappearance of “solid matter,” the project of explaining reality by means of deterministic natural laws governing cause and effect has been replaced in quantum theory by non-deterministic, probability-based models of explanation. The materialistic, causal billiard ball image of reality is like a mirage that evaporates into immateriality the closer we look at it.

The mechanistic world view has also been undermined by Einstein. Mechanistic billiard balls have their own intrinsic mass, shape, and velocity. The theory of relativity, however, dismisses the intrinsic nature of these qualities. Far from being properties of the things themselves, their quantities depend on the velocity of an observer. Even the mass—the matter itself—turns out to be interconvertible with energy. The effect of gravitational force, that epitome of a cause of motion, disappears as objects move without force along the shortest space-time line instead. In relativity theory not even the distinction between space and time remains absolute.

It is not only the mechanistic notion of a thing-in-itself that is problematic. The idea of a moment in time, a temporal point, is just as incoherent. Punctualists think of time as a flow of events that are absolute in themselves. A point in time is analogous to a thing-in-itself in that it is what it is independently of everything else. In Einstein's relativity theory, however, the passage of time depends on the velocity of the observer. Even more counter-intuitive is that what is a temporal interval to one observer may be a spatial distance to another. There are no events-in-themselves, but only events for an observer. That, at a particular moment, a mass has a specific velocity is not a "fact," if by fact we mean how the world objectively is: a "fact-in-itself." In-itselfness is as problematic for points of time, events, and facts as for objects. Events are no more defined by an internal essence than objects.

Our common sense idea of reality shares with mechanism the notion that a thing-in-itself makes sense. Whether we are scientists or just common-sense folk, we assume that what a thing is depends on itself alone. When considering an atom or a chair, we naturally think of a self-contained object that would remain what it is if transported to another galaxy. Even if everything else in the universe were to magically vanish, the solitary chair would still be a chair and a carbon atom would still be a carbon atom. Common-sense takes it for granted that each thing is what it is regardless of its context.

We come by this thought honestly. Perception and thought have evolved for a purpose: survival. Brains and nervous systems evolved to help organisms predict what is likely to happen next and decide what to do about it. Organisms are designed to be pragmatic, not to be impartial observers or speculative scientists. An amoeba's detectors detect food-for-amoebas, not carbon atoms or quarks. The temporal history of the amoeba species accounts for the functional definition of food, but the detectors detect food; they do not detect that history.

The same is true for all evolved organisms, including humans. Our eyes process what is useful about our world, not ultraviolet input or radio waves. If reality is quantum probability waves, our senses are oblivious to them. We see, not what is "real" in this scientific sense, but what is useful. And what are useful are stable, medium-sized objects—tigers, bananas, chairs, and \$20 bills. So it is natural that we live in a world of objects that appear to us as things-in-themselves. Like all organisms, we are "thing-oriented." The anthropologist, Ruth

Benedict, says that culture is as invisible as the air we breathe. It is not just traditional culture, however, that is hard to see; the context that defines the objects of any organism's world is, for pragmatic reasons, almost always ignored.

It is not surprising, then, that selves—one's own self and other selves—are taken for granted as just given. We do not attend to the contexts that define selfhood. As evolved entities concerned with our pragmatic dealings with our world, we assume that selves are context-free entities.

What is a Context?

This book sets out to suspend this evolved pragmatic attitude. By suspending this natural approach, we can see beyond not only the dogmas of mechanism, but beyond the even more fundamental idea of things-in-themselves. Once we step back from our natural attitude, we can come to realize that things are what they are only in their context, not in themselves. Without the context of monetary exchange, a piece of paper is not a \$20 bill. Without the context of a language, a sound is meaningless. Without their evolutionary context, an amoeba's organs are not food detectors. Without social norms of responsibility and temporal, narrative-based commitments, bodies and brains are not selves.

What I am maintaining is a general philosophical position about the nature of reality: a being is what it is because of its context.

Expressing this position is difficult. Because of its pragmatic role, language gives a privileged role to nouns. Nouns refer to objects, but even when we talk about events or actions we name them with nouns as if they were objects. Such objectification places a particular hurdle before us when we try to say what a context is. Despite "context" being a noun, contexts are not objects. A context is a set of relations between things, not just one more thing among others.

Imagine for comparison someone who, having learned that in English the sounds "The cat is on the mat" is grammatical, and that "Mat cat on" is ungrammatical, then asks whether the English language is grammatical or not. She would simply be confused. The "English language" is not itself an assertion: it is the context within which sentences are grammatical or not. So the question is out of order.

It is the same kind of mistake that leads one to ask what kind of object a context is. A context is not an object of any kind, despite being named by a noun. A context is a set of relations between the entities that are defined by that context.

Not being an entity or an event, a context cannot be a cause. It doesn't *do* or *make* anything. It is not an agent. When we say that it is the monetary exchange context that makes a piece of metal into \$1, we must not misinterpret "makes" in a causal manner. The Mint manufactures disks of metal, but the exchange system does not manufacture dollars. It is true that without the exchange context, the coin would not have any monetary value, but this must not be understood as a claim that the exchange system *causes*, in the mechanistic sense, the metal disk to be money. Being money is a normative, not a causal issue. Within the exchange system the metal takes on its monetary value because it *should* be exchanged for 100 cents.

In the case of dollars it is the present context that defines an entity or event. For selves, on the other hand, the defining context is a temporal one. For an analogy, consider a melody. The musical value of each note comes, not from its absolute frequency, but from the context of the previous (and anticipated) notes. Moving the melody up or down to a different key preserves the melody. The note's identity is the role it is supposed to play in the ongoing temporal flow of the melody. The frequency of the sound waves does not define the musical role just as chemical processes do not define biological functions. It is their temporal contexts, not the underlying physics, that define the notes and their musical value. Similarly, for selves: it is not brain processes that make up a self. It is the context, communal and temporal, that defines selfhood.

In the case of temporal contexts, the future contributes to the significance of the past. In the narrative of one's life, the meaning of an event depends on how the future unfolds. An early marriage, or a change of religion, may later be interpreted as a great success or as a dismal failure. As Aristotle put it, "Call no man happy until he is dead." The normative judgment that defines life events is always open to future reinterpretation. The significance of the first cod anti-freeze doesn't become clear for millions of years. This is not, of course, a causal relationship. I am not claiming that the future *causes* an effect in the past. It is the defining of what normative events or entities *are* that is spread over time when the context is a temporal one.

Temporal systems are dynamic. They are in flux. What appears at any moment is a time slice, a cross-sectional view, not a final resting point. The normative entities at a particular moment are the crystallizations of the past, but they are also the seeds of the future. The amoeba's food detector has its function at one stage of evolution, but in retrospect it can be seen as the beginning of the elaborate perceptual systems of mammals. The detector is to the future what its own evolutionary history is to its present.

In the case of an individual self, the temporal context is crucial. It is the self's past—its narrative, commitments, character formation, reputation, beliefs and so on—that defines its present being and its current world. One's current relationships, one's fulfilling of promises, the meaning of one's actions, and one's modes of perception are what they are only in the context of one's previous life and future projects.

It is only in this precise sense of context that one should understand my position about the nature of reality: entities are not things-in-themselves but are what they are only in their context, whether that context is current or temporal.

This contextual account of reality allows us to answer those who ask whether selfhood is real or only an illusion. Studies of cognitive science and of the brain do not unearth any object one might call a self, so some have made the claim that selfhood is an illusion produced by neural activity. It is hard to figure out precisely what this could mean. Typically, an "illusion" refers to a self perceiving something that doesn't exist. A person thinks they see an oasis when in reality it is a mirage. But if a self is not real, to whom does it appear as an illusion? If there is no real self to be deceived, how can a self suffer an illusion?

Perhaps the claim is just badly expressed by the term illusion. Rather than being about false perception, maybe the claim is a theoretical one: the theory that there are selves is false. Selves are unreal, like ghosts or witches. If, like mechanists, we think *reality* refers to independent hard objects, like billiard balls, then selves cannot be *real*.

The billiard ball way of looking at reality, however, is of limited use. It excludes not only selves, but cells, amoebas, and elephants from the realm of reality. To this approach, social entities such as nations, wars, and the English language count as unreal. The difference between a valid \$20 bill and a forgery vanishes: both are unreal. Far from clarifying the situation, the billiard ball concept of reality leads to paradox and confusion when applied

to many things, including selfhood. If instead we use “real” in my contextual sense of reality, then selves are just as real as biological cells, as organisms like gazelles, and as social objects such as parliaments or universities. The question, “Are selves real,” turns out not to be so much about the status of selves; it is primarily about how we use the word “reality.”

Systems

What is “real” depends on the system that contextualizes it. This approach is often referred to as “holism.” Holism maintains that systems set the contexts for their components and what they are. A whole is not just a collection of its parts: What parts are real depend on the whole they are parts of.

Our common sense thing-focused approach assumes that relations as secondary to things: first there are things; secondly they enter into relationships between them. My holistic claim is the opposite: systems are primary in that it is only within a system that entities can be what they are in the first place.

The most crucial feature of a system, as I am using the term, is that its components only are what they are in the context of the system. In contrast, the thing-focused approach assumes that systems are made up of pre-existing or independent parts that are initially things-in-themselves but secondarily cluster together to form larger wholes. A dining room suite, for instance, can be formed by putting together assorted, pre-existing tables and chairs. This approach, however, doesn't work for contextual entities. A chess set, for instance, cannot be understood by how wood fibres group together into rooks, queens, and knights. A molecular analysis of the parts gives no insight into what makes a rook a rook. It is only in the context of the game of chess that anything could be a rook. A language is not composed of independent words that a linguistic community accumulated over its history. Without the linguistic community words would be meaningless; they would not be words at all, only sounds. The notion that, in evolution, first there are bees and, independently, there are flowers and that later they form a relationship, misses the systemic relations required for either to exist.

My holistic claim is that selves, and indeed all organisms, must be understood as parts of a system. It is not that entities first come into existence and are then dropped into the world to establish relationships with other entities. Evolution is not simply the origination of

new species by chance mutations. Amoebas can only appear in parallel with the Umwelt that evolves with them. Ultraviolet perceiving bees require flowers with matching pigments. Cultural hominins would make no sense without a cultural tradition to define them. Over thousands of years the traditional group conformity system gradually morphed into a new system in which selves and the social institutions of their world developed in parallel. A self as a structure of binding commitments can only evolve within a community that holds individuals responsible for their actions and beliefs.

A self is a new kind of real entity, but it is not a prefabricated thing, like a billiard ball, that gets parachuted into society. It is real in the way that the parts of any system are real. The notion of a self makes sense only in a social system in which other selves hold individuals responsible for their assertions, beliefs, and actions. This is a general principle: entities are what they are because of their relationships to other entities within a system that contextualizes them.

Evolution moves slowly, so when I refer to a system I am talking about a structure that endures over a period of time. Such a system has a certain stability. It is not simply a slave to momentary outside forces; it maintains itself by determining what external influences to accept and how to respond to them. The homeostatic stability of the system of single cell organisms, discussed in chapter 1, lasted for billions of years. Tribes of early hominins and their cultural worlds formed relatively stable systems. A language is a system that maintains a certain identity by absorbing words or grammatical structures from elsewhere and giving them its own spin. English, for example, imported the French word “centre” by modifying it so that it was pronounced “center,” in accordance with the phonetic style of English. The self-system, of course, is my most central example of an enduring system. Any system survives by stabilizing itself in the face of external influences.

Of course, it may not survive. The dinosaur system was unable to survive the asteroid collision 65 million years ago. Most languages have lost their integrity and gone extinct over many thousands of years. Evolved systems are not necessary structures with transcendent destinies: they come to be under certain chance historical circumstances and evolve into other systems—or disappear—when circumstances change. While they exist, however, they

maintain themselves by ensuring that eruptions beyond the system impinge on them “only with respect.”

The self-system evolved out of cultural conformity in an analogous way. There was not some sudden discovery (by the Romantics, for instance) that all humans are created equal as responsible and free individuals. Instead, literary, legal, religious, educational, economic and many other factors developed in parallel with the gradual change of conformist individuals into selves who commit to their own personal norms. The self-system involves a self capable of binding itself to norms and a social structure within which being-bound is a possibility, a recognized status. Neither selves nor their worlds could appear without each other. It is the system as a whole that evolves.

As systems evolve, they establish normative contexts for their components. Some writers who accept the notion of systemic wholes are still tempted to misinterpret the whole's relationship to its parts as causal. Donald Campbell, for example, proposes that we think of a biological system as *downwardly causing* effects on its parts.⁵² In the case of the sunflower—my own example—they might say that the flower as a whole downward-causes various subordinate biochemical processes to occur in such a way that the sunflower turns towards the sun. A traditional causal mechanist, of course, would claim that the only true explanation of the event is that biochemical mechanisms within the flower upward-cause the plant to respond to light by turning towards it. But such reductionism sounds like a miracle! Why would the chemical mechanisms be ordered in such a way that this biologically useful behaviour occurs? The notion of downward-causation is introduced to explain such miracles. Rather than the biochemical mechanisms upward-causing the behaviour of the flower as a whole, they say the system as a whole downward-causes the subordinate process as effects.

I think this is an unhappy way to put it. Causal explanation, in the precise technical sense in which I've been using the term, relates an event, the cause, to a regularly resultant event, the effect, in accordance with a universal natural law. The ultraviolet pigment in a particular flower might be said to cause this bee to approach it, but it is overstretching the concept of cause to say that the bee-flower system causes the flower to have ultraviolet pigment, or causes anything, for that matter. Similarly, the historically developed self-system

should not be thought of as causing the existence of selves. A system is not itself the kind of entity that could act causally on its component objects.

The confusion arises from a loose use of the term “cause.” Many people use “cause” as a label for *anything* that explains. I think this use results in confusion. As I pointed out in the introduction, throughout this book I am using the term “cause” in the precise, restricted sense that mechanists developed: to explain an event causally is to show that it is the effect of a prior event in accordance with a universal, natural law. The central dogma of the seventeenth-century mechanistic approach is that the only valid way to understand anything is by causal laws.

To understand the relationship of wholes to their parts we do not need to invent a new kind of causality—downward causality. Rather we need to break out of the dogma that all understanding is causal explanation. We can understand a system as establishing the roles that its parts should play without imposing the notion of cause on this relationship. In the English language, for instance, the subject of a sentence and its verb should agree in number. The language as a system, however, doesn’t *cause* this agreement: the system sets the norm that such agreement is grammatically correct. We can understand this relationship without any appeal to causation, at least in the mechanistic sense. Causal explanation works well for billiard balls and planets. It is inadequate for understanding a linguistic system—or the nature of selfhood.

My account of systems abandons the dogma that causal explanation is the only valid mode of understanding. A system is not a cause: it defines meanings. Since selves, like organisms, cultures, and languages, are defined by their norms, their reality cannot be understood causally. The addition of downward-causation doesn’t solve the fundamental flaw: all causal accounts are norm-blind.⁵³

Norms

Selves are normative realities. This claim should be understood against the background of this holistic, systemic account. Entities within a system are defined by the norms of that system. But what, precisely, is a norm? A norm, as the term is used throughout this book, is a technical term, a term of art. I will first mention some ways the term can be misunderstood before offering my positive definition.

Norms are standards for what should or should not happen. For many people the primary model for what should and should not happen is morality: it is morally wrong to cheat or steal, for example. For a number of reasons, however, morality is a poor model for my concept of normativity. First, some consider moral norms to be universal and absolute: if murder is morally wrong for me, it is morally wrong for you and for everyone, in all places, times and cultures.

The normativity of selfhood, however, is not absolute, but contextual. This is true of all the normatively defined entities I have discussed. Amoebas should pursue food-for-amoebas and avoid poison. When I say that the amoeba *should* follow its norms, I am not referring to some absolute obligation, as if a god standing outside the system had commanded amoebas to pursue food-for-amoebas. This norm is specific to the amoeba-system and essentially referred to that system.⁵⁴ Nor is the amoeba-system itself absolute: there is no necessity that such a system come into existence. It is a contingent development that originates in the evolutionary history of the amoeba species, in a history of chance mutations and environmental circumstances.

Besides, if we think of norms as like moral norms, then use of the concept of norm to understand the evolution of selfhood would be wildly anachronistic; analyzing cells, amoebas, and their worlds by norms would be ludicrously anthropomorphic. Moral norms are more or less conscious; they are explicit and frequently couched in linguistically expressed rules. They therefore bind only responsible selves. Even domestic animals are not obligated by moral norms, much less amoebas. Moral norms are inescapably linked to evolved selves, so if we are to employ the concept of norm in earlier evolution, we need to free ourselves from the moral sense of the term and give an account of the nature of norms that does not rely on our contemporary moral intuitions.

Language offers a better model than morality for how to think about norms. Far from being absolute or universal, each particular language has come about by chance, historical events. Each is a system that sets up roles for the units of the language: the phonemes, words, and sentences. Phonemes are combined into words and sentences in accordance with syntactic and semantic standards unique to each language. Languages are holistic: it is not that a language encounters a set of pre-given words or sounds onto which it then imposes

an order. There are no words independent of language. The elements of a language are defined by the language as a whole system.

Thinking of the norms of a language can help free us from exclusively moral uses of the term “should.” There is communal pressure to speak “correctly” and this pressure can be expressed by statements like, “you should say ‘I am’ not ‘I is.’” But this “should” is not a moral one. Violating linguistic norms is not immoral.

Language also illustrates the nature of systemic norms. In actual speaking, linguistic rules are often broken. The linguistic structure is not a description of how the sounds are actually used: rather it lays down how the sounds *should* be used. A language sets ideals for each phoneme, but everyone's pronunciation is a little different. Similarly, actual speakers frequently violate syntactic rules and use words with meanings that vary from the semantic standard. The norms of a system are not a description of what actually happens: they are about what *should* happen.

The norms that define selfhood, like the biological, and cultural norms I have been discussing throughout this book, have these same characteristics: they are non-moral; they derive from a system that controls its inputs; they are not absolute, universal, or preordained, but come about by evolutionary chance; they point to what should happen, not what actually does. And these norms make sense only in the context of the system.

So I can now explicitly define a norm: a norm is the relationship of an entity to its constituting context. A norm is the role an entity should play in the system that defines it.

A normative entity is one whose nature derives from its contextual system. I don't mean that pre-existing things-in-themselves become normative when they take on a role within a system. I don't mean that the same thing could in principle be moved to a different system and become governed by an alternative norm. My claim is more fundamental: *What an entity is* is the role that it *should* play within the system. The norms define the very essence of the entity. It is the norm that constitutes its being.

A norm is not an extra object over and above the entity it defines. It is not a superficial add-on, like a coat of paint on an already existing object. A norm does not add something extra to an entity. Rather, talking about a norm is a kind of reminder: it draws attention to the

context within which the entity is defined, a context that may be forgotten when we focus on the entity itself in isolation.

This is the positive take-away from the language paradigm: it is the role that a word should have in English that constitutes it as this English word. Similarly, an entity that is food-for-amoebas is not a neutral, value-free object: it is a normative entity. If it wasn't that which should be pursued, it would not be food-for-amoeba. This *should*, this norm, derives from the food's role within the context of the amoeba-system.

A norm is not a cause: it does not *do* anything in the mechanistic sense. When I say that it is a functional norm that makes a bunch of chemicals into a heart, someone committed to the dogma that all understanding must be causal explanation would be tempted to interpret this "making" as the norm causing the chemicals to be a heart. This sounds like magic! Does a norm hover over a chemical structure and miraculously convert it into a heart? No! A norm is not a mystical object that causes anything as its effect.

The difference between defining the being of an entity and mechanistically causing it can be understood by considering a pocket calculator. A calculator displays " $2+2=4$." If broken, or badly programmed, it might output " $2+2=5$." There are causal processes in the electronics of the calculator that explain the outcome in either case, but these mechanisms do not explain why " $2+2=4$ " is correct and " $2+2=5$ " is wrong. The rightness or wrongness depends on the norms of arithmetic. The calculator does not establish the rules of arithmetic. If the calculator is working well, the electronic mechanisms enable it to follow these norms. If it fails to live up to the arithmetic norms, we throw it out as useless.

When a child learns arithmetic, their brain gets "programmed" to assert that " $2+2=4$." If the child asserts " $2+2=5$ " more than, say, half the time, they fail the course. The arithmetic norm, however, that $2+2=4$, not 5, is not derived from the child's brain. The child's neural processes, the causal mechanisms, *enable* them to get the answer right, to fulfill the norms. The neural processes, however, do not define the rightness or wrongness of the response.

Similarly, for the norms that define the self. Consider, again, the nature of promising. Society sets a norm that promises should be fulfilled. There are culturally learned brain processes that enable a self to follow this norm by fulfilling promises. If those processes do not result in promises being fulfilled most of the time, the community will no longer trust the

individuals and so their attempts at promising will fail. But the success or failure, indeed even the norms that promises should be fulfilled in the first place, can no more be explained by the brain processes than the mechanism of the calculator can explain why $2+2=4$. Causal mechanisms do not help us to understand norms.

This is true even for the simple biological norms of a single-cell organism. There is a homeostatic norm for saline concentration set by the cell as a unified entity. Causal processes in the membrane succeed or fail to maintain this norm. The homeostatic norm, however, is not itself an object that causes the concentration. An amoeba's food detector works by some chemical mechanisms, but it is not these mechanisms that make it into a *food detector*: that is a normative role defined by the amoeba-system. An early hominin's brain may cause vocal sounds or bodily gestures, but whether they have linguistic meaning or not is up to the culture's norms. The movements of passing a piece of paper to another person are caused by brain processes, but that the movement counts as the fulfillment of a promise to repay \$20 depends on communal and personal norms.

Norms are relationships that bind, not objects that cause effects. If norms were objects—norms-in-themselves—we might be puzzled about why they should be followed. If it were written in the stars that one should fulfill promises, any individual could ask why they should be bound by such a norm. Norms, as I am using the term, are not like rules written on stone tablets handed down from outside. They are relationships between components of a system. That an amoeba's food detector is bound by the norm that it should detect food is not due to some external imposition: if the detector were not bound by this norm, then it would not *be* a food detector. Norms are not external impositions: entities are only the kinds of beings they are in so far as they are bound by their defining norms. Puzzlement about how norms bind frequently stems from thinking of norms as a second order of extraneous objects that cause effects, rather than thinking of them as defining relationships. Norms define selves: it is not that selves initially exist and secondarily bind themselves to norms. It is the binding to norms that qualifies an individual as a self.

Norms are also non-causal in another sense. A norm cannot be causally imposed. Even an all-powerful god couldn't create me in a "promised" state. A hypnotist might cause me to believe I made a promise, but the belief would be false. Unless I committed myself to

the promise in the past, I am not in a promised state in the present. A god could cause an amoeba to pursue food, but a god cannot make it that the amoeba *should* pursue food. Only the history of species evolution that sets up the amoeba-system can bring about that contextual functional norm.

Unlike causality, normative binding is not deterministic. Norms are not inevitably followed. Not all promises are kept. Amoebas sometimes fail to pursue food. Occasionally one might exchange 19 dollars for a \$20 bill. This doesn't mean that normative binding is some kind of probabilistic, non-deterministic causality. It is not any kind of causality. However, while normative binding is not deterministic, it does require that the norms be followed enough of the time. What is "enough" depends on the entity. Only a very low failure rate is needed for a pocket calculator to be broken, to be no longer a "calculator." On the other hand, a self must have a high number of linguistic errors before they are dismissed as speaking gibberish.

While being bound by a norm is not deterministic, it is also not a magical relationship. Binding describes how entities are and the way they behave. How binding actually gets carried out in individual cases depends on some causal mechanism or other. There are biochemical mechanisms that transport sodium in the membrane of a cell. That's how the homeostatic norm is achieved. Without some such mechanism, homeostasis would not be maintained, and the cell would cease to exist. The norms of English are not neurological: meaning comes from the linguistic community. Yet there have to be brain mechanisms that enable any individual speaker to follow English norms. In principle, silicon processes could be used as a linguistic prosthesis to replace the brain's wetware. Such a linguistic prosthesis is a long way off, but its logical possibility reminds us that linguistic capacity must be understood as the ability to be bound by communal norms, whatever causal processes bring that about.

Causal mechanisms enable an entity to be bound by norms, but its status as an entity does not depend on *which* mechanisms do the enabling. Causes cannot explain norms. Electronic causes cannot explain why $2+2=4$. Brain mechanisms do not explain the norms of a language. Neural processes do not explain why a promise should be fulfilled. Brain states do not determine what a belief is about; only norms can do that. Selves are normative, not mechanistic entities.

To summarize my position, norms are the roles that components of a system should play within their system. A norm is not an isolated, atomic, thing-in-itself. Nor is it a mysterious or transcendental reality added to things-in-themselves. Norms are not a second set of realities added to or paralleling non-normative facts. Norms are not magical causes: they are not causes at all. Norms are how a systemic context defines what entities within the system are, that is, their reality.

Answering the question, “Are selves real?” requires defining clearly what “real” means. If reality refers to in-themselves, context-free, objects—billiard ball style—then selves are indeed illusions. But that concept of reality also implies that money, words, and organisms are not real. Indeed, on close examination, it also declares atoms, electrons, and, indeed, billiard balls themselves, to be illusions. This concept of reality leads us nowhere. It is a dead end.

My alternative is to use “real” to refer to normative entities defined by their systemic contexts. With this understanding of reality, tables, dollars, cells, and elephants are real. Normative selves are real in the same sense.

Status of my Account

This contextual understanding of reality has implications for the status of my account. My project has been to understand what a self is and how it could have come about. The task has a peculiar self-reference in that only selves can understand. An amoeba cannot understand what a self is. It cannot even understand what an amoeba is. It is not *into* understanding. A self, in contrast, can set out to understand itself and its world. Such a project is possible only because selves are evolved entities that exist in a particular historical period and in a world that parallels their capacities. Understanding is therefore always itself contextual: it cannot be seen as a search for the Absolute Truth.

For those who think of reality as made up of things-in-themselves, it is tempting to imagine a grasp of such reality that is context-free. Descartes proposes that we have access to a God-given “Natural Light” that we can rely on to deliver Truth. Such a grasp would require a viewpoint beyond the confines of biological, cultural, or linguistic limitations. A God's Eye View like this presupposes a viewer that transcends time, evolution and history, and that can perceive and think about reality-in-itself, reality that is what it is without reference to any

context. A God's Eye View and the notion of things-in-themselves are two sides of the same coin.

The coin, however, is counterfeit. There is no Archimedean point from which we can view reality from the outside. All beliefs, true or false, require a believer. As we've seen, linguistic beliefs are commitments attributed to a self by a linguistic community. Without the evolutionary story I have presented—or something like it—there can be no selves and so no such beliefs. All selves come into existence within a community that holds them responsible for their commitments. Without the self-system—selves and the Umwelt in which they live—there can be no account of anything.

Denying that there is no Archimedean point should not be interpreted as relativism. Typically, this term is used as a label for a claim that what is true for one person might be false for another. Broccoli tastes delicious to me, but it may be horrid to you. Taste is subjective. The contextual understanding of reality is not subjective relativism in this sense. That a dollar is worth 100 cents is not subjectively relative. A person who claims a dollar is worth 90 cents is wrong: their claim is objectively false. This is not a matter of subjective taste. The objects, the dollar and the cents, set the norm for truth or falsity. In the absence of the monetary exchange context, the claim that a dollar is worth 90 cents could not turn out to be true: in the absence of the context, there would be no dollars or cents in the first place. The point about contextualism is that the context defines what entities are, and therefore sets norms for the truth or falsity of assertions about them. It is precisely the context of norms that undercuts subjective relativism. In the absence of norms for truth, it would make no sense to claim that what is true for one person is false for another. Nothing could be true for anyone.

My theory of selfhood, however, *does* mean that my account of selfhood is not absolute. The account I am offering is an account of the nature of selves by a self whose understanding of selfhood and of evolutionary and personal history is a product of this very history. My account is itself contextual. This is not unique to my account. Any account, including one that disagrees with mine, must be held by responsible selves that have evolved over billions of years and that have learned from their culture how to adopt responsible commitments and beliefs. Amoebas do not hold a theory of selfhood; they cannot hold any theory. Conversations with parrots go nowhere! The very notions of account, of explanation,

and of understanding depend on evolved selfhood. Despite the dreams of mechanists, causal explanation is no exception: it too consists of linguistic beliefs—whether true or not—whose existence requires selves with their contextual situation. Even the theory of evolution is not an account from the outside: the theory itself can exist only in the context of evolution.

What I have laid out is what I believe is the best way of understanding who we are as selves embedded in time, in history and in tradition. Since our culture, our language, and our status as selves are functional products of our history, without this history we could not be who we are. In particular, I could not be writing this book—its “words” would be meaningless—and you could not be reading and interpreting it. Any account we offer depends on our community's “best lights,” not on any absolute “Natural Light.” Without evolved selves, educated in a culture, and equipped with language, there can be no theories whatsoever. Understanding is a situated endeavour. There is no God's Eye View, no Archimedean point: there are only us selves, contingent products of evolutionary chance, endeavouring to understand selfhood.

Chapter 9

Selfhood: Past and Future

My central point in this book is that a self is a normative entity that is unified over time, so let me recap this position.

I have argued that a self's self-identity cannot be understood as thing-like: the unity of a self over time is not the continuity of a body, a soul, or the cognitive processes of a brain. Temporal self-identity comes from the relationship between earlier events in which a self makes commitments and its current status of being bound by them.

These temporal commitments explain "ownness." An experience is *mine* in so far as it takes its meaning from the context of my life-history. This history is not just a chronicle of one event after another, nor is it simply a stream of memories. The current self is characterized by being bound to norms that originate in its prior commitments. It is the relationship of current norms to past commitments that constitutes the self's unity over time. It is due to this relationship that the action that fulfills the commitment is one's *own*.

This account overcomes the intellectualism of the punctual theory of the self. A self is not left isolated with its own current thinking as its only resource. Having a character means a self is not abandoned to the instant: its historically developed habits support it in its current decisions and actions. A self is emotionally inclined to live up to the habitual norms to which it has bound itself. A self does not live in an isolated silo: its norms are embedded in a social context. Having a reputation means that the expectations of others, the norms they expect a self to live up to, are essential elements in the life of a self. If its reputation has been well established, it mirrors the norms that make up the character of the self. It is through its character and reputation that a self relates to its world.

The world that a self inhabits is made up of entities that are meaningful to it—Umwelt entities. Beliefs, true or false, are about these entities: they are commitments to how the world is. What one believes is perceivable by others. Second level beliefs about one's own beliefs—

reflection—include the belief that one has a stream of beliefs, a stream that is often misinterpreted as a series of private inner events.

To be a self, in the normative way I am using the term, is to be an evolved entity that faces its world with a set of habitual personal and social norms. These norms are the context within which perceptions and beliefs should be interpreted, and actions should be evaluated. This context is temporal: its current normative system is an embodiment of the self's cumulative individual history and its culture's tradition.

Is there a "True Self"?

In my introduction I embarked on this analysis of selfhood by quoting the adages, "Be thyself," "To thine own self be true," and asking, "Why is that important? What does it even mean?" How could one not be oneself? History is replete with admonitions about finding one's True Self, escaping from self-alienation, or even overcoming the illusion of being a separate self. Mystics have proposed that one should transcend the everyday self and merge with God, Atman, or some ultimate source of being. Some contemporary psychologists urge therapy to overcome the "false" self. These adages and proposals come from so many different intellectual frameworks that finding a common meaning for "*true self*" is a hopeless task. The notion of selfhood I'm offering probably matches none of these frameworks.

My account, however, does distinguish the self as the set of norms to which one is committed from the factual individual who may or may not live up to these norms. It is a feature of all normative entities that they can succeed or fail to live up to their defining roles. The English language, for instance, is a set of norms, but what is factually spoken often varies from them. Selfhood is no different.

A self may fail to live up to its norms for a number of reasons. Brain pathology can prevent one fulfilling one's commitments. Even a pocket calculator can malfunction and say that $2+2=5$. Yet even if there is no hardware failure, selfhood is a learned status and, like most learning, is incomplete. Children often fail to fulfill commitments. As we mature, we get better, hopefully, but learning to be reliable is always a work in progress.

Another reason for failure is that, like beliefs, norms may be in conflict. A father might have a long-term norm of honesty, but in committing himself to supporting his child in a

dispute—a short-term norm—he might lie to protect her. In retrospect, he might feel guilty that he had not lived up to his lifetime ideal.

Noticing these kinds of failures could lead a person to feel that some particular actions do not reflect their real, enduring character. In a moment of reflection, they might notice the discrepancy and vow to be more faithful to the norms they have embraced. Perhaps occasions like this have led some people to speak of a “True self.” This analysis, however, presupposes my account of normative selfhood and is unlikely to cover the many different meanings of “Be Thyself” that are based on very disparate notions of the self.

When did Selfhood come about?

Selfhood is an evolutionary development. Without billions of years of biological and, later, cultural and linguistic evolution, selfhood could not have come about. Particular social conditions, as I explained in earlier chapters, are required for there to be unified entities responsible for their own norms, beliefs, and actions. Some writers have proposed specific historical moments for the arrival of such unified selves.

Julian Jaynes suggests that prior to about 3000 years ago humans hallucinated voices that they interpreted as the gods telling them what to do—the “bicameral mind”—a condition we now label psychotic. When people came to own these voices they became unified selves, which Jaynes refers to as the origin of consciousness.⁵⁵

Others date the arrival of selves to Ancient Greek philosophy starting with Socrates. Some of these philosophers challenged the notion that it is the gods who are responsible for the actions of individuals. Plato, for instance, held that souls should leave the shadowy cave of conformist opinion and justify their own beliefs in the bright light of reason. Could this be the moment when the idea of unified, responsible selves emerge?

Christianity, some think, emphasized the values of individuals before God. In particular, the Reformers in the sixteenth century preached that each person had their own individual relationship to God. Could selfhood be an offshoot of Christianity?

The Humanism of the Renaissance, and above all, of the Enlightenment, placed human persons at the peak of the universe, and some assume that it is these events that are the most fundamental factors in the development of selfhood.

Throughout this book I have insisted on gradualism and continuity, so from my point of view, there is little sense in trying to establish a specific date for the advent of selfhood. The periods mentioned above may have contributed to selfhood norms, but looking for *the* point when selfhood comes into existence is a fool's errand. Even if, as I have proposed, language is needed for the reflective, self-attribution of beliefs, we must remember that language, as communication based on communal norms, starts early in the evolution of hominins and develops gradually. The social conditions for selfhood do not arrive suddenly, out of the blue, but in many steps. Nor is the development one-directional and linear. A period favourable to selfhood may be followed by an unfavourable one. Even in favourable periods, only some in a society may be treated as self-responsible. Ancient Greek philosophers, for example, lived in a society based on slavery and on the subordination of women. Selfhood is fragile.

Will Selfhood Survive?

As a fragile phenomenon, the future of selfhood is not assured. It is not obviously in the interests of capitalist corporations that individuals commit to their own norms. In a consumer society the ideal consumer is one whose norms conform to whatever is profitable for producers. The goal of advertising is to manipulate consumers so that they do what the corporation wants, rather than what they want themselves. Military training and political propaganda have a similar aim. As advertising becomes more scientific and effective—think of neural research for advertising—one could imagine a future dystopian society in which everyone's values are manipulated by impersonal organizations so as to reduce rather than enhance individual autonomy. Some aspects of social media point in this direction. Kingwell asks, “Are we, with the energy of technological changes, moving into a new moment of human existence where the quest for selfhood, understood as the creation of stable personal identity, is over?”⁵⁶

Making individuals into cogs in the machine—that is, eliminating selfhood—might be to the advantage of the machine and could conceivably increase the fitness of the human species, though I am not advocating it! But how about the possibility that selfhood could be bad for individuals themselves? Ruminating on the past or anticipating the future can lead to anxiety and unhappiness. Gautama Buddha proposes a solution: he claims that only the present moment is real. The idea that there is a self as an enduring entity that is self-identical

in its past and in its future is an illusion. This illusion is the cause of all suffering and prevents individuals from achieving the ultimate happy state of living entirely in the present, the state of non-self, of nirvana. So could it be that selfhood is an aberration that we would be better without?

It is possible that Buddha is thinking of the self as a contextless object, like a billiard ball, and that it is this conception of selfhood that he is rejecting as illusory. My account, of course, also rejects any object-like notion of the self. The Buddhist declaration that a self as an in-itself object is an illusion does not undermine my claim that contextually defined selves are real entities.

This is not the place for a serious analysis of the many strains of Buddhism. Let me just ask for whom the state of non-self is valuable. In the absence of selfhood there may be no suffering, but who is around to benefit from the resulting happiness? Furthermore, in the absence of a self, there is no one to believe that the state of non-self is superior to selfhood.

In any case, I am not claiming that selfhood has an absolute value. My claim rather is that all norms and all values are contextual. Within a system entities have roles they should play. Selfhood is one of these roles and the self-system context sets norms for how selves should fulfill their roles. Within this context entities can succeed or fail, can be more or less successful. Outside such a context, questions of value or of success or failure are meaningless. As I claimed in chapter 1, a single cell organism sets norms for saline concentrations within it, but the organism does not set norms for whether the origination of cells should count as a success or a failure. The self-system is a context that sets norms for the value of selfhood. In the absence of a self-system, the question of the value of selfhood would be moot.

My account, therefore, should not be thought of as a defence of human exceptionalism. There is no Archimedean point from which one form of life can be declared as absolutely more valuable than other forms. Amoebas have value within the amoeba-system. Selves have value within the self-system. If the amoeba-system disappeared, there would be no basis for deeming the extinction of amoebas a failure. If selves, such as you and I were to fantasize about the extinction of selfhood, we might consider the event tragic on the basis of

our norms. Were the self-system to disappear in reality rather than in imagination, however, there would be no norm against which the extinction of selfhood could be considered a failure.

How might Selfhood Survive?

Evolution does not stand still. Like all evolved entities, selves might become extinct. More likely, they will evolve further. With the arrival of selves, evolution has not reached some final destiny, some ultimate stopping point. How might selfhood evolve in the future?

Normative entities, as we have seen, are defined by their contexts, not by their underlying processes. The exchange value of a dollar—which is what defines it—does not depend on whether it is made of silver, paper, plastic, or bytes. In chess, it is irrelevant whether a rook is made of wood, marble, or pixels on a screen: what defines it is the role it is supposed to play in the game. So far, selfhood has always been realized in a biological organism and has relied on brain and organic processes to be able to follow its defining norms. Is it possible that selves, could, in the future, be enabled by computational—silicon, digital, or quantum—processes?

Ray Kurzweil, in 1990, pointed out that digital computing power has been growing exponentially, and that computers would soon be capable of more calculation per second than human brains. In fact, by 2022, supercomputers have more than four times the brain's speed. (While the brain's speed has been estimated at 20 to 100 million billion calculations per second, the Fugaku supercomputer can manage over 400).⁵⁷ Speed, however, is not enough: only when we learn what computational processes the brain uses, will it be possible to duplicate them digitally. Then, In principle, by replicating the neural processes of an individual's brain a self could be transferred to a silicon body and live on.⁵⁸

Even the tasks of human programmers, Kurzweil points out, could be taken over by computers. The time will come when computers will be faster at designing computer programs than humans. From that moment on—the Singularity—computers will be able to outpace humans. Just as *Homo sapiens* out-competed earlier hominins, the future of evolution will be in the hands of computers.

Kurzweil's scenarios should not be dismissed as science fiction. Already we have figured out enough about how hearing works to be able to manufacture artificial cochlear

implants that are already in use by many. Experimental artificial retinas are close to ready for use. Note that such implants do not “simulate” the neural processes of hearing or vision: users really hear and see. Recent analyses of the neural processes of the hippocampus have allowed computer researchers to mimic the operation of a section of it made up of hundreds of thousands of neurons.⁵⁹ Since the hippocampus is a brain area required for memory, this development is of particular interest for selfhood. Duplicating the whole brain in silicon is still far in the future, but these developments indicate that the project is technically possible.

Interpreting these developments has led to much debate in recent decades. Those who think that it is intelligence that defines humanity ask whether Artificial Intelligences (AIs) are really intelligent, or just gimmicky imitations. The colloquial term “intelligence” is used loosely for many kinds of human activity, but if it is defined more precisely as the ability to solve problems, then AIs are clearly intelligent—indeed for some problems more intelligent than humans.

Others question whether AIs could be conscious. What is meant by “consciousness,” however, is even more fraught than intelligence. Some assume that consciousness refers to a mental, private state. But if we understand consciousness in this way, the question is in principle unanswerable. We cannot even know whether any human is conscious in this sense. So asking whether AIs could be conscious is a dead end.

A better question is to ask whether an AI could be a self. Given my account of selfhood, the question amounts to asking whether an AI could follow norms. The answer is an easy “Yes.” The hardware of a computer is designed to allow it to run software programs. Programs are normative: they dictate what the hardware *should* do to achieve its goal. In actual fact, the hardware may succeed or fail to carry out the programmer's instructions. Since the same software can be implemented on different hardware, like all normative entities a program has a certain independence from the causal processes that run it. A program is more than the movement of electrons through semiconductors. Computing is normative from the get-go.

So whether there could be “artificial” selves boils down to the question whether computers could take responsibility for following the norms that define selfhood.

A pocket calculator is supposed to be programmed by its manufacturer to respond to “2+2?” by displaying “4”. If the device is programmed incorrectly, the responsibility is the programmer's.

Als, however, are more sophisticated. Programmers do not dictate to chess computers, such as AlphaZero, what strategies to use to win chess games. The basic rules of chess, which pieces can move what way and what counts as checkmate, must be programmed. These are like the Als DNA. In machine learning, however, programmers do not dictate strategies. Rather, they program the computer to learn, that is, to figure out for itself what the best strategies are. The programmers themselves don't know what strategies the AI will come up with. Indeed, the only way the programmer's can discover the computer's strategy is, like its chess competitors, to observe how it plays.

If we ask, then, who is responsible when AlphaZero wins a match against a human chess master, the answer cannot be the programmers, whose skill at chess is inferior to that of a chess master. The ability to win—or lose—the game must be attributed to AlphaZero. Whatever programmed strategy AlphaZero uses is a strategy it has itself devised. In its learning, AlphaZero has programmed itself.

Selves, I have claimed, have some responsibility for their own character, for the norms they have adopted. If I am a dishonest person and fail to pay the debt I have promised, I cannot blame my DNA. The self I have become, while requiring DNA, is due to commitments I have made in the past. It is this learned character that results in me, my self, being responsible for my actions. If my actions were totally determined by my DNA, I would have no responsibility for them, I would not be a self. My personal history, the norms I have learned to adopt, is what makes me a self.

Following this same kind of analysis, we can say that AlphaZero's winning of a chess match is its own responsibility, not that of the programmers. Although the initial learning program came from the programmers—its “DNA”—the actions that enabled it to win the game came from its own history, its own learning. The strategies it used are AlphaZero's own norms for what its moves should be.

We can say the same for other Als. Contemporary translation Als have not been told how to translate sentences. They have learned by being trained on many examples. The

result, good or bad, is therefore the AI's responsibility. Some AIs have been trained to diagnose diseases, and can often do better than human doctors. If they get it wrong on occasion, and someone dies, the responsibility is the AI's, not the programmers'.

A chess computer can actually move pieces, not just inform humans what to move next. The AI can perceive a situation and act appropriately. Just like a human, or any other organism, what an AI can perceive depends on the sensors it is equipped with. Its sensors give it access to entities in its world. In other words, the AI lives in its own Umwelt, and its states are meaningful ones: they are about the entities in its world.

An AI, therefore can be said to have beliefs about its world. A diagnostic AI might conclude that a patient has cancer. Such a conclusion counts as a belief: it is a state that refers to a situation in the world, a state that could be true or false. An AI could also form deuterio beliefs about its own beliefs. During reinforcement learning, it may be informed that its diagnosis of cancer was incorrect. It could be useful for its learning, for improving its diagnosis in the future, for it to take account of the falsity—or truth—of its prior conclusion. It could then be said to have a belief about its own belief.

So am I saying that AlphaZero, and other AIs are selves? This is not a yes or no issue. I have claimed that there are degrees of selfhood. Young children, for instance, are to some extent selves. At what age we declare them fully responsible for their actions is relatively arbitrary. AIs, so far, are very limited: the worlds of entities they can respond to, and their repertoire of beliefs about them are therefore meagre. A chess computer is a specialist AI: it may form beliefs about what its opponent will do next, but it has no beliefs about the weather, the value of a dollar or about promises. Such items are not part of its world.

An AI with a richer repertoire of actions and beliefs, comparable to that of humans—call it a generalist AI—could perceive and act on the entities that make up our human world. Such an AI could not only play chess: it could diagnose diseases, perform surgery, converse intelligently with humans, successfully translate languages, and so on. That is, it would live in the same world as we do.

So perhaps we can say that AIs are on the way to selfhood. If generalist AIs become flexible enough to follow the norms for the vast arena of beliefs and actions that human selves can manage, then we could admit them into the community of selves.

Our initial question was *could* AIs become selves. For those who think that to be a self involves a spiritual mind or a mysterious consciousness unexplainable by science—the myth of interiority—it is impossible in principle for computers, however powerful, to ever become selves. My normative account of selfhood, on the contrary, leaves open the possibility of digitally-based selves.

Apart from the issue of AIs becoming selves, there is the reverse question of whether selves could become AIs. Kurzweil predicts the imminent transition of current human selves into computers. His timeline, however, is not very realistic. Besides, the shift is more likely to be gradual, with many intermediate stages. Indeed, it is already happening. Often electronic calculators have taken over many of the arithmetic capacities that our brains had only a generation or two ago. Some of our memory functions—calendar, contacts, spelling—have been outsourced to smartphones. Computer games can sometimes be controlled directly by brain impulses. So there has been a gradual shift of some brain functions to computers. We are already cyborgs.

Note that if selves someday came to rely on computing machines rather than biological brains for their existence, they would still be products of evolution. A biological heart is defined by its evolved function. An artificial heart relies on different, non-biological mechanisms to carry out its function. The function, however, remains the same—to pump blood. The situation is similar for selves. Selves are contextual entities, defined by the norms of their personal, historical and contemporary communal contexts. A shift from a biological infrastructure to a computer-based mechanism would not undermine these contextual defining features. As normative entities, selves would still be contextually defined by their evolutionary history.

The advent of selves was not predetermined. They are not the final destiny towards which life has been travelling since the beginning. To assume that we are the final goal would be hubris. We came into being through billions of years of evolution and myriads of chance events. What transpired could not have been predicted. Our one certainty is that life is never static, but where it is going next cannot be foreseen. Current selves are just a temporary resting point in an ongoing saga. What the future holds for selfhood is unpredictable.

1. Andrew Pohorille and David Deamer, "Self-Assembly and Function of Primitive Cell Membranes." *Research in Microbiology* 160, no. 7 (2009): 449-56.
2. Ulanowicz labels the process "autocatalytic mutualism." Robert E. Ulanowicz, *A Third Window: Natural Life beyond Newton and Darwin*. Templeton Foundation Press (2009): 76.
3. As Merleau-Ponty puts it, "Between the exterior world and the living organism, there is an insertion of a whole that orders, coordinates and interprets ..." Maurice Merleau-Ponty and Dominique Séglaard, *Nature: Course Notes from the Collège de France*. Northwestern University Studies in Phenomenology and Existential Philosophy. Evanston, Ill: Northwestern University Press (2003): 171.
4. "Here's the point: a creature's ecology must not be confused with its environment. The environment that creatures live in is common to each and every one of them—it's just 'the world.' ... By contrast, a creature's ecology consists of whatever-it-is-about-the-world that makes its phenotype viable. That's to say: it is constituted of those features of the world in virtue of which that kind of creature is able to make a living in the world. In effect, the notions 'environment' [sic] and 'phenotype' are interdefined." Fodor, Jerry and Massimo Piattelli-Palmarini, *What Darwin Got Wrong*. New York: Farrar, Straus and Giroux (2010): 140.
5. Ibid., 143.
6. Marina von Uexkull (Translated by Joseph D. O'Neil), *Foray into the Worlds of Animals and Humans : With a Theory of Meaning*. Minneapolis, MN, USA: University of Minnesota Press (2010):179.
7. Morten Tønnessen, *Umwelt Transition and Uexküllian Phenomenology*. Estonia: Tartu University Press (2011): 57.
8. Flowering plants first appear about 130 million years ago. Some explanation like this is the most likely, although the co-evolution of flower perfume and insect scent detectors is also a possible factor.
9. Francisco Varela, Evan Thompson and Eleanor Rosch, *The Embodied Mind: Cognitive Science and Human Experience*. MIT Press (1991): 202.
10. "The archaeologist Dr Ceri Shipton from the Australian National University ... said [recent research] made a good case for *Homo erectus* and several more species of hominin and other animals emerging at a time of a drying climate 2.3-2 million years ago. ... 'This fits with the idea of our genus being adapted to the savannah and in particular exploiting the big game that is available on grasslands, which they would then butcher using stone tools,' Shipton said." <https://www.theguardian.com/science/2020/apr/03/earliest-known-skull-of-homo-erectus-unearthed-by-australian-led-team>.
11. Jennifer A. Parkinson *et al.*, "Meat on the Menu: GIS Spatial Distribution Analysis of Bone Surface Damage Indicates that Oldowan Hominins at Kanjera South, Kenya had Early Access to Carcasses," *Quaternary Science Reviews* 277 (2022) 107314.
12. A. Agam *et al.* "Estimating Temperatures of Heated Lower Palaeolithic Flint Artefacts." *Nature Human Behaviour* (2020). <https://doi.org/10.1038/s41562-020-00955-z>. See also: <https://www.sapiens.org/archaeology/hominins-fire-tools/>.

13. Russell B. Cuttsa, *et al.*, "Thermal Curved-fragments: A Method for Identifying Anthropogenic Fire in the Archaeological Record." *Journal of Archaeological Science* 106 (2019): 10–22
14. Wenda Trevathan, "Birth, Obstetrics and Human Evolution." *International Journal of Obstetrics and Gynaecology* 109 (November 2002): 1199–1206.
15. Wrangham claims that domestication has a genetic component. Richard Wrangham, *The Goodness Paradox*, Vintage (2019): 28.
16. RIM Dunbar, "Neocortex Size as a Constraint on Group Size in Primates." *Journal of Human Evolution* (1992): 469–493.
17. Michael Tomasello, *A Natural History of Human Thinking*. Harvard University Press (2014): 36.
18. Andrew Whiten *et al.*, "Emulation, Imitation, Over-imitation and the Scope of Culture for Child and Chimpanzee." *Philosophical Transactions of the Royal Society B* (2009) 364: 2417–2428. [doi:10.1098/rstb.2009.0069](https://doi.org/10.1098/rstb.2009.0069).
19. "Handaxes were first made by our ancient ancestors, members of the hominin family about 1.76 million years ago, as part of the Acheulean tradition toolkit of the Lower Paleolithic (a.k.a. Early Stone Age), and they were used well into the beginning of the Middle Paleolithic (Middle Stone Age) period, about 300,000–200,000." "Acheulean Handaxe: Definition and History," ThoughtCo.com.
20. "In contrast to an Oldowan tool, which is the result of a fortuitous and probably extempore operation to obtain one sharp edge on a stone, an Acheulean tool is a planned result of a manufacturing process. The manufacturer begins with a blank, either a larger stone or a slab knocked off a larger rock. From this blank he or she removes large flakes, to be used as cores. Standing a core on edge on an anvil stone, he or she hits the exposed edge with centripetal blows of a hard hammer to roughly shape the implement. Then he or she works it over again, or retouches it, with a soft hammer of wood or bone to produce a tool finely chipped all over consisting of two convex surfaces intersecting in a sharp edge. Such a tool is used for slicing; concussion would destroy the edge and cut the hand." *Stone Tool*, Wikipedia. https://en.wikipedia.org/wiki/Stone_tool.
21. K. W. Carsten *et al.*, "Oxytocin Promotes Human Ethnocentrism," *Proceedings of the National Academy of Sciences* 108 (4) (January 25, 2011): 1262-1266. <https://doi.org/10.1073/pnas.1015316108>.
22. Leslie C. Aiello and R. I. M. Dunbar, "Neocortex Size, Group Size and the Evolution of Language," *Current Anthropology* Vol. 34, No. 2. (Apr., 1993): 184-193.
23. R.I.M. Dunbar *et al.*, "Performance of Music Elevates Pain Threshold and Positive Affect: Implications for the Evolutionary Function of Music," *Evolutionary Psychology* 10(4) (2012): 688-702.
24. *Ibid.*
25. Takayuki Nakata and Sandra E. Trehub, "Infants' Responsiveness to Maternal Speech and Singing," *Infant Behavior & Development* 27 (2004): 455-464.
26. "In this regard, it is a remarkable fact that despite decades of research in psychology and neuroscience in which animals have been trained to do elaborate tasks, there is not a single report of an animal being trained to tap, peck, or move in synchrony with an

auditory beat.” Aniruddh D. Patel, *Music, Language, and the Brain*. Oxford University Press (2008): 409.

27. Marcel Gauchet, *The Disenchantment of the World: A Political History of Religion*. Princeton University (1997).

28. The analogy of society and biological organism is, of course, at least as old as Plato. Some worry that the analogy justifies totalitarianism, with the functional parts obeying the head in a way that violates the freedom and autonomy of individuals. This criticism, however, presupposes that the individuals who make up society are already autonomous selves. My claim is that early human groups did not suppress autonomy; there was not yet any autonomy to suppress. Free selves come at the end of the evolutionary process; they are not there in some Hobbesian beginning.

29. M. A. Gorzelak et al., Inter-plant Communication through Mycorrhizal Networks Mediates Complex Adaptive Behaviour in Plant Communities. *AoB Plants*, 7, plv050. (2015). [doi:10.1093/aobpla/plv050](https://doi.org/10.1093/aobpla/plv050).

30. Iain Morley, *The Prehistory of Music*. Oxford University (2013): 313.

31. Ibid. 321.

32. Charles Darwin, *The Descent of Man*. Classic Literature Library ebook: 366. <https://charles-darwin.classic-literature.co.uk/the-descent-of-man/ebook-page-366.asp>.

33. While I am focusing on sounds as the bearers of meaning, the analysis applies to any form of culturally normative language: writing, signing, digitally encoded sentences, flag signalling and so on.

34. “There’s glory for you!” “I don’t know what you mean by ‘glory,’” Alice said. Humpty Dumpty smiled contemptuously. “Of course you don’t — till I tell you. I meant ‘there’s a nice knock-down argument for you!’” “But ‘glory’ doesn’t mean “a nice knock-down argument,”” Alice objected. “When I use a word,” Humpty Dumpty said, in rather a scornful tone, “it means just what I choose it to mean — neither more nor less.” “The question is,” said Alice, “whether you can make words mean so many different things.” “The question is,” said Humpty Dumpty, “which is to be master — that’s all.” Lewis Carroll, *Through the Looking Glass*, Chapter 6 (1872): 364. https://www.sabian.org/sabian_alice.php.

35. I am not, of course, suggesting that any early hominin groups spoke English! Using equivalent English expressions is simply the easiest way to explain the point.

36. This approach to meaning borrows from Robert Brandom, *Making It Explicit*. Harvard University Press. (1998).

37. Noam Chomsky, *Cartesian Linguistics: A Chapter in the History of Rationalist Thought*. New York: Harper & Row (1966).

38. JJ Bolhuis, I Tattersall, N Chomsky, and RC Berwick “How Could Language Have Evolved?” *PLoS Biol* 12(8) (2014): e1001934. [doi:10.1371/journal.pbio.1001934](https://doi.org/10.1371/journal.pbio.1001934).

39. <http://electronics.howstuffworks.com/gadgets/high-tech-gadgets/speech-recognition2.htm>.

40. “As our findings indicate that songbirds have the ability to discriminate the grammatical rules of context-free languages, a cognitive ability previously supposed to be unique to humans, our results cast doubts on what is currently considered to be a unique

characteristic of human language.” (Kentaro Abe & Dai Watanabe, “Songbirds possess the Spontaneous Ability to Discriminate Syntactic Rules,” *Nature Neuroscience* 14 (2011): 1067–1074. [doi:10.1038/nn.2869](https://doi.org/10.1038/nn.2869)).

41. JJ Bolhuis, I Tattersall, N Chomsky, and RC Berwick “How Could Language Have Evolved?” *PLoS Biol* 12(8) e1001934 (2014):1. [doi:10.1371/journal.pbio.1001934](https://doi.org/10.1371/journal.pbio.1001934).

42. Ibid. 1.

43. Ibid. 1.

44. Martin A. Nowak, Joshua B. Plotkin and Vincent A. A. Jansen, “The Evolution of Syntactic Communication.” *Nature*, vol. 404 (30 March 2000).

45. Jean-Jacques Rousseau, *The Social Contract*. Baltimore, MD: Penguin Books. (1968): Chapter 1.

46. Daniel C. Dennett, *Consciousness Explained*. Little, Brown and Co. (1991): 418.

47. “The Myth of the Mental is just this transcendental claim that every way we relate to the world must be pervaded by conceptual, rational, mental activity.” H. Dreyfus (2009): 2.

48. Isaiah Berlin, *The Crooked Timber of Humanity: Chapters in the History of Ideas*. London: John Murray. (1990): 57-58.

49. Descartes understood that, as an immaterial unity, the mind cannot literally be a container with ideas bumping into each other. “Interior space” and “outer world” are metaphors, though what their literal sense is, Descartes never made clear.

50. David Premack and Guy Woodruff, “Does the Chimpanzee have a Theory of Mind?” *The Behavioral and Brain Sciences*, Vol.1 (4) (1978-12): 515-526.

51. Sam Harris, <https://mindfulnessexercises.com/self-is-illusion/>.

52. Donald T. Campbell, “Downward Causation in Hierarchically Organised Biological Systems”. In [Francisco Jose Ayala](#) and [Theodosius Dobzhansky](#) (Eds.), *Studies in the philosophy of biology: Reduction and related problems*. London/Basingstoke: Macmillan. (1974) 179–186.

53. The emphasis on causal explanation since the sixteenth century may well be due to the vision of early modern thinkers that science would give humans control over nature. Francis Bacon, for instance, proclaimed in 1620 that “the aim of knowledge is to benefit human life.” Knowledge of causes, he claimed, allows us to conquer nature; ignorance of causes prevents us from achieving effects. “Human knowledge and human power come to the same thing, because ignorance of causes frustrates effect. For Nature is conquered only by obedience; and that which in thought is a cause, is like a rule in practice.” Francis Bacon, *The New Organon*. Cambridge University Press (2000): Book I, Aphorisms III p. 33.

But not all activity or knowledge has to be valued for its results. A grasp of the relationship of a linguistic system to its components does not give us any more power over individual assertions, nor over the language as whole. Music can be used to promote the sale of consumer products, but this is a kind of prostitution: enjoying music is valuable for itself. The kind of understanding of life and of selfhood I am seeking in this book is not aimed at any kind of technological control. It has a philosophical goal: to enrich our lives by increasing our self-understanding.

54. Merleau-Ponty calls such norms “immanent norms.”

“Immanent norms are thereby inscribed in experiences and function as internal rules that guide experience, and are also constituted and retained within experience, i.e., through repeated interactions with the environment that results in skills or habits, such as in the acquirement of new motoric and perceptual sense/meaning. Basso explains this with a citation by Merleau-Ponty, who states that there is no external unfolding or pre-existing reason or norm to begin with, and there is no external condition of a norm, but only ‘the birth of a norm’ within the interaction of the bodily subject and the world. Cf. Basso, 2012: 168f; cf. Merleau-Ponty, 2012: 62” (Maren Wehrle, “There Is a Crack in Everything. Fragile Normality: Husserl's Account of Normality Re-Visited,” *Phainomenon* 28 (2018): 60n.)

55. Julian Jaynes, *The Origin of Consciousness in the Breakdown of the Bicameral Mind*. Houghton Mifflin. (1976).

56. Mark Kingwell, “Language Speaks Us: Sophie’s Tree and the Paradox of Self,” in Paul Socken (ed.), *The Edge of the Precipice*. McGill-Queen's University Press (2013): 119.

57. “Using visual processing as a starting point, robotics expert Hans Moravec of Carnegie Mellon institute estimated that humans can process about 100 trillion instructions per second (or teraflops) [source: Moravec]. But Chris Westbury, associate professor at the University of Alberta, estimates the brain may be capable of 20 million billion calculations per second, or around 20 petaflops [source: Westbury].”
<https://science.howstuffworks.com/life/inside-the-mind/human-brain/computer-intellectual-ability.htm>.

58. Ray Kurzweil, *The Age of Spiritual Machines*. Viking Press (1999). See also Ray Kurzweil, *The Singularity is Near*. New York: Viking Books (2005).

59. In 2017, the Soltesz lab published a report in eLife on its virtual model of a rat’s hippocampal compartment CA1 replete with 338,740 neurons.
<https://stanmed.stanford.edu/2018fall/neuroscientists-creating-virtual-hippocampus.html>.

Bibliography

- Abe, Kentaro, and Dai Watanabe. "Songbirds Possess the Spontaneous Ability to Discriminate Syntactic Rules." *Nature Neuroscience* 14, no. 8 (August 2011): 1067–74.
<https://doi.org/10.1038/nn.2869>.
- Agam, Aviad, Ido Azuri, Iddo Pinkas, Avi Gopher, and Filipe Natalio. "Estimating Temperatures of Heated Lower Palaeolithic Flint Artefacts." *Nature Human Behaviour* 5, no. 2 (February 2021): 221–28. <https://doi.org/10.1038/s41562-020-00955-z>.
- Aiello, Leslie C., and R. I. M. Dunbar. "Neocortex Size, Group Size, and the Evolution of Language." *Current Anthropology* 34, no. 2 (April 1993): 184–93.
<https://doi.org/10.1086/204160>.
- Bacon, Francis, Lisa Jardine, and Michael Silverthorne. *The New Organon. Cambridge Texts in the History of Philosophy*. New York: Cambridge University Press, 2000.
- Basso, Elisabetta. "On Historicity and Transcendentality Again. Foucault's Trajectory from Existential Psychiatry to Historical Epistemology." *Foucault Studies*, September 17, 2012, 154–78. <https://doi.org/10.22439/fs.v0i14.3904>.
- Berlin, Isaiah. *The Crooked Timber of Humanity: Chapters in the History of Ideas*. Second Edition. Princeton ; Oxford: Princeton University Press, 2013.
- Bolhuis, Johan J., Ian Tattersall, Noam Chomsky, and Robert C. Berwick. "How Could Language Have Evolved?" *PLoS Biology* 12, no. 8 (August 26, 2014): e1001934.
<https://doi.org/10.1371/journal.pbio.1001934>.
- Brandom, Robert. *Making It Explicit: Reasoning, Representing, and Discursive Commitment*. 4. print. Cambridge, Mass: Harvard University Press, 2001.
- Campbell, Donald T. "Downward Causation in Hierarchically Organised Biological Systems." *Emergence: Complexity and Organization* 15 (1974): 179–86.
- Carroll, Lewis, and Charles Santore. *Alice's Adventures in Wonderland: The Classic Edition*. Kennebunkport, Maine: Applesauce Press, 2017.
- Chomsky, Noam. *Cartesian Linguistics: A Chapter in the History of Rationalist Thought*. 3. ed. Cambridge: Cambridge Univ Press, 2009.
- Cutts, Russell B., Sarah Hlubik, Ross Campbell, Jana Muschinski, Pamela Akuku, David R. Braun, David B. Patterson, Joseph J. O'Brien, Ervan Garrison, and J.W.K. Harris. "Thermal Curved-Fragments: A Method for Identifying Anthropogenic Fire in the Archaeological Record." *Journal of Archaeological Science* 106 (June 2019): 10–22.
<https://doi.org/10.1016/j.jas.2019.03.006>.
- Darwin, Charles. *The Descent of Man*. Volume 1 Volume 1, 2011.
- De Dreu, Carsten K. W., Lindred L. Greer, Gerben A. Van Kleef, Shaul Shalvi, and Michel J. J. Handgraaf. "Oxytocin Promotes Human Ethnocentrism." *Proceedings of the National Academy of Sciences* 108, no. 4 (January 25, 2011): 1262–66.
<https://doi.org/10.1073/pnas.1015316108>.
- Dennett, D. C. *Consciousness Explained*. Back Bay Books. Boston: Little, Brown, 1991.
- Descartes, René, and John Cottingham. *Descartes: Selected Philosophical Writings*. Cambridge ; New York: Cambridge University Press, 1990.
- Dreyfus, Hubert L. "Overcoming the Myth of the Mental." In Hubert L. Dreyfus and Mark A. Wrathall. *Skillful Coping: Essays on the Phenomenology of Everyday Perception and Action*. First Edition. Oxford, United Kingdom: Oxford University Press. (2014): 104–24.

- <https://doi.org/10.1093/acprof:oso/9780199654703.003.0006>.
- Dunbar, R.I.M. "Neocortex size as a constraint on group size in primates." *Journal of Human Evolution* 22 (1992): 469–93.
- Dunbar, R.I.M., Kostas Kaskatis, Ian MacDonald, and Vinnie Barra. "Performance of Music Elevates Pain Threshold and Positive Affect: Implications for the Evolutionary Function of Music." *Evolutionary Psychology* 10, no. 4 (October 1, 2012): 147470491201000420. <https://doi.org/10.1177/147470491201000403>.
- Fodor, Jerry, and Massimo Piattelli-Palmarini. *What Darwin Got Wrong*. New York: Farrar, Straus and Giroux, 2010.
- Gauchet, Marcel. *The Disenchantment of the World: A Political History of Religion*. New French Thought. Princeton, N.J. Chichester: Princeton University Press, 1999.
- Goldman, Bruce. "Neuroscience team is building a virtual hippocampus." *Stanford Medicine*. Accessed March 10, 2022. <http://stanmed.stanford.edu/2018fall/neuroscientists-creating-virtual-hippocampus.html>.
- Gorzelak, Monika A., Amanda K. Asay, Brian J. Pickles, and Suzanne W. Simard. "Inter-Plant Communication through Mycorrhizal Networks Mediates Complex Adaptive Behaviour in Plant Communities." *AoB Plants* 7 (2015): plv050. <https://doi.org/10.1093/aobpla/plv050>.
- Grabianowski, Ed. "How Speech Recognition Works." *HowStuffWorks*, November 10, 2006. <https://electronics.howstuffworks.com/gadgets/high-tech-gadgets/speech-recognition.htm>.
- Jaynes, Julian. *The Origin of Consciousness in the Breakdown of the Bicameral Mind*. Mariner Books. Boston New York: Houghton Mifflin, 2000.
- Kris, Hirst K. "Acheulean Handaxe: Definition and History." *ThoughtCo*. Accessed March 9, 2022. <https://www.thoughtco.com/acheulean-handaxe-first-tool-171238>.
- Kurzweil, Ray. *The Age of Spiritual Machines: When Computers Exceed Human Intelligence*. New York, NY: Penguin Books, 2000.
- . *The Singularity Is near: When Humans Transcend Biology*. New York: Viking, 2005.
- Merleau-Ponty, Maurice, and Donald A. Landes. *Phenomenology of Perception*. Abingdon, Oxon ; New York: Routledge, 2012.
- Merleau-Ponty, Maurice, and Dominique Séglaard. *Nature: Course Notes from the Collège de France*. Northwestern University Studies in Phenomenology and Existential Philosophy. Evanston, Ill: Northwestern University Press, 2003.
- Merleau-Ponty, Maurice, and Colin Smith. *Phenomenology of Perception: An Introduction*. Routledge Classics. London: Routledge, 2006.
- Morley, Iain. *The Prehistory of Music: Human Evolution, Archaeology, and the Origins of Musicality*. Oxford, United Kingdom: Oxford University Press, 2018.
- Nakata, Takayuki, and Sandra E. Trehub. "Infants' Responsiveness to Maternal Speech and Singing." *Infant Behavior and Development* 27, no. 4 (December 2004): 455–64. <https://doi.org/10.1016/j.infbeh.2004.03.002>.
- Nowak, Martin A., Joshua B. Plotkin, and Vincent A. A. Jansen. "The Evolution of Syntactic Communication." *Nature* 404, no. 6777 (March 2000): 495–98. <https://doi.org/10.1038/35006635>.
- Parkinson, Jennifer A., Thomas W. Plummer, James S. Oliver, and Laura C. Bishop. "Meat on the Menu: GIS Spatial Distribution Analysis of Bone Surface Damage Indicates That Oldowan Hominins at Kanjera South, Kenya Had Early Access to Carcasses." *Quaternary Science Reviews* 277 (2022): 107314. <https://doi.org/10.1016/j.quascirev.2021.107314>.
- Patel, Aniruddh D. *Music, Language, and the Brain*. New York: Oxford University Press, 2010.
- Pohorille, Andrew, and David Deamer. "Self-Assembly and Function of Primitive Cell Membranes." *Research in Microbiology* 160 (2009): 449–56.

- Premack, David, and Guy Woodruff. "Does the Chimpanzee Have a Theory of Mind?" *Behavioral and Brain Sciences* 1, no. 4 (December 1978): 515–26. <https://doi.org/10.1017/S0140525X00076512>.
- Rosenberg, Karen, and Wenda Trevathan. "Birth, Obstetrics and Human Evolution." *BJOG: An International Journal of Obstetrics and Gynaecology* 109, no. 11 (November 2002): 1199–1206. <https://doi.org/10.1046/j.1471-0528.2002.00010.x>.
- Rousseau, Jean-Jacques, Susan Dunn, and Gita May. *The Social Contract: And, The First and Second Discourses*. New Haven: Yale University Press, 2002.
- Socken, Paul, ed. *The Edge of the Precipice: Why Read Literature in the Digital Age?* Montréal; Kingston: McGill-Queen's University Press, 2013.
- MC Stiner, R Barkai and A Gopher, "Cooperative Hunting and Meat Sharing 400-200 kya at Qesem Cave, Israel." *Proceedings of the National Academy of Sciences* 106 (32) (2009): 13207-13212.
- "Stone Tool." On *Wikipedia*, February 25, 2022. https://en.wikipedia.org/wiki/Stone_tool.
- Strickland, Jonathan. "Will Computers Overtake Humans in Intellectual Ability?" *HowStuffWorks*, March 22, 2010. <https://science.howstuffworks.com/life/inside-the-mind/human-brain/computer-intellectual-ability.htm>.
- Thompson, David L. *Daniel Dennett*. New York: Continuum, 2009.
- Tomasello, Michael. *A Natural History of Human Thinking*. Harvard University Press, 2014.
- Tønnessen, Morton. *Umwelt Transition and Uexküllian Phenomenology*. Estonia: Tartu University Press, 2011.
- Uexküll, Marina von, and Joseph D. O'Neil. *Foray into the Worlds of Animals and Humans : With a Theory of Meaning*. Minneapolis, MN, USA: University of Minnesota Press, 2010.
- Ulanovicz, Robert E. *A Third Window: Natural Life beyond Newton and Darwin*. Templeton Foundation Press, 2009.
- Wehrle, Maren. "'There Is a Crack in Everything'. Fragile Normality: Husserl's Account of Normality Re-Visited." *Phainomenon: Journal of Phenomenological Philosophy*, no. 28 (February 5, 2019): 49–76.
- Whiten, Andrew, Nicola McGuigan, Sarah Marshall-Pescini, and Lydia M. Hopper. "Emulation, Imitation, over-Imitation and the Scope of Culture for Child and Chimpanzee." *Philosophical Transactions of the Royal Society B: Biological Sciences* 364, no. 1528 (August 27, 2009): 2417–28. <https://doi.org/10.1098/rstb.2009.0069>.
- Wrangham, Richard. *The Goodness Paradox*. Vintage, 2019.

Index

| | |
|--|--|
| Aboutness (Intentionality)..... | |
| about a mammoth..... | 74 |
| about food-for-amoebas..... | 25 |
| aboutness..... | 24, 26, 74 |
| aboutness defined..... | 25 |
| aboutness, linguistic..... | 74 |
| as fitness..... | 27 |
| due to social usage..... | 120 |
| implications of a belief..... | 120 |
| is a functional relation..... | 27 |
| not causal..... | 24 f. |
| not factual..... | 26 f. |
| not representation..... | 26 |
| problem for dualists..... | 26 |
| reference to things..... | 82 |
| requires normativity..... | 121 |
| umwelt entities..... | 146 |
| Abraham..... | 92 |
| Acheulean hand axes..... | 49 |
| Anthropomorphic..... | 7, 29, 45, 64, 138 |
| Aristotle..... | 30, 102, 132 |
| artificial heart..... | 18, 58 |
| Artificial Intelligence (AI)..... | |
| AI beliefs..... | 154 |
| AI responsibility..... | 154 |
| AlphaZero..... | 153 f. |
| artificial intelligence..... | 58, 75, 152 |
| brain's speed..... | 151 |
| computing is normative..... | 152 |
| conscious AI..... | 152 |
| final destiny..... | 151, 155 |
| generalist AI..... | 154 |
| neural networks..... | 75 |
| silicon brains..... | 152 |
| singularity..... | 151 |
| Autobiographies..... | 94 |
| Beethoven, Ludwig van..... | 105 |
| Beliefs..... | |
| a cooperative matter..... | 121 |
| animal beliefs..... | 115 f. |
| attribution of belief..... | 117, 120, 122, 149 |
| belief..... | 8, 12, 84, 86 f., 96, 112, 115 f., 119, 122 ff., 133, 144, 146 |
| belief that..... | 115 |
| beliefs are about the world..... | 121 |
| beliefs as normative states not factual objects..... | 126 |
| communal meaning of beliefs..... | 120 |
| conflicting beliefs..... | 124 f. |

| | |
|---|---|
| deutero beliefs..... | 117 f., 122, 125, 154 |
| distributed..... | 122 |
| dualist..... | 115 |
| enduring beliefs..... | 118 |
| lying..... | 120 |
| not transparent..... | 121 |
| racism..... | 119 |
| reflection..... | 12, 84, 118 f., 125 f., 147 f. |
| self-attribution..... | 117, 122, 149 |
| silent assertions..... | 120 |
| stream of beliefs..... | 12, 118, 122, 126 |
| Benedict, Ruth..... | 130 |
| Berlin, Isaiah..... | 105 |
| Billiard balls..... | 6, 129, 133, 137 |
| Bolhuis, Johan J..... | 76, 80 |
| Brain..... | |
| brain..... | 8 f., 35, 51, 57, 63, 71, 75 f., 79 ff., 90 f., 96, 130, 140 ff., 146 |
| brain processes..... | 9, 75, 79, 81, 113, 122, 132, 140 f. |
| brain size..... | 40 f., 52 |
| brainwashing..... | 87 |
| linguistic assertion not a brain process..... | 81 |
| neural process..... | 49, 80, 140, 151 f. |
| social brain hypothesis..... | 41, 51 |
| Buddha, Gautama..... | 149 |
| Campbell, Donald..... | 136 |
| Cartesian..... | 79 ff. |
| Cartesianism..... | 27, 95 |
| Cause..... | |
| causal explanation..... | 6, 136 f., 145 |
| causal mechanism..... | 6 f., 10, 22, 79, 140 f. |
| cause..... | 6 ff., 11, 16 f., 20, 132 |
| cause defined..... | 7, 137 |
| dogma of causal explanation..... | 137 |
| norm-blind..... | 137 |
| value-free..... | 28 |
| Chess..... | 114, 134 |
| chimpanzees..... | 41, 47 f. |
| Chomsky, Noam..... | 74 ff., 80 |
| Codfish..... | 20 |
| Conformism..... | |
| conforming behaviour..... | 56 |
| conformist individuals..... | 136 |
| conformist opinion..... | 148 |
| conformity..... | 46, 56, 63, 99, 107 f., 135 f. |
| Context..... | |
| context defined..... | 131 f. |
| context defines selfhood..... | 132 |
| context not a cause..... | 132 |
| context of entities..... | 128 |

| | |
|--|---|
| contextual reality..... | 18, 29, 37, 58, 71, 74, 85, 133 ff., 138 f., 142, 144, 150, 155 |
| forgetfulness of context..... | 114, 127 |
| future context..... | 132 f. |
| melody..... | 55, 132 |
| not behaviourism..... | 115 |
| Cooperation..... | |
| based on learning..... | 37 |
| between generations..... | 12, 36, 42, 65 |
| challenge from game theory..... | 42 |
| competition vs. cooperation..... | 41 |
| contract theories..... | 43 |
| cooperation defined..... | 36 |
| cooperation vs. competition..... | 42 f. |
| division of labour..... | 56 |
| endorphins..... | 51 f. |
| evolution of cooperative groups..... | 35, 45, 55 |
| free-riders..... | 42, 45 |
| grooming..... | 51 f. |
| group cooperation..... | 11, 37, 39, 50, 66 |
| groups as units of evolution..... | 42 |
| many forms..... | 39 |
| multi-cellular organisms..... | 43 |
| operating together..... | 36 |
| origin of..... | 37, 39 |
| outsourcing..... | 40 |
| religion..... | 53 ff., 63 |
| shamanistic rituals..... | 53 |
| slime mould..... | 42 |
| survival of the kindest..... | 41 |
| Culture..... | |
| actions defined by traditional norms..... | 59 |
| behaviour patterned by culture..... | 47 |
| being a member of a tribe is a normative status..... | 58 |
| campfire as cultural entity..... | 40, 59, 61, 71 |
| categorizes a world..... | 59 |
| child within culture..... | 47 |
| conservative..... | 38 |
| cultural meanings of natural things..... | 60 |
| cultural myths..... | 92 |
| culturally defined entities..... | 60 |
| culture..... | 42, 62 |
| culture defined..... | 37 |
| culture sets norms..... | 103 |
| defines roles..... | 56 |
| hominin individual is defined relationally..... | 63 |
| inherited from ancestors..... | 62 |
| inheriting norms by contagion..... | 49 |
| militaristic culture..... | 108 |
| over-imitation..... | 48, 54 |

| | |
|--|--|
| role of leader..... | 61 |
| story-telling cultures..... | 90 |
| things in culture's world..... | 60 f. |
| tradition..... | 11, 38 f., 42, 49, 55, 59, 62 f., 72, 81 f. |
| Darwin, Charles..... | 10, 15, 23, 30, 33, 68, 106 |
| Deamer, David..... | 13 |
| Dennett, Daniel C..... | 91 |
| Descartes, Rene..... | 8, 10, 26, 70, 80, 88, 95, 99, 112, 115, 143 |
| determinism..... | 7 ff. |
| Deutero defined..... | 99 |
| Dionysus..... | 105 |
| Doyle, Conan..... | 91 |
| Dr. Jekyll and Mr. Hyde..... | 5, 87 |
| Dualism..... | 6, 8 ff. |
| Dualist..... | 9, 26, 36, 87 |
| Dunbar, Robin..... | 41, 51 f. |
| Einstein, Albert..... | 129 |
| Emotion..... | |
| emotion..... | 37, 45 f., 48 f., 57, 61 ff., 67 f., 72, 94, 96, 146 |
| emotion defined..... | 50 f. |
| emotional bonding..... | 46, 51 |
| Emotional contagion..... | 48 f. |
| emotional identification..... | 50, 53 |
| emotional responses..... | 49, 56 |
| non-deliberative behaviour..... | 50 |
| oxytocin..... | 50 |
| primary..... | 50 |
| secondary..... | 50 f. |
| social emotion..... | 50 f. |
| Emulate vs. imitate..... | 48 |
| Enthusiasm..... | 105 |
| Entities..... | |
| enduring self..... | 86, 88, 90 |
| enduring system..... | 103 |
| normative entity..... | 12, 58, 96, 117, 128, 139 f., 146 |
| organism as unified whole..... | 15 |
| unity..... | 13 ff., 19, 22, 30, 37, 51 f., 54 f., 59, 81, 89 f., 92, 111, 125, 146 |
| what entities are..... | 135 |
| Evolution..... | |
| adaptationism..... | 15, 30 ff., 38 |
| biological evolution..... | 62 |
| co-evolution..... | 31 ff. |
| competition vs. cooperation..... | 31, 33 |
| contingency..... | 30, 34 |
| ecological niche..... | 27 |
| evolution..... | 10 f., 15, 20, 22 f., 30 |
| evolution of cooperation..... | 41 |
| evolution of humans..... | 34 |
| group as unit of evolutionary selection..... | 56 |

| | |
|-------------------------------------|---|
| human evolution..... | 35, 37, 39, 45, 78 |
| lactaze..... | 78 |
| mutual enaction..... | 32 |
| preservation..... | 37 |
| simplistic theory of evolution..... | 31, 78 |
| species..... | 30 |
| system as a whole evolves..... | 32 f. |
| unit of evolution..... | 42 |
| Experience..... | |
| amoeba's experience..... | 29 |
| experience..... | 29, 71, 88, 90 ff., 102, 112, 118, 122 f., 128, 146 |
| experience is meaningful..... | 124 |
| experience is what I believe..... | 123 |
| Fata Morgana illusion..... | 124 |
| framing of the situation..... | 123 |
| illusions..... | 124 |
| sense-data..... | 122 f. |
| Fact-value distinction..... | 28 |
| Fodor, Jerry..... | 27 |
| Freedom..... | |
| advertising..... | 87, 149 |
| autonomy..... | 11, 36, 46, 54, 86 ff., 149 |
| binding enables freedom..... | 107 |
| cogs in the machine..... | 149 |
| determinism..... | 104 |
| dissociative identity disorder..... | 87 ff. |
| evil neurosurgeon..... | 86 |
| freedom..... | 5, 7 ff., 12, 36, 56, 85 f., 97, 104 ff., 110 |
| freedom from constraint..... | 104, 107 |
| gradual accomplishment..... | 107 |
| hypnosis..... | 86, 90 |
| inspiration and possession..... | 105 |
| not novelty..... | 108 |
| propaganda..... | 87, 149 |
| romantic idea of freedom..... | 105 |
| slave..... | 36, 86, 104, 149 |
| Freud, Sigmund..... | 94 |
| Function..... | |
| biological functions..... | 11, 17, 56, 63, 132 |
| cod anti-freeze..... | 20 |
| crystallized history..... | 20, 30 |
| dollar as functional thing..... | 25 |
| function defined..... | 17 |
| functional understanding..... | 19 |
| functional unity..... | 17, 19 |
| functionally defined thing..... | 25 |
| heart as functional organ..... | 22 |
| historical context..... | 19 |
| non-anthomorphic notion..... | 29 |

| | |
|-------------------------------------|--|
| normal vs. abnormal..... | 21 |
| not explained by physical laws..... | 19 |
| Gauchet, Marcel..... | 53 |
| Gold, value of..... | 28 |
| Gorzelak, Monika A..... | 64 |
| Harris, Sam..... | 128 |
| Heart..... | 114 |
| Hobbes, Thomas..... | 43, 45 ff. |
| Holism..... | 134 |
| Holmes, Sherlock..... | 91 |
| homeostasis..... | 13 f., 142 |
| Hominins..... | |
| hominin lineage..... | 83 |
| Homo erectus..... | 39, 41, 52 |
| Homo habilis..... | 39 |
| Homo heidelbergensis..... | 44 |
| Homo sapiens..... | 10, 44, 81, 91, 151 |
| Neanderthals..... | 44 |
| Human exceptionalism..... | 10, 115, 150 |
| Humpty Dumpty..... | 70, 74 |
| Identity..... | |
| children..... | 38, 40, 47 ff., 56 f., 66, 92, 154 |
| Dissociative Identity Disorder..... | 88 f., 125 |
| group identity..... | 37, 44 ff., 54 f., 57 f., 60, 63, 65, 92 |
| mineness..... | 87, 89 f., 94, 96, 105, 111, 125 |
| music and dance..... | 51 f. |
| ownness..... | 87 f., 92, 96, 146 |
| prince and cobbler..... | 89 |
| punctualism..... | 88 |
| self-domestication..... | 40, 56 |
| self-identity..... | 12, 85 f., 90, 97, 101, 110, 125, 146 |
| sleep-walker..... | 88 |
| team sport analogy..... | 46 f., 51 |
| tribal groups..... | 38 f., 41, 55 |
| tribe analogous to species..... | 42 |
| Imitation..... | 48 f., 51, 55 |
| In-itselfness..... | |
| axe-in-itself..... | 59 |
| fact-in-itself..... | 130 |
| gold-in-itself..... | 28 |
| meaning-in-itself..... | 70 f., 74 |
| object-in-itself..... | 110, 127 |
| reality-in-itself..... | 143 |
| self as object-in-itself..... | 127 |
| things-in-themselves..... | 23, 29, 128, 130 f., 134, 139, 143 f. |
| Incest..... | 57 |
| Infanticide..... | 40 |
| Internalization..... | 79 f. |
| Jaynes, Julian..... | 148 |

| | |
|--|---|
| King..... | 83 |
| Kurzweil, Ray..... | 151, 155 |
| Language..... | |
| Adamic theory of meaning..... | 69 |
| Alice in Wonderland..... | 70 |
| bird songs..... | 76 |
| body language..... | 66 f., 71 |
| communication..... | 12, 64, 67 f., 73, 76 f., 81, 83, 149 |
| communication defined..... | 64 |
| contextual theory of linguistic meaning..... | 71 |
| continuity with emotional communication..... | 65 |
| contributes to cultural identity..... | 66 |
| cooperative enterprise..... | 66 |
| culturally distinct languages..... | 71 |
| evolution of language..... | 74, 76 |
| gestures..... | 65, 67 f., 71, 141 |
| gibberish..... | 70, 74, 82, 109, 142 |
| governed by cultural norms..... | 65, 67 |
| importance for cooperation..... | 65 |
| inferences of assertions..... | 72, 83 |
| language..... | 35, 45, 63, 71, 76 ff., 83, 93, 109, 135, 138 |
| language defined..... | 64 |
| linguistic brain processes..... | 79 |
| linguistic meaning..... | 68 ff., 77, 79, 81, 83 f., 96, 119, 121, 141 |
| meaning-in-itself..... | 70 f., 74 |
| mentalist theory of meaning..... | 70, 73 |
| non-syntactical language..... | 77 f. |
| objectivity..... | 82 |
| one kind of communication..... | 65 |
| origin of language..... | 78 |
| poverty of the stimulus..... | 75 f. |
| prosody..... | 67 |
| recursion..... | 75 f., 78 |
| rehearsal..... | 80 |
| semantics and syntax..... | 75, 81 |
| signing..... | 69 |
| speech recognition technology..... | 75 |
| strong minimalist thesis..... | 75 |
| tower of Babel..... | 69 |
| truth and lies..... | 73 |
| verbal language..... | 71 |
| vocables..... | 67 |
| wolf coming..... | 72 |
| Laws of nature..... | 6 |
| Learned behaviour..... | 37 |
| Leibniz, Gottfried Wilhelm..... | 69 |
| Lions and gazelles..... | 116 |
| Locke, John..... | 89 |
| Luther. Martin..... | 107 |

| | |
|------------------------------------|--|
| Māui..... | 92 |
| Meaning..... | |
| linguistic meaning..... | 68, 84 |
| mentalistic theory of meaning..... | 79 |
| set of inferences..... | 71 |
| Mechanist defined..... | 7 |
| Memory..... | |
| cobbler's memories..... | 89 |
| concepts of memory..... | 102 |
| habits..... | 79 f., 96, 104 |
| memory unifies the self..... | 89 |
| non-intellectual habit..... | 60 |
| second nature..... | 79 |
| stream of memories..... | 89 f., 93, 146 |
| Mind..... | |
| cogito..... | 8, 95, 104 |
| consciousness..... | 9, 12, 29, 36, 45, 102, 104, 112, 126, 148, 152, 155 |
| immaterial/spiritual mind..... | 9 f., 26, 96, 104, 112 ff., 117, 155 |
| inner eye..... | 113, 119 |
| interiority..... | 112 |
| introspection..... | 125 |
| mind..... | 8 ff., 45, 70, 74, 79, 90, 99 |
| mind-body dichotomy..... | 104 |
| more than the body..... | 58, 113 f. |
| Myth of Interiority..... | 12, 112, 114 f., 118, 122, 126, 155 |
| private..... | 12, 70 f., 74, 86, 95, 112, 114, 116 ff., 147, 152 |
| representations..... | 26, 113 |
| secrecy..... | 119 |
| self-consciousness..... | 118 |
| soul..... | 9 f., 22, 87 ff., 96, 99, 110, 113, 146, 148 |
| stream of consciousness..... | 12, 112 f., 118, 122, 126 |
| theory of mind (animal)..... | 115 f. |
| "in the mind"..... | 112 |
| mirror neurons..... | 49 f., 68 |
| Modern Science..... | 6 ff. |
| Narrative..... | |
| belief-desire explanation..... | 91 |
| chronicle..... | 91 f., 146 |
| conscious status of narrative..... | 94 |
| Freud's "talking cure"..... | 94 |
| inadequacy of narrative..... | 95 |
| myths..... | 92 |
| narrative theory..... | 90 ff. |
| Sherlock Holmes..... | 91 |
| status of the narrative..... | 94 |
| Normative Binding..... | |
| binding illustrated..... | 97 |
| commitment..... | 80 f., 85, 96 ff., 103, 107 ff., 133, 144, 146, 153 |
| feeling of obligation..... | 98 |

| | |
|--|---|
| not a causal relationship..... | 20, 97 |
| not narrative sequence..... | 97 |
| perception is normative..... | 97 |
| promising..... | 97, 108, 140 |
| reference to past..... | 97 |
| Norms..... | |
| about what should happen..... | 139 |
| communal norms..... | 61, 72 f., 79 ff., 120, 142 |
| define actions..... | 59 |
| define essence of an entity..... | 139 |
| language as model..... | 138 |
| moral norms..... | 138 |
| norm defined..... | 139 |
| norm is not a cause..... | 140 |
| norms not magical..... | 142 |
| norms of arithmetic..... | 140 |
| norms-in-themselves..... | 141 |
| not a superficial add-on..... | 139 |
| not deterministic..... | 142 |
| not external impositions..... | 141 |
| not like rules..... | 141 |
| personal norms..... | 12, 96, 98 f., 107, 111 f., 118, 136, 141, 147 |
| pocket calculator..... | 140 |
| second level norms..... | 99 |
| should..... | 140 |
| social norms..... | 12, 79, 99, 106, 147 |
| tennis..... | 99 |
| Nowak, Martin A..... | 77 f., 82 |
| Orangutans..... | 46 |
| Organisms..... | |
| agent..... | 17 |
| amoeba..... | 15 ff., 28 f., 33, 37, 56, 58, 62 f., 81, 130, 133, 135 |
| correct concentration..... | 14 |
| food detector..... | 15, 17 ff., 23 f., 56, 81, 130, 133 |
| food-for-amoebea..... | 18 ff., 23, 25, 28 ff., 83, 113 |
| food-for-amoebea a functional thing..... | 25 |
| multicellular organisms..... | 55 |
| organism's world..... | 29 |
| organisms..... | 11, 13, 33, 35, 58 f., 134 |
| organisms as unified entities..... | 34 |
| protocell..... | 13 f. |
| receptor..... | 14 ff. |
| single-cell organisms..... | 11, 13 |
| toy amoeba..... | 18, 62 |
| ultra-violet detector..... | 31 |
| unified whole..... | 15, 17 |
| vital force..... | 14, 114 |
| Pain..... | 49 f., 68, 88 |
| Phenomenological approach..... | 29 |

| | |
|--|---|
| Plato..... | 87, 148 |
| Pohorille, Andrew..... | 13 |
| Premack, David..... | 115 |
| Psychoanalysis..... | 94, 126 |
| Punctualist..... | 88, 130 |
| Reality..... | |
| billiard ball reality..... | 128, 133 |
| contextual sense of reality..... | 134 |
| events..... | 130 |
| fact-in-itself..... | 130 |
| in quantum theory..... | 129 |
| in-itselfness..... | 130 |
| materialist..... | 128 f. |
| mechanistic conception of reality..... | 10 |
| nature of reality..... | 131 |
| pragmatic medium-sized objects..... | 130 |
| theory of relativity..... | 129 |
| thing-oriented..... | 130 |
| things-in-themselves..... | 128, 130 f., 133 f., 139, 143 f. |
| time in theory of relativity..... | 130 |
| transcendental reality..... | 143 |
| Umwelt-things..... | 33 |
| Reid, Thomas..... | 90, 101 |
| Reincarnation..... | 87 |
| Reputation..... | |
| loss of entitlement..... | 109 |
| loss of reputation..... | 109 |
| reputation..... | 12, 44, 96 f., 108 ff., 133, 146 |
| reputation defined..... | 108 |
| temporal structure..... | 110 |
| Romulus and Remus..... | 92 |
| Rousseau, Jean-Jacques..... | 86 |
| Sacredness..... | 53 |
| Self..... | |
| arrival of selves..... | 148 |
| as a communal relationship..... | 109 |
| as a normative status..... | 110 |
| as aberration..... | 150 |
| as status not object..... | 57 f., 104, 110 f., 142, 147 |
| authenticity..... | 5, 97 |
| bicameral mind..... | 148 |
| Buddhism..... | 150 |
| character..... | 12, 87, 96 ff., 119, 133, 146, 148, 153 |
| character defined..... | 98 |
| character is cumulative..... | 101 |
| character is learned..... | 100 |
| character is structural..... | 103 |
| character not intellectual..... | 102 |
| character of a self..... | 98 |

| | |
|--|---------------------------|
| character stability..... | 100 |
| cogs in the machine..... | 149 |
| degrees of selfhood..... | 154 |
| develops gradually..... | 99 |
| emergent entity..... | 34, 96, 110 |
| extinction of selfhood..... | 150 |
| habit defined..... | 56, 80 |
| habits..... | 102 f., 111, 146 |
| illusion..... | 128, 133 |
| meanings of “Self”..... | 85 |
| more than the body..... | 115 |
| nirvana..... | 150 |
| normative entity..... | 85 |
| not a punctual entity..... | 104 |
| origin of selfhood..... | 45 |
| person..... | 86 |
| protagonist in a story..... | 91 f. |
| responsibility for own character..... | 101 |
| same self over time..... | 88 |
| second nature..... | 102 |
| selfhood 6, 10, 12 f., 34 ff., 63, 69, 84 ff., 89 f., 92, 95 f., 99, 108, 112, 133, 137, 144, 146, 150 f., 155 | |
| selfhood is fragile..... | 149 |
| survival of selfhood..... | 149, 151 |
| temporal nature of the self..... | 102 |
| true self..... | 5, 12, 147 f. |
| virtual entity..... | 91 |
| Shakespeare's Polonius..... | 5 |
| Socrates..... | 5 |
| Status of my account..... | 143 |
| Stevenson, Robert Louis..... | 87 |
| System..... | |
| amoeba-system..... | 29 |
| bee-flower system..... | 33 |
| co-evolutionary systems..... | 33 |
| contingent..... | 138 |
| emergent system..... | 36, 62 f. |
| enduring system..... | 135 |
| evolves as a whole..... | 136 |
| functions defined by system..... | 34 |
| holism..... | 134 |
| language as a system..... | 137 |
| language paradigm..... | 140 |
| not downward-causation..... | 136 |
| self-system..... | 96, 110, 121, 127, 135 f. |
| stability..... | 135 |
| system not a cause..... | 132 |
| system, most crucial feature..... | 134 |
| systems define their components..... | 23, 34, 134, 137 |
| temporal systems..... | 133 |

| | |
|-------------------------------------|---|
| Taboo..... | 53, 57 |
| Thing defined..... | 24 |
| Thoreau, Henry David..... | 105 |
| Tick..... | 27 |
| Time..... | 145 |
| Tomasello, Michael..... | 43 ff., 47, 54 |
| Tønnessen, Morton..... | 29 |
| Totem..... | 54 f. |
| Trees communicate..... | 64 |
| Uexkull, Marina von..... | 27, 29 |
| Ultra-violet flowers..... | 31 |
| Understanding..... | |
| a situated endeavour..... | 145 |
| Archimedean point..... | 144 f., 150 |
| God's eye view..... | 128, 143 ff. |
| holism..... | 134 |
| natural light..... | 143, 145 |
| nature of understanding..... | 11 |
| no absolute truth..... | 128, 143 |
| subjective relativism..... | 61, 144 |
| systemic mode of understanding..... | 34 |
| understanding..... | 7 |
| understanding is contextual..... | 11, 34, 143 f. |
| Universal natural law..... | 7, 136 |
| Value not an add-on..... | 29 |
| Varela, Francisco..... | 32 |
| Woodruff, Guy..... | 115 |
| Wordsworth, William..... | 105 |
| World..... | |
| bees' Umwelt..... | 31 |
| construction of the world..... | 98 |
| linguistic Umwelt..... | 82 |
| things in my world..... | 98, 113 |
| things-for-an-organism..... | 28 |
| Umwelt..... | 36, 59 f., 62, 82, 98, 103, 116, 119, 135, 146, 154 |
| Umwelt defined..... | 29 |
| Umwelt thing..... | 30, 60 ff. |
| Umwelt things value-laden..... | 28 |
| world-for-amoeba..... | 83 |
| world-for-hominins..... | 82 |
| world-for-organism..... | 28 |