THE SELF THE SOUL AND THE WORLD: AFFECT REASON AND COMPLEXITY

Avijit Lahiri

Copyright

‘The Self The Soul and The World: Affect Reason and Complexity’

(a self-published book)

Kolkata (India): 21 November, 2022; 27 March, 2023.

avijit.lahiri.al@gmail.com

Contents

Chapter 1

Laying the foundations

1.1 Should you be interested in this book?

This book is primarily about the *human mind* and how it faces the world. It raises a number of issues, some of which are in the nature of *interpretations* of mine, relating to the *self* and the *soul*, and to the way individuals and larger social groups interact with the world in their journey in life. As for an individual, she finds herself surrounded by other individuals in a larger society, and a large chunk of her activities is devoted to the task of negotiating through complex social interactions. In addition, human beings are incessantly engaged in efforts to fathom the mechanisms of *how Nature works*. An important part of this engagement relates to investigations in science and technology. We will examine how our social interactions shape our self and our soul, and how the constitution of the self and the soul set the terms of our inquiries into Nature.

In this, we exert care so as *not to be inconsistent* with the vast literature on the topics covered in this book. This requirement greatly constrains us against taking liberties with how we frame our interpretations. In the end, the ideas and interpretations we present in this book will have to pass the test of hanging together with published literature, and of making up a coherent whole. Whether the book passes that test will be decided by its readers.

In this opening chapter, we state a number of basic positions that constitute a starting point for more detailed considerations to follow in subsequent chapters. The two notions of central relevance that play a pivotal role in this entire book are the ones of *affect* and *complexity*. It is affect that demarcates 'proto-concepts' from concepts, 'proto-beliefs' from beliefs, and the 'proto-self' from the self, where these are of foundational significance relating to our mental life — each of these terms is explained in later sections of this book. As for complexity, one feels awed as one contemplates the *mind* facing *Nature* at large — two impossibly complex systems confronting each other. But more of that later.

Speaking of affect, we will examine closely the strange and complex relation between affect and *reason* in the way the mind operates.

1.2 The brain and the mind

Every human being is blessed with a mind. And the mind is, to a large extent, a product of the brain. It is often said that the mind is a redundant concept — as and when we come to possess a complete knowledge of the workings of the brain we will not need the mind to explain our behavior. In this book, we will stay away from this line of thought, though we will not explicitly attempt to explain why — there are things that depend on one's point of view and are best left implied, since points of view cannot be argued over. What is relevant here is to note that the brain, together with the biological environment it is embedded in, is an enormously complex system, and the mind — the arena of a huge array of complex processes itself — is an *emergent phenomenon* within the context of this complex system.

1. Ontologically speaking, the mind can be looked upon as a *property* of the complex system referred to as the brain (see sec. 1.15).
2. Not all points of view are arbitrary constructs of the mind since they are often supported to some extent by observations and reasoning. However, there still remain *gaps* between the points of view and these attempts at rational reasoning. One can recognize the gaps if one is honest and discerning enough.
3. We will dwell upon the concept of 'emergent phenomena' at greater length later in the book (section 4.2.8).

The mind is a most remarkable thing — if one can call it a 'thing' — all its activities are, in some way or other, *goal-directed*, or purposeful, though there also arise *conflicts* among the innumerable processes of the mind, each having its own 'purpose'. To what goal or purpose does the mind work, and how did this purposefulness arise? We will briefly engage with these questions of central relevance later in this book.

> Goal-directed activities of the mind are often referred to as *top-down* ones. There exist, on the other hand, *bottom-up* or 'stimulus-driven' processes where the brain forms representations of stimuli or signals received from the world. The latter serve the purpose of the mind *perceiving* the world. However, perception always involves integration with existing representations, which results in the *interpretation* of the signals, a top-down activity itself. In other words, bottom-up processes are always associated with top-down ones and end up with goal-directed activities of the mind. Put differently and in general terms, top-down processes constitute the response of the mind to bottom-up ones.

As for the purpose the mind works for, an incomplete answer is: to maintain and improve our life-processes. This raises the question as to what the term 'life-processes' means and, then, what does 'improvement' consist of. One can get trapped in an infinite regress here and even then the question will not be answered completely. For the time being, we'll have to settle for an apparently circular answer: the mind is disposed to make us 'contented' (and, by implication, to make us avoid whatever causes aversion), where contentment is sustained only when complemented with *regulation*. In other words, the mind is ceaselessly engaged in regulatory activities while seeking contentment in the form of satisfaction of a huge set of *preferences*. This, on the face of it, brings us back to square one, but we'll have a sense of what it means by and by.

One evening, as I walk along a narrow lane that goes past a depressed-looking locality, I am in a rather downcast frame of mind. On my way, I hear a tune being sung softly by some lady, perhaps to make her baby sleep, and a memory rises in my mind of an almost identical tune that my mom used to croon to make me *sleep some fifty years back. Suddenly I feel so elated I almost burst with happiness.*

What put the purposefulness in our mind in the first place? The answer to this too is not easy to come by, though one often gets away by naming *evolution* as the causal factor. We too will accept this as an explanation of sorts, while desisting from building upon it for now.

In any case, the tendency of the mind to be goal-directed in seeking contentment will be referred to as its *adaptive* disposition — one that is remotely connected with the phenomenon of adaptation in biological evolution.

The mind, in the course of accumulating experience, generates a huge set of preferences by the operation of the *affect* system (sec. 1.5, 2.4), and sets its goal as the satisfaction of these preferences. Section 5.2.2 briefly mentions the factors responsible for the generation of the preferences. Incidentally, the term 'preferences' is used to include an equally important set of *aversions* too — ones that the mind strives to avoid. Incidentally, goal-directedness is, in a sense, a quite general feature in Nature (refer to sec. 5.2.1).

Speaking of the efforts of the mind to make us satisfied or happy, it is a plain truth that happiness often eludes us. This, in part, is due to the peculiar nature of 'happiness' that the mind is capable of generating (at times, we may even be unaware of the happiness, since affect is generated below the level of consciousness) and, at the same time, owes its origin to *conflicts* in the goals that various mental processes aim at. As a result of these conflicts, happiness remains a distant goal most of the time and, what is more, there arises the pervasive need of *regulatory* processes in the mind so as to manage the conflicts.

In other words, the mind is not a homogeneously directed collection of purposeful processes, all with mutually consistent goals. More precisely, every mental process has its *own* locally-defined goal, and myriads of conflicts arise between these locally defined goals. It is only a huge and parallelly-running regulatory activity that gives the mind an overall direction and coherence.

The mind is constituted to be a *reactive* system — it *responds* and reacts to inputs received from the world, all the while remaining *spontaneously* active in generating this response. In a protracted process of evolution and in the developmental process of an individual, the mind gets equipped with an exquisite ability to respond to inputs received from the world, in a manner that manifests itself as a quest for contentment and well-being, and for all-round regulation of conflicts. Oftentimes, the response of the mind takes the form of *anticipating* events that happen around it.

In this context, here is an important little note to remember: the mind is more a *process* than a structure. While the brain does seem to provide a structure in which the mind operates, things don't happen quite like that. The neuronal organization of the brain evolves ceaselessly throughout life — the more the mind responds, the more the structure gets modified. Indeed, the very response generates patterns in the brain that act as an effective modification of the structure. Thus, when one speaks of, say, the 'unconscious mind' or the unconscious 'layer' of the mind, it should properly be interpreted as unconscious *processes* in the mind. This understood, we will, at times, defer to common usage at some cost to precision. The term 'consciousness' stands for the capacity of the mind to be aware of objects and events (circularity again!) in *relational terms* and, in particular, to be aware of certain processes occurring in the mind, namely the very processes we refer to as conscious thought (however, there understandably is no such term as 'unconsciousness' — strictly speaking, the terms 'conscious' and 'unconscious' define each other implicitly).

Speaking of the distinction between the process view and the structure view, it is also important to refer to the relation between the so-called 'states' of the mind and mental processes. We often use expressions such as 'I am not in the right state of the mind for undertaking this job', where the state of the mind is meant to be something that lasts in time. The term 'process', on the other hand, means something that keeps on changing. In reality, a 'state' of the mind is not a permanent condition but is associated with a process that occurs on a *longer* time scale. This is typical of the activities of the brain and the mind: there is a *large spectrum of time scales* on which processes occur. In the ultimate analysis, this is symptomatic of both being *complex* systems (refer to sections 1.13, 4.2, 4.3.11).

1.3 Mental processes, thoughts, and concepts

Processes in the brain arise from the activities — of electrical and chemical nature — in the billions of neurons in it, and other associated physiological chains of events. Only some of those activities appear as operations of the mind. The mind, moreover has a *conscious* and a *non-conscious* component. We are mostly not aware of the huge flux of processes constantly going on in the mind — those we *are* aware of constitute our *thoughts*. Non-conscious and conscious processes are, in many respects, structurally analogous, one major similarity being the goal-directed, integrative, and regulatory nature of both. What is more, conscious processes are always associated with non-conscious ones, while not all non-conscious processes get reflected in awareness.

1. In this book, we will not distinguish between the terms 'non-conscious' and 'unconscious', though a distinction is at times drawn with reference to the submerged 'depth' in the mind a process is associated with. In this context, a third term — 'pre-conscious' is also invoked for some purposes. We will use the terms 'non-conscious' and 'unconscious' interchangeably. In addition, the term 'sub-conscious' that assumes relevance in certain areas of discourse, will not be used in this book as one having specific significance.
2. An enormous number of neural processes take place every moment that regulate the chains of physiological events in the body. These are not commonly recognized as activities of the mind — for instance, it is not the mind that regulates the functioning of one's kidneys, though it does get disturbed when the kidneys do not function well. On the other hand, a large number of voluntary and involuntary motor activities involving the motions of parts of our body are commonly deemed to qualify as functions related to the mind; many of these are included in what is referred to as *automaticity* — a capacity of the mind in virtue of which we come to possess skills that can be invoked without requiring conscious thought. At times, the conscious use of a skill many times over converts it into an automatic one. Later in the book (sections 2.2, 2.3.2), we will have a brief look at the types of processes that constitute unconscious and conscious mental activity.

The fundamental building blocks of thought are *concepts*. Concepts are generated in the conscious layer of the mind by means of *perceptions* of various kinds. In particular, perceptions can be of internal or external origin. Internal perceptions are the ones that are produced by processes of the mind itself — ones that can be unconscious or conscious, while external perceptions are generated from the world around us. The latter are registered in the mind by means of our sense organs and include, in particular, information of various complexity communicated to us by our fellow humans, and records of our scientific observations. Whatever their origin, perceptions primarily generate excitations in neuronal assemblies in the brain, after which the *information* carried by these excitations gets *processed* by being associated and integrated with other mental ingredients already lodged in the brain. We will gradually get acquainted with the multifarious types of mental ingredients that one may have. All these ingredients are fundamentally of the nature of excitation patterns of various types in multitudes of structured neuronal assemblies, of which we will have more to say later. It is this processing of the information content of excitation patterns in neuronal assemblies that constitute mental processes — ones that do the job of regulation (of conflicts) and of goal-directed functioning that the mind is so adept in.

> As we continue, we will be well advised to constantly keep in mind the distinction between the brain and the mind. The brain produces the mind, but in a manner that may continue to remain non-determinable.

The processing of information referred to above results in the accumulation of the stock of excitation patterns in the brain, some of which correspond to *concepts* — the fundamental ingredient of all thought that one can have. A concept is a representation of an object, an event, or a composite of objects and events. It is important to recognize that not all records of objects or experiences qualify as concepts. For the time being, we mention that a concept is a mental entity that an individual can be aware of, i.e., concepts are ingredients of *conscious* processes — ones, moreover, that evolve dynamically. On the other hand, there may be a stupendous variety of other dynamically evolving records of analogous kind that result from perceptions of various types and remain outside the scope of awareness — for the sake of concreteness, we will refer to these as *proto-concepts*.

In this book, we will have much more to say of processes occurring in the non-conscious and conscious layers of the mind, and of proto-concepts and concepts. In addition, we will have a lot to say of the distinction between the *self* and the *non-self* where, once again, the process view is more useful to adopt than the structure view. The two pairs of categories, namely, proto-concept and concept on the one hand, and self and non-self on the other, have a subtle correlation between them. That correlation will be made clearer by referring to *affects* and *emotions* in the human mind.

1.4 The non-conscious and the conscious

Activities of the non-conscious mind have been noted and commented upon for a long time, starting from antiquity. A new era was inaugurated as Freud told the world that the unconscious has its own 'logic' that may appear strange and distorted when compared to the 'rational' modes of reasoning followed by the conscious mind. Freud's theoretical and empirical contributions were all but banished from the field of academic psychology during the period from the thirties to the sixties of the last century, when the behavioral school gained in ascendancy and then was eclipsed itself by new trends emerging in psychology. The implicit modes of learning and of a number of other activities of the mind were again recognized, but now in a new context. Even before these new beginnings, the great chemist and philosopher Michael Polanyi had famously proclaimed: '...we can know more than we can tell' ([107]), thereby indicating the relevance of implicit memory and learning.

The emerging emphasis on modes of implicit functioning pointed toward an unconscious layer of the mind that was somewhat different from the Freudian unconscious made up of repressed drives and desires, and spoke of a 'cognitive unconscious'. Within decades, it was well recognized that the cognitive unconscious was responsible for a large number of adaptive functions that were earlier assumed to be exclusively in the domain of 'higher', conscious thought. Of special relevance was the complex task of social cognition that requires quite sophisticated functioning of the mind and is to a large extent undertaken by the unconscious ([8]). In particular, social cognition requires that *decisions* be made quickly in the mind. This resulted in attention being focused on implicit processes in decision making and similar other cognitive activities of the unconscious mind.

Earlier, it had been noted in psychological experiments that decisions are often made by individuals *not* in accordance with 'norms of rationality' and that, more generally, such norms were routinely ignored by individuals in various types of reasoning tasks. This led to a renewed hunt for identifying the nature of the reasoning processes made use of by the non-conscious mind. One expected such processes to differ from the 'logical' ones familiar in discourse and deductions carried out consciously, but the 'logic of the unconscious' was not easily grasped, and the 'irrationality' of individuals in real life situations remained unexplained.

A number of clues to the 'irrationality' of the unconscious were suggested, such as the one of 'bounded rationality' ([122]) which leaves us wanting when confronted with complex problems that were computationally intractable. Another important clue pointed to various types of *bias* ([72]) residing in the mind (the *belief bias*, and the *confirmation bias* are commonly cited instances) that caused our decisions to deviate from what norms of rationality would require. This ushered in a new era in the economic theory of consumer preference that is supposed to have a great role to play in modern-day market economics. Pretty soon, the role of *unconscious neural processes* was brought out as being of overriding relevance in determining our preferences and decisions in a wide spectrum of circumstances. And before long, the enormous importance of *affect* and *emotions* in our mental life took the center-stage in psychological literature and discourse.

> Of course, the actual sequence of developments in the various relevant disciplines based on which all the important insights and theories were arrived at was not anything like the neat progression suggested by the above sketchy outline. It never comes out like that. However, it sometimes helps to have a scheme of things built up notionally so as to cast reality in a simpler light.

We will talk of affect and emotions (at times one refers to *affects* — in the plural) at greater length in later sections in the book (see, in particular, sec. 1.5 below, and sec. 2.4 in chapter 2). And, we'll see how these are relevant in the emergence of the self. And, Oh, what about *consciousness*, then? How does it emerge and how is it related to the unconscious and to the emotions? How does it help us in accessing and appraising the phenomena pertaining to reality at large? We will have a look at these as we move ahead. Finally, the *soul* will be introduced as *the capacity of the conscious mind to reflect upon* *the self*, partly submerged in the unconscious, and to *reconstruct* it, at least in parts. The reconstitution is not necessarily a liberation in the commonly used sense of the term, but is certainly a great example of the purposefulness and regulatory activity of the mind. It is in the context of the soul that terms like ethics and morality — ordinarily thought to be rather nebulous in content — acquire significance. We will see how these can constitute higher forms of regulation in our mental life.

1.5 Affect and emotions: a primer

The term 'affect' carries a number of different, though overlapping, meanings in the vast literature devoted to it in psychology and neurobiology. At times, it is meant to stand for emotional feelings in an individual — an awareness of emotions having been activated, often with either strong or subdued bodily reactions such as tearful eyes, faster heartbeat, goosebumps, and tensed muscles.

However, such bodily reactions are often not under our conscious control, and appear automatically as a result of unconscious and innate mental reaction to inputs received from the external or the internal (mental) world of an individual.

> As an instance of an internally generated perception producing a pronounced bodily reaction, think of an imagined accident that a dear one may be supposed to have faced (say, *son or daughter out on a long drive in the hills and — who knows — perhaps caught in a landslide*) — just the supposition is often sufficient to suddenly cause a dry mouth.

In the present context we will use the term 'affect' to refer to basic and innately operative mental responses — ones 'hardwired' into the mind in virtue of evolutionary processes — that may or may not produce noticeable bodily reactions (which nevertheless appear, perhaps as subdued ones). These mental responses, in turn, activate *emotions*, the latter being complex mental markers that get associated with the experience arising out of a perceptual process. In a subsequent recall of the experience, the emotions are re-enacted (either strongly or weakly) in the mind. Affects, in other words, constitute the *core* of emotions.

The term ‘experience’ used above in connection with the idea of an emotion is to be understood in a broad sense. It includes the context in which the experience is gained, along with the state of mind at the time of it, and an ordering in the salience of the various factors involved. *A child has the happy experience of seeing his mother draped in a red dress singing on the terrace. After the lapse of many years, as an aged person, he sees a young lady draped in a similar red dress singing, and a happy nostalgia floods his mind.*

”Whereas affect is biology, emotion is biography.” Nathanson [98]

However, there exist differing accounts of what the ‘core’ is made of. Thus, several authors identify a number of *basic* emotions, some of which have positive valence and some negative, where the emotions add color to the valence and where each of these may occur in a mild form or a relatively strong one.

In this book, we narrow down the definition of the term ‘affect’ to a still more basic level, using it to refer to just the bare valence of a stimulus, i.e., a mental marker signifying whether the stimulus is pleasant or unpleasant where, in addition, the *strength* of an affect is often of relevance (commonly referred to as the ‘arousal level’). At times, the basic emotions referred to above are counted as affects since these are also hardwired in our mind, having been put there in the first place in the process of evolution. A number of other emotions may be assumed to result from the simultaneous operation of these basic emotions, having been ‘programmed’ in the form of neuronal circuits in the course of development of an individual. Finally, we often experience composite emotions (*love and hate*, *pride and embarrassment*) generated by a complex mix of simpler emotions.

Almost all our experiences, beginning with those from right after birth, generate affect (in the above sense of valence of emotions, along with strength of arousal) and emotions, and these get stamped upon memory as an experience gets stored for later recall. As we will see, the involvement of affect and emotions in the representations of objects, persons, and events in our mind converts such representations into constituents of a huge and composite neuronal excitation pattern of a very special category — one that generates the psychological engine referred to as the *self*.

At times, the terms ‘affect’ and ‘affective’ will be used in an inclusive sense in this book, with implied reference to emotions in addition to their valence (along with the strength of arousal), where this reference is to be picked up from the context. Thus, as we see below, the self is generated predominantly by association with the affect system where, to begin with, this involves mostly the affective core of valence alone, with more complex emotional links accumulating over time. The term 'affect system' itself will, at times, be used to include emotional associations pertinent to a context.

1.6 Affect: a complementary view

In the foregoing paragraphs, we have talked of two uses of the term 'affect': *one*, in the sense of the valuation-based core of emotions that marks an experience as desirable (pleasant, rewarding) or undesirable (aversive, punishing), and *the other* as the more complex emotional marker where an experience is uniquely stamped with a mix of emotions in which, at times, only one (or a few) of these may stand out as being dominant. In the remainder of this book, we will time and again refer to the huge and unique set of *preferences* that characterize the self of an individual, originating in the affective core of the experiences faced during her developmental history (this implies that the set of preferences itself is an evolving one) and also the vastly complex emotional history that gives the self a more complete identity.

However, there exists an alternative and complementary view of affect, largely initiated by Jaak Panksepp, that will also be pivotal to the view of self and soul that we develop in this book.

1. It was Panksepp who coined the term 'affective neuroscience' in the first place.
2. In the published literature on affect and emotions, the two views — one outlined in sec. 1.5 and the other mentioned in the present section, mostly originating in the work of Panksepp — are, at times, posed as competing ones. In this book, we adopt a broader view where ideas from each of the two approaches are retained without conflicts or inconsistencies showing up.

The two approaches to affect and emotions mentioned above will be referred to as the *reward*-based and the *seeking*-based ones, since the terms 'reward' and 'seeking' have acquired special connotations in the context of the two. This distinction helps one recognize the complementary ways of looking at the affect system that the two approaches bring to the fore. In a manner of speaking, the two views of affect differ in the basic affect systems and mechanisms that one needs to invoke in explaining the complex emotion-laden behavior of humans. As a consequence, the neural correlates of the basic affect components also differ indicating, perhaps, a future broader scheme in correlating the psychological and neuronal aspects of affect.

> The reference to *basic* mechanisms is not necessarily of fundamental significance. The brain and the mind are incredibly complex systems. The behavior of a complex system can be looked at from many points of view — indeed, an infinity of those. What is considered as 'basic' in describing the behavior depends on the way we look at the system, and on the context in which we seek to explain its behavior.

In the seeking-based view one speaks of seven basic emotional systems generated by the evolutionary process that humans share with most other mammals, namely, *seeking, rage, fear, lust, care, panic*, and *play*. Among these, the first four are supposedly of earlier evolutionary origin while the remaining three are shared by mammals, with their broader social orientation [37].

> In explaining the exquisitely complex array of behavior patterns in humans, various attempts have been made at identifying a set of 'basic' emotions, combined effects of which produce more complex emotional states. However, the mind is a complex system with vastly complex responses in different contexts, so that *various different* explanation schemes may be found necessary, depending on the way we look at emotional states and their effect on behavior. In other words, it may not do to identify one single set of emotions as *the* basic one. Various emotion studies have singled out distinct sets of emotional states as being of basic and fundamental relevance. Thus, three emotional states, namely, *happiness, sadness*, and *stress* have been identified in [61], the three being modulated respectively by the activity of the neurotransmitter systems dopamine, serotonin and norepinephrine (refer to sec. 2.4.1.3 in chapter 2 for a brief sketch of neurotransmitters and their role in generating emotions).

1.7 On proto-concepts and concepts

During a short period immediately following birth, the entire mental world of an infant remains saturated with affect, with only a few basic emotions gradually coloring that world, somewhat like the canvass of an artist. During this time, the mental processes of the infant are almost entirely unconscious, with consciousness creeping in relatively slowly as it passes through successive stages of development. The newborn senses its mother or the primary caregiver almost as part of itself, without any effort or ability relating to conscious appraisal. Thus, its mother and other people in close vicinity interacting with it get represented in its newly evolving mind not as objects identified distinctly and separately, but as *proto-concepts* generated with affective tags. A proto-concept represents something unique, such as the mother's face, or her dress, or even the colorful and sweet-ringing bell-toy given to it for keeping its attention engaged.

But the mother keeps on changing her (i.e., mother's) dress from one day to the next, and the dress ceases to be a unique object, though her dresses are still sensed as objects of affect, with only a dim perception of a *category* made up of unique instances of objects. Similarly, when a child is made to see a rolling object described to her as a 'ball' she, in her joy and wonder at observing a new and interesting object, registers it in her memory as a unique entity, with nothing but affect and emotion attached to it. As she comes to observe a few more similar instances of rolling objects and is instructed that each of those is a ball, the proto-concept is gradually transformed into a concept, where, strictly speaking, the latter is a *category*.

There are two basic aspects to a proto-concept — *association* between specific items in the world (where the process of category formation is at an initial and rudimentary stage), and an *affective* tag, though the two may not go together. Consider once again a child's first experience of a rolling ball. The round shape, the red color, and the peculiar motion on the floor — all these get associated in the mind of the child into one single perception in the form of a proto-concept. The association is, moreover, in the nature of a *simple* one where there is no precedence of one aspect of the perception over another excepting the one of affective salience in the mind of the child. What is more, the idea of the ball in the mind of the child refers to a unique object and does no involve any association to other objects in the world. Among the two features of simple association and affect, the former is the typical mode in which perceptions get registered in the unconscious mind while the latter has a special significance — that of being typically the way a perception gets integrated into the *self*-system.

Every perception, of however simple a nature, is actually a composite one. Thus, in the example above, the perception of a ball as a rolling object is associated with ones of color and size. To start with, all these are mixed together in a proto-concept saturated with affect, with vague and latent links to emerging proto-concepts of color and size, considered as separate entities, gradually appearing as experience accumulates. As further instances of rolling objects and objects of various colors and size are perceived, several aspects of perception begin to take more enriched shape in the mind all at once, each connected to others by means of associations — ones that begin to acquire greater complexity.

Stated schematically, one has the following picture: simple associations are the hall mark of unconscious perception; affective links over and above the associative ones are responsible for the cumulative generation of the self-system — one that is partly rooted in the unconscious; and finally, as the various aspects of perception become endowed with more enriched relations among them, *proto-concepts get transformed into concepts*, the basic ingredients of consciousness.

In the first experience of a red colored ball rolling on the ground, or in a few early repetitions of that experience, the aspects of shape, color, and the rolling motion are all fused into one single composite perception. Discounting the affective link to the emerging self, the association between these aspects is of a simple nature, devoid of additional structure. With gathering experience, it gradually transpires that the red color is shared by other objects, that the round shape is also similarly shared, that the peculiar rolling motion is due to the round shape (association of a *causal* nature), and so on. In other words, a proto-concept gets transformed into concepts as a result of two basic types of processes — one of each aspect of perception being seen to be shared by numerous items of diverse description (*category formation*) and the other of diverse *types* of association being registered (such as the one of causation, or one aspect implying another). The resulting set of concepts in the mind form a richly connected conglomerate with multiple links of association and correlation between them.

An affective tag remains associated with each concept in most cases, forming, in a manner of speaking, its core, while other defining associations get piled up layer after layer upon the core as links are established in the mind to a host of other emerging concepts.

Thus, a concept arises as the representation of some entity belonging to the world (which includes the mental world within our head) that the mind can survey as an *independent* entity, in terms of its relations to other concepts. *A proto-concept is associative (and often affective) in content, while a concept is part of a more complex network of relations.*

If it were not for the concepts, we would not have been able to make sense of the huge mass of signals from the world around us invading our senses every moment of our life. Concepts act as templates or organizing centers that make it possible for us to attach *meanings* to bundles of signals in relation to other bunches of signals already received by us. As we categorize the world in terms of concepts, we actually go beyond the raw information carried by the signals, which would otherwise have appeared to be chaotic to our senses (refer to [9], chapter 5). While concepts are formed out of specific bundles of signals, our senses receive large numbers of such bundles every moment. The representations formed in the brain out of these signals are then *integrated* to form perceptions of *phenomena* and *experiences*. This is how concepts allow us to *interpret* the world in phenomenal terms.

It is this network of relations that holds the key to the distinction between the non-conscious and the conscious, and also between the self and the non-self.

In summary, concepts lodged in the mind of an individual form a hugely complex network representing relations among them. The relations between concepts may be of various different *types*. The most elementary of all relations is the one of *simple association*, where two concepts associated with each other stand in a kind of equal partnership. As mentioned above, such relations of simple association are typical of the proto-concepts formed in unconscious mental processes. On the other hand, more complex relations (such as the one of *implication* or of *causation*) add layers of *meaning* to the proto-concepts, transforming those to the status of concepts. The most basic level of 'meaning' is the one imparted by affect and emotions that eventually remain the fundamental ingredients of the self-system of an individual.

Deep and insightful elaborations of the idea of concepts and their relevance in our mental world are to be found in [65], [9].

1.8 Neuronal assemblies and dynamical excitation patterns

How are the objects and events of the world represented in the mind? This is the fundamental unsolved problem in neuroscience and psychology. There hangs a looming doubt as to whether it will ever be solved in precise and explicit terms.

On a personal note, I consider it a divine blessing that the human mind is fundamentally inscrutable. This, however, is no more than a belief or a point of view of mine.

However, the broad contours of a correlation between perceptions of the world (which includes the mental world in it) and the neuronal configurations of the brain — the hardware of the mind — are not difficult to guess. This is based fundamentally on the states of neurons in the brain and their mutual connections across synapses. The neurons number up to billions, and the number of their possible pairwise connections runs into trillions. One can then only wonder at the enormity of the number of possible neuronal *circuits* that take part in the representations of perceptions, thoughts, and the ingredients making possible all the regulatory and purposeful activities of the mind.

The states of neurons in the myriads of neuronal circuits are based on the generation of electrical excitations in them and on the states of interneuronal synapses, the latter being controlled by chemical means through the action of neurotransmitters. While a single neuron can be either 'on' or 'off', neuronal circuits can be in an almost infinite number of *spatio-temporal configurations* based on the generation of dynamical excitation patterns in these circuits.

> Even the 'on' and 'off' states of a single neuron appear in the form of spikes running down the length of its axon, which can be looked upon as elementary instances of spatio-temporal patterns. As the spikes run successively through a number of neurons across the intervening synapses along the neuronal circuits, the variety and complexity of the resulting spatio-temporal patterns increase by leaps and bounds.

A single neuron or a connection between a pair of neurons is nothing more than a *microscopic* constituent of the brain, but neither of these represents any of the significant structures that generate the activities of the *mind*. All available evidence support the assumption that mental processes are all based on the states of *macroscopic assemblies* of neurons in the brain.

> " Advances in behavioral and cognitive neuroscience, and more recently systems and affective neuroscience have finally provided a foundation for conceptualizing diverse large-scale, psychologically relevant systems of interfunctioning neuronal nets that act together to generate coherent behavioral responses and primal psychological experiences." [150]

These assemblies are organized in a hierarchical manner and interact with one another through connecting paths. The existence of such interactions is known by weak electromagnetic signals from various different regions of the brain, that can be recorded and analyzed experimentally.

> All this is analogous to the fact that a single logic gate is but a microscopic constituent in the context of the running of a computing machine characterized by an *architecture* based on *macroscopic assemblies* of the gates. Various possible configurations of numerous assemblies of the gates make possible the running of software *programs* by the computing machine, where a program acts toward the realization of some 'goal' in the form of some desired end result.

> There is a strong school of thought that aims at explaining the human mind on the model of a computing machine. A major development in this area is the realization of machines based on *artificial intelligence* (AI), in which numerous activities of the mind have been reproduced. AI is one of the most potent forces of modern-day

civilization. However, there still remains the fact that even as the activities of a mind can be reproduced to a large extent by an AI machine, *each mind is unique in its developmental history* which means that there can be no *universal* AI machine that can reproduce all actual actions of human minds under all possible perceptual contexts. This, however, is an area where speculation lies beyond the scope of the present book.

The interconnections between various neuronal assemblies, including *re-entrant* connections, are responsible for the innumerable *modes* of mental activity, including conscious thought which is essentially *deliberative* in nature. All mental activity corresponds to the generation of dynamical excitation patterns in macroscopic neuronal assemblies within the brain, though the nature of this correspondence cannot be formulated in explicit and precise terms. In particular, this correspondence depends on the innumerable events making up the developmental history of an individual. Indeed, the neuronal connections and circuits themselves are experience-dependent — a fact referred to as the *plasticity* of the brain. What is more, the generation of an excitation pattern in the course of some activity of the mind depends on the *existing* patterns in numerous relevant circuits, which means that the patterns themselves are equivalent to structural features — in other words, the 'hardware' and the 'software' are inextricably locked in a mutually determining and evolving dynamic — the foundational fact in all operations of the mind.

All the ingredients, including concepts, preferences, beliefs, dispositions, and memories, that make possible the various operations of the mind are, in the ultimate analysis, based on dynamical excitation patterns in myriads of possible combinations of interacting neuronal assemblies. This is the basic position we adopt in this book, and constitutes the bedrock on which all our subsequent considerations will be implicitly assumed to rest — even though it will itself be of no direct relevance to these considerations.

1. When we speak of, say, belief in general, it is in the nature of a concept (an abstract one generated by introspection and discourse) that can be referred to in conscious deliberation. On the other hand, a concrete experience is marked with a concrete affective marker, a concrete belief and is stored as a concrete memory, all of which are unique and, depending on circumstances, may be associated with the unconscious layer of the mind.

2. Ingredients of the mind such as the ones mentioned above are the proximate causes of *behavior*.

In this book, we distinguish between the *mental* and the *psychological* — the latter refers to processes associated with the *self* (see sec. 1.9 below).

1.9 The non-conscious, the self, and the conscious: a composite picture of the mind

The *self* is a concept of central relevance in psychology, though an elusive and slippery one in that it is spoken of in many senses and from many points of view, without one single idea standing out and entailing others. In this book, we use the idea of the self as a central engine of our mental activities — the *psychological* entity that gives identity to our mental being. This, of course, is no solution to the problem of explicitly understanding and explaining what the self is, and has a circularity ingrained in it. One hopes that we all have a shared and implicit understanding of the idea of self, and discourse from various different points of view can only enrich that shared understanding even as the latter is not made explicit and precise.

The self is best understood in terms of its *developmental history*, i.e., as a process evolving all through life, starting from the time immediately after birth though, strictly speaking, the pre-natal period is also to be included in this history. As the baby is born, it appears as an entity in a world that is distinct from itself — though the baby itself has no sense of that distinction in the context of its survival and growth. In other words, the distinction is not manifest in its own behavior since objects in the external world attain independent relevance in its response to that world only with passing time. The world of the newborn is limited to the primary caregiver, along with a few other individuals attending to its needs and a few objects in its surroundings such as items of its daily use. All these — in addition to its physiological processes — satisfy its biological needs and are marked with affective identity.

In other words, only a very limited set of entities constitute the world of the newly born, all of which are identified with the bodily self of the baby by means of positive valence, while negative valence may also be generated. With time, the baby comes to possess a mind of its own, made up of affective preferences, dispositions, and proto-concepts. This also marks the emergence of the self which, at this early stage, is entirely submerged in the unconscious — indeed, consciousness is yet to emerge in the mind of the newborn since it is yet to take stock of the world in *relational* terms that it perceives beyond what the affect system and the rudimentary capacity of establishing associations tell it.

Given this initial emergence of the self, further development of the self-*system* in the mind takes place in accordance with processes and principles that can be summarized as follows:

The self-principle: *The self is generated iteratively by an affective-conceptual linking of current perception with the existing self-structure, i.e., the one generated in prior developmental history; the initial emergence of self occurs almost entirely in affective terms.*.

Here the affective link may be due to affect produced along with perception or else, may be generated from the perception by means of integration with diverse other mental processes and associated planning and *seeking* behavior (see [150] for background). Other than an affective link, a conceptual link with the extant self-structure may also serve to integrate a current perception into the self-system of an individual. In the latter case, the self continues to be an affect-based system because of its developmental history, including the one of its initial emergence.

At any moment of time the world presents numerous perceptual inputs to a growing baby, among which some turn out to be congenial to it, such as, mother singing, mother preparing bed, new toys with exciting colors and forms, new and fragrant smells, new facial expressions of appreciation and affective interplay, and so on. A whole new world opens up rapidly before its eyes, with novelty appearing at every turn, promising security, wonder, and pleasure. At the same time, unpleasant objects and events also acquire affective meaning on the negative side. All this gets integrated to the existing self of the baby in the form of proto-concepts, preferences, desires, aversions, and drives, and in accordance with the self-principle mentioned above.

> In the case of an insecure or a neglected baby, the self-system develops predominantly by means of *negative* affect. What is more, there develops stresses in its mind and body that leave deep and long-lasting effects on its self — ones that get manifested in untold ways as it grows up. At the same time, too much of security and pleasure denies the baby the joy and wonder of acting upon the world, finding its way in it, and discovering new experiences, i.e., in other words, the development of its self-system is thwarted and its mental life eventually gets tarnished with too many negative feelings generated by lack of fulfillment — it is to be remembered that fulfillment is the reward of effort.

To repeat, our perception of the world is based on representations of objects and events in the form of spatio-temporal patterns generated by dynamical excitations in neuronal assemblies. While the world is made up of objectively existing entities and the mind and its processes are also objective in the sense of being emergent phenomena (see sec. 4.2.8) based on neuronal excitations, perceptions constitute a very special type of process in which the mind receives and registers — in the form of representations — stimuli from the world. Objects and events appear as proto-concepts formed, on the one hand, by means of elementary associations between small numbers of representations and, on the other, by way of the representations generating affect, where the latter mechanism leads to the formation of the self immersed in the unconscious. For the sake of convenience, we refer to this early self almost wholly embedded in the unconscious as the *proto-self*.

With growing experience and the concomitant development of organized structures in the brain, objects and events present themselves *in relation to one another* by means of multifarious types of complex correlations — ones that get registered and represented in the developing mind over and above the simple associations characterizing the proto-concepts, along with their affective links. All these generate dynamical excitation patterns representing components of a newly developing system — the *non-self*. It is the non-self that, in contrast to the ever-growing self-system, represents various entities of the rest of the world that now constitute *objects of contemplation*. As this process of *objectification* of entities of the world proceeds, the mind begins to register the presence of an independent *reality* existing apart from its own self, without overt reference to its affective experience — this marks the emergence of *consciousness*.

With emerging consciousness, the proto-self expands into the self-system of an individual, based on neuronal excitation patterns and processes, some occurring within the unconscious layer of the mind and some others in the conscious. The self-principle mentioned above continues to operate, and the self continues to evolve in a complex manner, acting as the motive force that generates our behavior. With accumulating experience, the self acquires a structure, based on distinct roles played by distinct classes of dynamic excitation patterns — all based on interacting neuronal assemblies. This leads to the notion of the *ontological self* (alternatively referred to as the *objective self* or the self-*system* in this book). From the point of view of psychology, all the multifarious dynamical patterns contributing to the constitution of the self generate functions distinct from but related to one another. These *ingredients* (also referred to as *resources*) of the self include the network of concepts with proto-concepts residing at their core, a huge and evolving set of preferences generated from the affective valence of perceptions, self-linked memories, the store of self-linked beliefs and heuristics (see sec. 1.11below), and other psychological dispositions such as drives, desires, commitments, plans, and the like. All these diverse ingredients generate self-linked *thoughts* in the mind.

> As mentioned above, we distinguish between *mental* and *psychological* processes by way of noting that the latter constitute a sub-class of the former in being *self-linked* ones.

All along, the one distinguishing feature that the self possesses from the rest of the mind is the one of *affect constituting its core*. This is a result of the operation of the self-principle stated above, coupled with the evolution of the self from the proto-self, which is wholly affective. Growing upon this core, the self has a 'component' embedded within the conscious layer of the mind where the *relations* between the various parts or constituents of the world acquire significance and endow the self with a *dual* quality — with affect constituting one pole and relation-based perception the other. This generates a fundamental tension that makes the self the *psychological driving force* of an individual.

Speaking of the self, one faces the issue of the self as the *observer and the knower*, commonly referred to as the *subjective self*, or as the subjective 'I', to be distinguished from the ontological self or the objective 'me'. The latter is to be identified with the objective self mentioned above, while the former is a matter of some confusion. In this context, one has to recognize that *perception* is fundamentally in the domain of activity of the mind at large, and not just of the self, which constitutes a subsystem of the mind. It is the mind that perceives the world, and only a subset of all our perceptions relates to the self. Thus, it is the mind that fundamentally constitutes the perceiver and knower of the world, and the self acts as the perceiver in a more limited sense. Only those perceptions that get tied to the self by the operation of the self-principle — i.e., those that are affect-linked in some measure — can be said to be within the domain of the self. In other words, the self-linked perceptions constitute a subclass of all perceptions generated in the mind. It is this set of self-linked perceptions that is commonly assumed to be the purview of the subjective 'I'. Whie accepting this position, we distinguish between the *mind as perceiver* and the *self as perceiver* since the former commands a greater range compared to the latter (see, in this context, section 2.7.9).

Incidentally, it is crucial to recognize a class of perceptions as perceptions *of* the self (i.e., ones where the self is the perceived object), where the mind makes a map of the objective self. To be sure, such a map is often only partial and fragmentary, but then, *all* our perceptions are partial and fragmentary. It is the ongoing process of generation of these perceptions of the self that constitutes our *self-awareness*. While self-awareness is produced by the mind as a whole, the very process in which it is generated activates the affect system and self-awareness now gets owned by, precisely, the subjective self. It is in this sense that the ontological self gets objectified and is perceived as the objective self. Based on this perception, the conscious self now generates a *self-narrative* that it presents to itself (as an object of deliberation), and also to others — though the two narratives often differ considerably from each other.

We are on tricky ground now since we are referring to a set of distinct but closely related notions. There are, to repeat, perceptions by the mind at large and, among those, perceptions by the self. Next, the perceptions by the non-conscious mind differ significantly from those made consciously. Among the perceptions by the self, again, we have to distinguish between non-conscious and conscious ones. These distinctions are not idle ones since perceptions are processed into *interpretations* of the world, which determine the way we *respond* to it by means of behavior, where the latter, in a broad sense, includes further mental processes and thoughts. Further, along with the question of who acquires the perception (the mind at large or the self embedded in it) it is also relevant to ask as to *what* is the entity that is perceived where, once again, we have to distinguish between perceptions of aspects of the self on the one hand, and those of non-self entities on the other. In later sections of this book, when we speak of perceptions — all based on dynamical excitation patterns in neuronal assemblies — we will have to remind ourselves of these distinctions since the meaning and sense will often have to be picked up from the context.

Here is a partial summary:

The conscious and unconscious processes of the mind include ones that *map* the various parts of the world we live in — the latter as simple associations or, alternatively, in affective terms, and the former in more complex relational terms. The mapping appears in the form of representations made up of dynamical excitation patterns in the brain. At any point of time the *state of the mind* corresponds to a vastly complex aggregate of excitation patterns across neural assemblies — one that evolves dynamically.

The mind can even map its *own* activities in virtue of feed-back and re-entrant signals of diverse descriptions in the workings of the neuronal system. In particular, the mind can make use of this ability to access the workings of the self looked upon as an object of survey, thereby producing *awareness of the self*.

Finally — to conclude this section — in this book we adopt the position that the self, the non-self, our conceptual world, and our consciousness, all are, to a large extent, *socially generated* — essentially by means of our interactions with the world. In this, our social interactions mediate in our interactions with the rest of nature (see sec. 1.15).

One important insight into the working of the mind involves the recognition that the unconscious and the conscious layers of the mind *work in analogous manners*. Unconscious and conscious mental processes are mostly purposeful and integrative in nature, and are subjected to regulatory constraints. Indeed, in this book, we implicitly adopt the position that the unconscious and the conscious layers of the mind, are continuous with each other — both undertake the fundamental job of perceiving the world, interpreting it, and forming adequate response. This implies that consciousness is no miracle — it is the outcome of a well-orchestrated developmental process.

Unconscious processes occur mostly independently of one another, with only local interactions between neuronal assemblies, while conscious processes involve large-scale integrations across brain regions.

It is the analogous nature of unconscious and conscious processes of the mind that principally stood in the way of a deep understanding of the unconscious for a relatively long time since numerous unconscious processes (especially the cognitive and adaptive ones) were wrongly interpreted to be conscious ones.

All this, of course, does not mean that there is no need for a 'theory' of consciousness that is yet to emerge.

1.10 Concepts and the conceptual network

We have seen that concepts correspond to entities of the world that can be subjects of thought or of awareness, and are always related to or associated with other concepts. Indeed, a concept is defined by means of relations and associations. In contrast, a proto-concept is defined by means of simple associations alone, possibly along with affective links and, in most instances, refers to a unique object where relational aspects to other objects are not of relevance. Thus, when a baby perceives its mother as an object of affect, it does so without separate focus on her (the mother's) bodily parts (the face, hands, hair, etc.) though some part or other of the mother's body, such as her face, may be (and often is) an object of affect in its own right. In other words, the mother's face may *not* be registered *in relation to* the mother recognized as an object of appreciation. However, the relation does emerge in the mind of the baby by and by, and 'mother' and 'mother's face' appear as objects in their own right, related to each other.

The process continues, with more and more concepts emerging along with their mutual associations and correlations, and soon constitute a vastly complex assembly in the mind. Remarkably, as a concept (say, A) is defined in terms of its relation to a number of other concepts (say, B, C, ...), the definition of some of these other concepts (say, B) may, in turn need A as a referent. In other words, the set of concepts is *mutually* defined. This is analogous to a set of implicit equations in mathematics where, among a set of variables, each is defined in terms of the others. For instance, referring to the case of just two variables (x, y), one may have equations of the form $x = f(y), y = g(x)$, where f, g are given functions. On the face of it, such mutual dependence may appear to be a contradiction ('x determines y and y determines x') but still the pair of equations may have a perfectly well-defined solution for x, y (for instance, the equations $x = y^2$, $y = 2x$ imply $x = \frac{1}{4}, y = \frac{1}{2}$, discounting the trivial solution $x = y = 0$). One may alternatively think of words in a dictionary where there is no basic set of words in terms of which all the other words are defined, since all words are defined by means of their mutual relations.

The analogy between concepts generated in the mind and words in a dictionary is not fortuitous, since words are nothing but explicit linguistic representations of concepts. Of course, concepts constitute a much broader category than words — even a phrase or a sentence may represent a concept. A concept may have a highly complex content and may require a lot of description for a complete definition (for instance, a mathematical, scientific, or philosophical concept).

What is interesting and important to note is that truly speaking, a concept cannot be defined completely with perfect precision since it always consists of a fuzzy contour and an *implicitly* defined nucleus. The latter constitutes an implicitly perceived and affective core of the concept, which is thus a composite entity made up of an affective-implicit and a relational-explicit part — likewise, a word is never defined unambiguously in a dictionary since there always remains subtle nuances and vagueness of meaning that makes the definition open-ended. In reality, no concept nor a word can be defined with utmost precision in a finite description. However, such lack of precision notwithstanding, concepts are entities in the mind that can be defined by and large in *verbal* terms in a *language*, i.e., these are entities that can be freed to a large extent from personal and individual referents. In contrast, proto-concepts defined in terms of unconsciously acquired associations (partly or wholly affective) are pre-verbal and pre-linguistic entities, with implicit rather than explicit content.

Put differently, concepts are entities that can, to a large extent, be *communicated* by one to another. While a concept may originate from a proto-concept by means of gestures involving the presentation of multiple instances of an entity, such gestural communication is not wholly affective (thus, instruction by father is less affective than that from mother). After a certain stage the formation and proliferation of concepts occurs mostly through language-based communication and instructions, and through accumulating experience in the world.

As mentioned earlier, concepts are the basic building blocks in conscious thought. Put differently, consciousness is intimately related to concept formation. While concepts are characterized by explicitly defined relations, the same relations serve to *distinguish* between them. In particular, concepts make possible *the differentiation between the self and the non-self*. The former is constituted by means of affective links, while non-self is fundamentally made up of objects of the world and persons other than the self, making up the social and natural environment of an individual. Thus, in an ultimate sense, consciousness develops by means of social interactions (coupled with our interactions with the natural world) and is based on affect-free perception — in a relative sense — of objects of the world.

However, nothing pertaining to reality can be ascertained in absolutely precise terms. This is reflected in the fact that relational perception is never affect-free and always involves implicitly acquired associations, which is why every concept has a proto-concept as its core that constitutes an implicit content even though the concept itself is defined in explicit verbal terms. And likewise, the distinction between the non-conscious and the conscious is not absolute — the two merge continuously into each other.

The assembly of concepts lodged in the mind evolves ceaselessly in time, with each concept getting enriched by the emergence of new relations established with other concepts that generate successive layers of *meaning*, and by one set of relations nested within another, which is symptomatic of category formation. At any point of time, the entire assembly of concepts may be described as a *conceptual space* having an immensely complex structure. An alternative description is in terms of a conceptual *network* with the concepts as *nodes*, linked to other nodes in the network by means of associations and relations of a more or less complex nature. What is of great importance to note is that the network of concepts is a *multi-layered* (see section 4.2.1) one since there can be several *types* of links connecting one concept with another.

1.11 Beliefs as correlations between concepts

Bertrand Russell observed that *belief* constitutes "the central problem in the analysis of mind" (quoted in [123]). Likewise,

> "I argue that beliefs are the mind's software prescribing our behaviours, decision-making, and emotions." Smith [123]

Beliefs are of crucial relevance to the mind, enabling us to navigate through an uncertain world and making us bear the fundamental existential anxiety that life induces in us. They establish *correlations* between concepts and are made use of in the making of *decisions* and *inferences* — two major instances of goal-directed acts undertaken by the mind.

Our existence and survival in the world require us to make decisions and inferences ceaselessly, where the results of these processes are to have a good degree of match with the reality out there. This is the fundamental problem of how we *adapt* to the reality we are confronted with. A major part of responsibility in this process of adaptation is taken up by the non-conscious mind while the conscious mind also does not fail to play its part, notably in such fields as science and technology, economics, administration, and law. We will take up the issue of how we make decisions (refer back to sec 1.4 to begin with) and inferences in later sections of this book. It is in this context that *beliefs* turn out out to be of fundamental relevance.

Beliefs are meant to tell us things about the world. This includes both our external and internal worlds where the latter, i.e., the mental world is, to a large extent, accessed and mapped by the mind itself. In order to tell us how the world is and how to make decisions and inferences in it, the mind makes use of its repertoire of beliefs. The latter are often tied with emotions and affect, and it is the joint operation of the two that guides and shapes our life. Added to this, we are to recognize that beliefs are lodged at various depths in our mind, and can be held either unconsciously or consciously.

Many of our beliefs are commonly treated with a considerable degree of dismissal by saying that these are half-baked items of knowledge in that they have only a loose fit with reality and, at times, may even fly in the face of facts. However, there commonly resides a huge and pervasive web of beliefs in the mind of an individual, distributed across a wide spectrum in respect of their credibility. At one end of the spectrum are those beliefs that acquire the status of knowledge on the basis of confirmation, verification, and justification and, at the other, there are beliefs that are patently false. Analogous to the case of concepts, beliefs have a history of evolution in which they undergo revision in the light of evidence when some of these end up acquiring the status of knowledge. Even so, there remain beliefs that indefinitely resist revision regardless of evidence. The reason behind their intransigence is that they are tied with *emotions* (see [95] for background).

As mentioned above, apart from their degree of credibility and the degree of resistance to change, beliefs can also be classified on the basis of whether these operate at the unconscious or the conscious layer of the mind. We are little aware of beliefs lodged in the non-conscious depths of our mind and the power they wield over us.

Looked at from the point of view of establishing correlations between proto-concepts in the non-conscious mind, beliefs establish only rudimentary ones corresponding to simple associations. Thus, the sound of familiar footsteps conjures up in the mind of the newborn the smiling face of the mother even without the ability to associate footfall in a general way with the persons of other individuals in the household and also without any other related association that the conscious mind would later establish in a routine manner (for instance, *footfall* being *the result of impact of foot against floor*). Additionally, what is of great relevance is that the unconscious beliefs are mostly associated strongly with emotions and affect, and are impervious to the influence of 'external' evidence or logic. For instance:

Noting the stealthy mannerism of my next-door neighbor and the way he looks (seemingly) furtively at my children I have unconsciously formed a resentment against him and, in spite of having later found that he is a mild and honest person, I cannot help keeping a distance from him.

In this example, I am aware of my belief in a vague way and can even see the irrationality of it, but still the belief is lodged sufficiently deep inside me and is tied to anxiety and a negative affect generated unconsciously. In other words, one can be aware of the effect of a belief but the roots of the latter can still be inaccessible to him.

While the unconsciously held beliefs establish only rudimentary correlations among proto-concepts and are often tied to the affect system, still these play the role of *rules* governing our behavior, though ones of an elementary nature. Broadly speaking, a rule plays the role of an *if-then* type of correlation among concepts (*if footfall then mother* or, *if neighbor then avoid proximity*).

With all this in mind, we will refer to unconsciously held beliefs as *proto-beliefs* by way of analogy with proto-concepts. Briefly, these establish correlations of a rudimentary nature among proto-concepts in the unconscious layer of the mind, are often tied with emotions and affect, and resist revision in the face of contrary evidence, acting as 'if-then' type rules of an elementary nature.

Along with the proto-beliefs lodged in the unconscious mind, there are beliefs held consciously, though the two are of an analogous nature. The consciously held beliefs establish more complex correlations among concepts and, once again, act as rules governing our thought and behavior, where the rules may also be of considerable complexity. For instance:

I am of the opinion that the present government is a corrupt one and, like all corrupt governments, is undermining our democratic institutions; as such, there is no option other *than pulling it down from its seat of power.*

In other words, consciously held beliefs are analogous to proto-beliefs while having wider conceptual connections. As for their affective ties and their disposition to resist revision, consciously held beliefs have a mixed character. In spite of their wider and more complex conceptual connections, these are not free of affective ties, and it is affect that still makes them intransigent. The wide spectrum of beliefs referred to above is correlated with the degree of affective involvement and it is the same involvement that determines whether and how far these *spur us to action.*

We have noted that, at one end of the spectrum of beliefs are those that acquire the status of knowledge through a process of confirmation, verification, and justification and, at the other, there are beliefs that are patently false. Many of these unreliable beliefs, even though consciously held, are impervious to logic and rationality, precisely because these ar affectively tied ones, where affect is a strange thing that depends on details of the developmental history of an individual and has no direct relation to rules of rationality originating 'externally'. In other words, consciously held beliefs are generally of a dual nature — on the one hand they establish complex correlations among concepts belonging to a stupendously wide network and, on the other, they have emotional ties. This makes the beliefs act as rules that are *partly rational and partly irrational*, the relative importance of the two being a function of the degree of the emotional link. As a belief confronts evidence and logic, it may get more and more divested of emotions and approach the state of being counted as an item of knowledge, or else, it may be shielded from the process of revision by means of emotions exerting a brake.

What is of relevance in this context is that, the more a belief is resistant to revision in consequence of its emotional pulls, the more it is capable of provoking a person into *action*— belief, in contrast to knowledge, spurs one to action that may often answer to the description of being 'irrational'. Emotions have their own 'logic' that seldom defers to norms of rationality. Thus:

No amount of reasoning will convince me that the decision of the authorities to fire our *unfortunate colleague is justified — I am even prepared to tender my resignation in protest if she does eventually get the sack.*

The certainty of knowledge based on evidence and logic quenches the restlessness of mind while the very lack of certainty characterizing a belief based on emotion generates an anxiety to compensate for that lacuna by means of action.

Proto-beliefs and beliefs based on emotions and affect are linked to the self of an individual and are unique to her in that these depend on her developmental history. In contrast, a proto-belief or a belief that is not overtly dependent on affect is likely to be shared by other individuals belonging to a larger social group. For instance, imagine a red flag stuck on one side of a busy road. Almost everybody walking along it will be found to take a detour without thinking twice, under the spell of an unconscious belief that the flag must have been posted to issue a warning to people walking by. This is a shared proto-belief (based on a simple association between concepts — *red flag ... caution*), though not overtly affective in nature. Likewise, there is a belief held by a considerable number of people that palmistry is an effective art for foretelling what the future has in store for a person. Both these are shared beliefs — one held unconsciously and the other consciously. Later in the book (chapter 3) we will see how the distinction between self-linked and shared beliefs assumes relevance in the making of decisions and inductive inferences.

Talking of beliefs, we refer to the fact that these act as links between the nodes in the network of concepts lodged in the mind of an individual, where the network is a multi-layered one. In other words, there may be more than one type of beliefs setting up the correlations between concepts. For instance, consider:

The tiger is a dangerous animal; and again, *the tiger is a most graceful animal in its natural habitat.*

In the first statement the concept 'tiger' is related to 'animal' by means of the third concept 'dangerous', while in the second, the same two concepts 'tiger' and 'animal' are related in a different manner, by means of a different set of mediating concepts. These are two instances of beliefs, with none qualifying as certain knowledge. Such complex structure arising due to numerous different types of beliefs establishing correlations between concepts are of relevance in the context of *belief revision*. Thus, among the different layers of correlations establishing links between given concepts, one or more may get modified by force of new evidence coming to light, while some others may remain unchanged, perhaps because of emotional baggage. The concepts so linked may then appear in a new light, being endowed with new meaning and significance.

The case of *theory revision* in science is analogous and, indeed, constitutes a special case. From a broad point of view, theories in natural science are instances of beliefs that have passed through rigorous tests of justification, where these establish complex correlations in an elaborate and widely flung multi-layered network of concepts. While justified within given contexts, these theories are still defeasible and do not commonly carry the stamp of certainty, even within these limited contexts. As a result, theories get revised, sometimes mildly and sometimes quite radically. In the process, some of the multiple layers of correlation are modified or replaced with new layers while some others may survive, with the consequence that many of the concepts in the existing theory may appear in a completely new light. The emerging new theory appears in a dual relation to the earlier one — the two theories remain tied to one another by means of the surviving layers while *they are no longer on the same footing* because of the newly emerging layers. Later in this book, we will have occasion to dwell upon this aspect of theory revision in the context of *incommensurability* of successively emerging theories in science.

The conscious mind, based on the network of concepts, ceaselessly and pervasively *objectifies* the entities of the world by incorporating these into that infinitely flung network. As part of this capacity, it possesses the ability of thinking about, or deliberating on, beliefs as *objects* of contemplation (*"whatever you may say, I will never change my belief that people are fundamentally honest"*). This gives a hugely complex and nested structure to the conceptual space in which beliefs, as *rules* describing the workings of reality, become the object of focus in our quest for new ideas. The view that beliefs are in the nature of rules setting up correlations between concepts is of great significance, since inferences are made by making use of rules. As mentioned above, a rule, generally speaking, is a proposition of the *if-then* type (*if animal and if tiger then dangerous*).

In this context, we will distinguish between rules of uncertain validity and those of acknowledged credentials, the latter representing scientific theories. As an instance of the former we may cite a proto-belief that establishes an 'if-then' type of rule between specific or unique proto-concepts (*if footfall then mother*), without being used as a *general* rule. In contrast, as a consciously held belief is objectified, it turns into an instrument where it is used to established correlations between analogous sets of concepts in analogous contexts. Rules of the widest generality are those of *logic* since these appear as abstract ones capable of correlating concepts (that may include other beliefs as concepts of a greater degree of complexity) without regard to the meaning of these concepts. In other words, a great consequence of the emergence of consciousness, based on the infinitely extended network of concepts, is the capability of the mind to objectify beliefs acting as rules and to employ rules of various degrees of generality in creating correlations of an enormous degree of complexity.

Finally, in closing this section on beliefs, we mention that a class of beliefs of great relevance in the inferential and decision-making activity of the mind, is made up of *heuristics*. Heuristics are, in a manner of speaking, beliefs formed and discarded *on the fly*, much as a big computer program makes use of subroutines stored in a huge compilation of 'objects', or somewhat like *lemmas* in the proof of a mathematical theorem. Yet another analogy is that of a physicist or a chemist implicitly making use of a symmetry principle in arriving at some important theoretical result.

As we will see later, heuristics are generated in abundance in the course of decision-making and inferential acts of the mind. Every 'successful' act of such type generates a set of heuristics that the mind makes use of in subsequent occasions, while an 'unsuccessful' act also generates heuristics that serve as warnings, telling the mind to desist from proceeding in certain directions. In other words, heuristics have the dual significance of being generated in inferential processes and, at the same time, of being used as ingredients or resources (once again, mostly in the form if *if-then* rules) in such processes.

Looked at as beliefs, heuristics are often ones of proven credential, though there does exist a vast store of heuristics of doubtful validity too in the mind that the latter makes use of in its ceaseless inferential and decision-making activities. We will have occasion to dwell upon the role of heuristics in later sections (sec. 3.2) in the book at greater length where we will see how these can serve as *self-linked* mental ingredients.

1.12 Our conception of reality

The fundamental responsibility that the mind carries on its shoulder is to make meaning of the myriads of perceptual inputs received from reality at large, and to respond to those in accordance with the meanings that emerge and the preferences and beliefs that exist. In this, it forms representations of the perceptual inputs, integrates those with the existing set of information and beliefs, and forms decisions and inferences, in accordance to which it triggers actions, generating *behavior* — the way we act back on the world.

The representations that are formed in the first place are not *complete* and perfect facsimiles of reality. At any point of time, we perceive only a tiny cross-section of reality — a view of the heavens through a telescope, a scenic beauty, a family reunion, or even a sudden and revealing view of our own mind. Even so, we do not perceive all the details of the slice of reality so presented: the infinite details on the face of the moon, the detailed neural processes going through the minds of the family members participating in the reunion, or each leaf of the trees we view in the beautiful natural scene before us.

What is of importance to note, even the limited information received from the limited part of the universe we behold is not registered and responded to exactly *as it is*, because it is first integrated with the information that already exists in the mind, as the *processed* form of information *received in the past*. In other words, the response of the mind is always partial and in the nature of an *interpretation* of perceptual inputs received from reality out there. Can we, at least, decode the way this interpretation is done? Unfortunately, there is no way this question can be answered in the affirmative. Briefly, the interpretation is time-dependent and is determined by the entire *develop-*

mental history of an individual, on the basis of which beliefs, preferences, and attitudes are formed and it is the set of such ingredients of the mind that gets involved in the interpretation.

If our interpretation and response to reality is of such a convoluted nature, how can we ever hope to be accurate in that interpretation? This is one big question, to which no 'satisfactory' answer can be given (in this context, see sec. 3.2.2.7). Human existence has never been satisfactory in any unqualified sense — our perceptions always carry contrary messages for us. Happiness resulting from the interpretation of some part of nature is for ever blemished by terrible distress arising from the same process of interpretation applied to other parts that we are incessantly confronted with. One reason why this is so is that we interpret *social reality* in a way different from how we respond to the rest of nature.

Reality appears to us in a *phenomenal* form. That phenomenal reality differs from the 'real' reality out there. In the following, we will refer to the latter as the *noumenal* reality (sec. 4.3.2), borrowing from Kant, though Kant's use of the term carried a set of different connotations. Finally, our concepts that represent tiny bits of reality, correlated with one another by means of beliefs, make up a mental world different from either of the two — the noumenal reality and its phenomenal counterpart. It is the phenomenal reality that is the immediate source of all our mental ingredients including our concepts, preferences, and beliefs.

> The noumenal reality is self-determined, and infinite, while the phenomenal reality is contextual, and results from our perception of parts of the noumenal reality. In a sense, the noumenal reality is the source of our perceptions while the phenomenal reality is the result of those perceptions.

1.13 Complexity

In this book, we will have a lot to say about reality — both the noumenal reality and its phenomenal appearance — and how we access it and decipher it by means of our mental resources and ingredients. This is one foundational question in the entire world of science and philosophy. In engaging with this issue we have to be cognizant of *complexity*.

Complexity is all around us all the time but is something that has come under the lens of focused scientific inquiry in relatively recent times, in contrast to systematic and sustained investigations on *idealized* or *simplified* systems. Simplification is the name of the game, and quite understandably so. Humankind, with its remarkable cognitive abilities, has achieved very impressive success in understanding simplified systems which has essentially been the story of progress in the sciences so far.

In the case of systems whose complexity is of essential relevance and cannot be ignored on any count in some given context or other, progress has been mostly confined to empirical findings. On the other hand, science has excelled in looking at *models* of complex systems which are, in a manner of speaking, simple in nature and allow us to establish relatively simple regularities of their behavior that, moreover, compare favorably — in some cases, *very* favorably — with observations. The two most recalcitrant systems that have been met with are: the human mind and human behavior in the social context, and *Nature at large*, without reference to idealized and simplified parts of it.

The *theory* of complexity aims at characterizing complex systems and describing their behavior, including the way they evolve in time. This theory is currently in a very rudimentary stage of development, though important clues have been unearthed. It is of essential relevance in the context of real-life systems where complexity cannot be discounted in the interest of simplification. We will give a sketchy outline of some aspects of the theory later in this book (sec. 4.2) — a theory that is likely to help us in our quest to understand and comprehend our cognitive journey in this complex world of ours.

A complex system is usually one made of a large number of *subsystems* or components that interact with one another, where there may be numerous different *types* of interaction among these constituent subsystems. In this context we mention that complex systems are, at times, fruitfully represented in terms of *networks* in which the subsystems appear as *nodes* and the interactions as *links* between nodes. A network with more than one *types* of links is referred to as a *multi-layered* one. As mentioned earlier (sec. 1.11), multi-layered networks will be seen (sections refrevisions-sec, 4.3.10) to be especially relevant in the context of *theory revision* and the *incommensurability* of successive theories applied to models of reality.

> An alternative way of looking at complex systems is that these are capable of exhibiting a great variety of of *modes of behavior*, depending on the *context* these are viewed. All descriptions of a complex system are necessarily partial and contextual — while all such partial and contextual descriptions are mutually related, there is no final or ultimate description that can be collated from these.

1.14 The soul

We will now talk briefly of the *soul* and how it relates to our proposed study of the human mind in relation to the reality confronting it.

The soul is one of the most widely discussed subjects in philosophy, spiritual studies, and religion. It means many things to many people in many different contexts. In this book, we look at the soul as the *capacity* of conscious thought to access the self-system — one that evolves through unconscious and conscious processes — and to ceaselessly keep on *reconstructing* that self. This is aimed at resolving *conflicts* between various parts of the self, which has a complex and dynamic structure, and also those between the self and the enveloping reality, especially the social reality defined by human interactions of an exceedingly complex nature.

As mentioned earlier, the self is not a homogeneous being but a structured one. We have talked earlier of the adaptive functions of the mind, which imply the functioning of an *adaptive self*. The adaptive self is made up of many modules that carry on diverse functions in perceiving and interpreting reality, and planning our response to that reality. All these modules, based on interactions between neuronal assemblies distributed in the brain, form a complex system, and are frequently in conflict with each other, requiring overall regulation. The adaptive self, incidentally, is partly based on unconscious affective processes and partly on conscious, deliberative ones, the latter being also affect-linked in part. Among the two, the former occur in independent and localized neuronal assemblies while the latter involve large scale integration among distributed brain centers.

Further, the adaptive self resides beside what we will refer to as the *psychic* or the *repressed* self — a part of the self closer in spirit to the *Freudian* construct. While we will not have occasion in this book to refer to the Freudian self as such, the psychic self shares with it the feature of being the repository of repressed dispositions and tendencies. The latter appear as the mind tries to resolve the conflicts alluded to above, in which process a number of conflicts fail to get resolved, thereby putting the mind, along with the entire biological system, at risk.

> Conflicts are routinely resolved in the mind by means of interactions, possibly involving re-entrant connections between neuronal assemblies, where *emotions* play a seminal role.

The dispositions resulting from unresolved conflicts are repressed, i.e., forcefully restrained from getting expressed in our behavior and causing undesirable instabilities either within the mind or in the environment in which the mind resides. The repression may be unconsciously achieved, as in the case of turmoils inflicted on the mind in childhood days of an individual, or may result from a conscious effort (*I realize that I am under a fatal spell of attraction to a lady who is happily married and has children in the family; what is more, I have my own family to consider — I take great pains to repress and conceal my feelings, which results in an inner turmoil*).

In a sense, the psychic self acts as a force that tends to pull us back in the journey through life, even as the adaptive self tends to find solutions to myriads of problems that we are incessantly confronted with. On the face of it, the psychic self generates only torment, turmoil, and inner pain. At the end of the day, however, even the psychic self can be looked at as an adaptive and goal-directed device since it provides a way where conflicts are prevented from coming out in the open and generating a global instability in the existence of an individual or a society.

How are the repressed conflicts moderated and rendered harmless? This is a question that hounds and haunts mankind, with no satisfactory answer in sight — individuals and communities seek the answer in their own way. Ethics, morality, spiritualism, culture, religion, all have evolved around this single fundamental question of human existence. We will, in this book, dwell upon it only very briefly, citing the role of *consciousness* in this context. It is consciousness, and consciousness alone, that can *reflect* on the predicament of human existence and can engage in the effort — whether adequate or not — to come out of it. The rest is one big unknown.

The capacity of the conscious mind to access the repressed dispositions so as to regulate them and render them innocuous will be referred to in this book as the *soul*. More generally, the soul enables the conscious mind to critically look at the self and to reconstruct it. It is of paramount importance in our perilous journey through life. On it hangs the fate of human existence.

1.15 Modes of existence: individuals and social groups

The set of preferences and aversions of an individual is not all. Individuals are, at the same time, members of numerous *social groups* such as families, groups in workplaces, groups subscribing to religious doctrines, political groups, groups sharing common cultures, and so on. Our existence and our activities as individuals and as members of various social groups can be looked at as so many *modes* in which we confront reality.

All these groups have their own ways of generating sets of preferences and aversions that make a vastly complex system with pervasive conflicts among them. The history of humankind is the history of how various groups, cultures, and societies have interacted and behaved in the presence of these conflicts, in situations where the latter have proved themselves unyielding to efforts towards harmony and coherence.

There is a great deal of analogy between a group and an individual in the way the two respond to reality. In order to appreciate this, one has to realize that, in order to exist and to function, a group has to have a 'mind' of its own — it is not for nothing that we often talk of a 'group mentality' or of 'mob psychology'. As in the case of the individual, the 'mind' of the group is goal-directed, conflict-ridden, and engaged in regulatory activities.

Analogous to the mind of an individual, the collective 'mind' of a group does not have a material existence, but is very real nevertheless. The mind of the individual works on the basis of processes going on in neurons — the microscopic constituents of brain activity. Likewise, the 'neurons' on which the mind of a group is based are, precisely, the *individuals* in the group. Just as the mental activities of an individual are made possible by interactions between neuronal assemblies, likewise the 'mental' functioning of a group (its collective perceptions, interpretations, decisions, inferences, and resulting behavior) depends on interactions between effectively operating subgroups within the group or society in question where the term 'effective' has complex connotations, depending on such things as the defining characteristics of the group, its environment, the purpose that it is meant to serve, the *power relations* inside it, and so on (think of benevolent organizations, political groups, religious sects, cliques and conspiracies in a society; each of these subgroups can, generally speaking, be looked upon as a social group in itself).

> The mind can be looked upon as a *property* of the brain that has a phenomenal existence (see sec. 1.2, and chapter 6 for elaboration). This is analogous to saying that the repulsion between two electrons is a property of theirs. A property does not have a material existence but expresses how the interactions characterizing an entity get expressed in its behavior. Likewise, the 'mind' of a social group can be looked upon as an expression of the interactions between dominant individuals and subgroups within it, and is perceived in the form of its behavior under various contexts.

Further, analogous to case of the mind of an individual, that of a group is also riddled with latent and overt conflicts of various descriptions. Some of these are resolved by means of regulatory interactions within the group, while many others are repressed in the form of simmering discords. Some of the conflicts even come out in the open and require to be smothered by more serious intervention — at times, the conflicts explode within a group so as to lead to a renewed regrouping within a larger collective formation.

Finally, just as the mind of an individual evolves within a larger social environment, so does the 'mind' of a group. Moreover, analogous to the case of an individual, a group has 'non-conscious' and 'conscious' layers in its mental formation, and a 'self' and a 'soul'. These concepts appear to be tenuous, but are to be understood in a discerning approach (see sec. 5.2.8 for further considerations).

Many of the features of the workings of the mind of an individual and those of groups nested within one another are captured in the phenomenology and theory of *complex systems*.

1.16 The 'phase space' of the mind

Finally, we outline a fruitful analogy between the incessant dynamic processes within the mind and the dynamics of a complex system, where the latter is conveniently discussed with reference to a *phase space*.

> A good way to describe the complex evolution of a dynamical system is to refer to its 'phase space' where the latter can be described in mathematical terms. Pertinent to the present context are results relating to *nonlinear dynamical systems* where one can have a system with great many *modes* of behavior and scenarios involving transitions between these modes under small perturbations imposed on the system. Such transitions are symptomatic of the system being afflicted with *instabilities* of various types. A system may go through a series of successive instabilities, ending up with very complex behavior in its phase space.
>
> More generally, such transitions between distinct modes of behavior across instabilities is typical of *complex* systems. However, the instabilities are usually in the nature of *local* ones, where they are confined to low-dimensional subspaces of the phase space, involving a relatively small number of variables among those describing the state of the system, and where the system as a whole retains its integrity.

At any point of time, one can notionally speak of the 'state of the mind' in terms of sets of concepts, memories, thoughts, preferences, beliefs, drives, and various other dispositions of the mind, where these keep on evolving in a complex manner in what can be loosely described as a 'phase space' of an *enormously large* dimension. In other words, at any given time, there proceed in the mind a large number of distinct mental processes (many of these being unconscious ones) along a great many independent 'dimensions' and there may take place instabilities in some of these, causing a switch-over to new modes of behavior. The mind as a whole retains its integrity even as its evolution in time emerges as a complex one involving local instabilities, complemented with global stability.

As we will see (sec. 2.4), such instabilities are typically associated with the activation of *emotions* in the human mind. At the same time, the emotions themselves are responsible for the overall regulatory activities within the very same mind.

The above idea of describing mental processes in analogy with the dynamical evolution of a system in a phase space of a large number of dimensions, and also in analogy with the behavior of complex systems will be invoked later in this book. Complex systems have a wide repertoire of behavior analogous to nonlinear dynamical systems (see sec. 4.2.1), where both share the feature of transitions induced by instabilities intervening between regimes of stability. In addition, complex systems can be multi-layered, with more than one types of interaction between the nodes of a network, and are characterized by the feature of *co-evolution* (section 4.2.1). All these are of great relevance in the context of processes in the human mind.

Chapter 2

The unconscious, the conscious, and the self: revisiting the mind

2.1 The composite picture — an overview

We begin this chapter by reviewing the notions mooted in chapter 1 and briefly explaining those so as to arrive at a reasonably complete picture of the mind that will serve the purpose of this book.

The mind is based on ceaseless activity in the stupendously complex system of the brain, including its supporting physiological systems. While complex, the brain is nevertheless a structured whole, made up of macroscopic neuronal assemblies interacting with one another through neural pathways and circuits. The mind has an unconscious and a conscious layer of processes that make up a single coherent whole. It may be mentioned that the very definition of these two components of the mind involves a circularity in that the notion of the unconscious depends on the conscious, in being a set of processes that we are not aware of — in other words, the unconscious is in the nature of a construct, though not an arbitrary or a vacuous one.

The mind has a cognitive-adaptive function. It perceives its environment by means of representations formed of the multitudes of signals received from the world. In this, the mind also forms representations of the activities of itself by means of feedback signals of various descriptions. The representations, that are in the nature of maps of the world, are then processed into interpretations, where the latter are based on proto-concepts and concepts (see sec. 1.7), as well as on preferences, beliefs, and dispositions, where all these ingredients of the mind assign *meaning* to these representations. And finally, based on the interpretations, the mind generates behavior by way of response to its environment where, once again, the mind itself forms part of its own environment. In generating behavior, the mind integrates and regulates myriads of signals generated within it so as to avert the destructive effect of conflicts and to act as a coherent whole.

The unconscious mind and the proto-concepts formed of its perceptions precede the formation of concepts and of consciousness, where the precedence is to be understood in both phylogenic and ontogenic terms. With gathering experience, concepts are formed by layers of *relational meaning* accumulating on the pro-concepts that now constitute the core of the emerging concepts. Unconscious processes make use of only the core while paying little heed to the relational meanings inherent in concepts. Consciousness, on the other hand, makes use of the pervasive relational network of concepts, and has roots entrenched in the unconscious.

As mentioned in sec. 1.10, the relational meaning of a concept is analogous to the meaning of a word as expressed in a dictionary in terms of its relation to other words.

Proto-concepts are formed by simple associations among the perceptual inputs and by means of affect generated from these inputs. A large number of unconscious mental processes do not depend directly or significantly on affective associations (refer to sec. 2.2). On the other hand, a very important class of processes do depend significantly on affect and constitute the ones that make up the *self* — the unconscious component of it. The self constitutes a subsystem of the mind as a whole and, like the latter, has its unconscious and conscious 'components'. As in the case of the mind, the two 'components' of the self are to be be distinguished only notionally — they have overlapping and analogous activities, performed in two distinct styles.

While being initiated within the unconscious layer of the mind by affective means, the self-system evolves and expands with the emergence of conscious processes that make use of a pervasive network of concepts and greatly enhance the range of activity of the mind, including its overall regulatory role. The evolution of the self occurs layer by layer in an iterative manner with accumulating experience, as one layer envelopes a preceding one in accordance with the *self-principle* stated in sec. 1.9: as the mind receives perceptual inputs and concepts (with proto-concepts as their core) are formed and included into a vast network, those perceptions that are linked significantly with the extant self-structure get registered as new accretion to the self. This link to the existing self-system may be established directly by affective means or, alternatively, by means of significant conceptual linkages with only indirect affective connect. Given that the proto-self originates entirely in affective perceptions, it follows as a consequence that the subsequently evolving self-system remains always bound to affect.

Based on this picture of (a) the evolution of the conscious mind from the unconscious that is essentially linked with the evolution of proto-concepts into concepts, and (b) the evolution of the self from the proto-self, we will now be able to comprehend in clearer terms the overall adaptive functioning of the mind, based on processes of integration and regulation where these processes result in a ceaselessly active and coherent system.

The incessant adaptive activity of the mind is made possible by *ingredients* (or *resources*) that resemble the ingredients making up the self (refer back to sec. 1.9) with the distinction that, unlike the case of the self, the activity of the mind at large is not necessarily based on affect. For instance, beliefs of various descriptions constitute a major ingredient involved in the activities of the mind, but these need not necessarily be self-linked ones. Thus *shared* beliefs are made use of in diverse mental activities such as reasoning and inference-making. Looking at beliefs as rules establishing correlations between concepts, a particular set of beliefs of great relevance are to be found in scientific theories where the beliefs acquire the highest certificate of credibility, surpassed only by *mathematical* rules of inference. While scientific theories are defeasible, and get altered in the face of new evidence gathered in a wider field of observation, mathematical rules of inference are of a yet more universal nature.

Apart from beliefs (and heuristics), and other ingredients (such as those mentioned in sec. 1.9) made use of by the mind , the latter runs on a huge set of *housekeeping operations*, analogous to the operating system required to run a computer. These pertain primarily to the registering of myriads of perceptual inputs received from the world (where, as mentioned earlier, inputs originating in the mind itself are also to be included), storing of processed signals into memory systems, retrieving from memory, and operations that integrate signals of various types originating in diverse neuronal assemblies, the integration being effected by means of interactions among these assemblies. Along with these operations of integration, the mind has also the capacity to *regulate* possible conflicts between signals of diverse types by the activation of appropriate feedback loops that smother some signals and allow some others to be mutually integrated so as to generate behavior.

In this context, a set of housekeeping operations of major relevance is the capacity to generate *affect and emotions* (refer back to sec. 1.5 for a brief introduction) from signals received from various parts of the world, including parts of the mind itself. The emotions themselves, along with their affective core of positive and negative valence act as a set of essential ingredients in mental processes, especially in processes relating to the self. These will be discussed further in sec. 2.4 below.

We now turn to the elaboration of a number of major activities of the unconscious and conscious layers of the mind, indicating how these dovetail into a coherent whole. This will involve some repetitions of material presented earlier but, at the same time, will give us some degree of understanding of the way the human mind works.

Section 2.2 below will give an outline of a number of processes and functions relating to the unconscious mind, while sec. 2.3 will be devoted to the working of the conscious mind, though our survey in both sections will be of limited scope, compatible with the aims of the present book. Sec. 2.4 is devoted to the working of the affect system (the neuroscience and psychology of affect and emotions), which is so important in the context of this book.

2.2 The ways of the unconscious mind

The unconscious mind is a fixture we share with a lot of our predecessors on the evolutionary tree. With the progress of evolution, the unconscious mind has acquired more and more modules, giving it the ability to carry out diverse activities independently and in parallel, as a result of which an organism (an animal or a human being) acquires a greater degree of adaptive power where, once again, the explanation of the term 'adaptive power' will be left implied. This *model* of the unconscious mind is of relatively recent origin since, up to a point of time, it was only the Freudian unconscious that came under the purview of what was an attempt at a scientific and disciplined discourse.

Even as the processes in the unconscious mind run independently and in parallel — modeled by *parallel distributed processing* (PDP) in learning algorithms — there occur independently running processes aiming at integrating and regulating all these component processes. Indeed, processes of the mind are based on *interactions* between neuronal assemblies that have the dual aspect of (relative) independence and interactions stamped on them. The mind operates within a set of *constraints* defined by the biological requirements and social and material existence. It is this set of constraints that, in the ultimate analysis, determines the integrative and regulatory processes within the assembly of independently running and (at the same time) mutually interacting modules of the unconscious mind. What is interesting and important to note is that the modules for integrative and regulatory processes also run in parallel to the other modules so that, at the present level of knowledge, it can be asserted that the unconscious mind is *not centrally organized.* The goal-directed action of the mind — whether non-conscious or conscious — is not due to any centrally processing unit but is a result of the directed activity of every individual subsystem of the mind, based on a similarly unidirectional activity of the components of the brain.

The activities of the unconscious mind are of enormous diversity. Some of these activities are specific to the unconscious mind itself and are not associated with conscious activity, while some others run *in conjunction with* processes in the conscious mind.

Among the latter are some that are in the nature of housekeeping jobs needed for conscious mental processes to run smoothly. For instance, when one tries to retrieve from memory the details of some past episode, that process of memory recall activates an array of unconscious processes for it to be completed. On the other hand, there are some conscious processes that involve a kind of back-and-forth transfer of information processing between the conscious and the non-conscious. Indeed, unconsciously running activities are necessary for each and every consciously running process. As a useful analogy, one may think of a computer program where the main program, which is a sequential one, keeps on activating a number of subroutines in parallel and the subroutines in turn keep on transferring information to the main program. Something like this happens when, for instance, we try to construct a story by recalling a sequence of past events from memory.

The most fundamental function of the unconscious mind is, in general terms, the *processing of signals* of diverse types, received from the external and internal worlds, the latter being parts of the neuronal system itself. One major end-result of the processing is the generation of *perceptions* of various kinds. Also of great relevance is the production of proto-concepts and proto-beliefs (refer back to section 1.11), the latter in the form of rudimentary correlations between proto-concepts. The proto-beliefs and proto-concepts are used in the cognitive-adaptive activities of the mind. In addition, the unconscious mind produces a huge array of *preferences* and *aversions*, based on the activities of the affect system. At a fundamental level, the affect system operates unconsciously and spontaneously so as to classify our perceptions at a pre-verbal or pre-linguistic level and stamp these with markers that continue to remain when the perceptual experiences are recalled from memory. The preferences form the valence-based cores of *emotional* coloring of a more complex kind. From a broad point of view, the set of preferences can be looked upon as proto-beliefs (*if mother then good*, *if pain* (in stomach) *then bad*).

Thus, generally speaking, the unconscious processes of the mind are of two kinds — ones that do not depend directly on affect and others that are affect-driven.

The proto-concepts, proto-beliefs, and the preferences (along with aversions) form the basis of other dispositions lodged in the unconscious mind, namely, a set of drives and desires, and inhibitions based on avoidance, that generate behavior in the case of a baby and act as modulators of behavior in the case of an adult for whom the conscious mind wields enormous influence too. Most of these ingredients of the unconscious mind get incorporated into the self-system of an individual. The latter will be discussed at some length in sec. 2.7.

> The processes running in the unconscious mind have been grouped in three major classes ([8]): perceptual, evaluative, and motivational. The first of these correspond to the generation and initial processing of representations produced out of signals received by the mind; the second to affective processes, and the third to what Panksepp has termed the *seeking* behavior, which involves the formation of drives. Of course, all the three types of process have their counterparts in the conscious mind as well.

In generating the array of proto-concepts, proto-beliefs, preferences, and associated dispositions, the unconscious mind remains continuously engaged in integrating diverse sets of signals into stable perceptual packages and in regulating conflicts that may arise routinely. For instance *mother approaching with a smiling face* and a *cramp in the stomach* generate contrary feelings (one pleasant and one aversive) that eventually find a resolution in some specific behavior (*jumping onto mother's lap while whimpering with tearful eyes*). Once again, such processes of integration and regulation are relatively rudimentary in nature and localized in specific brain regions, assuming proportions of a greatly enhanced magnitude, over distributed locations in the brain, in the context of the conscious mind.

All these activities of the unconscious mind are based on the basic capacity, generated in a protracted evolutionary process, of ceaselessly establishing associations between perceptual inputs and generating behavior based on those associations. This capacity gets enhanced step by step during the developmental history of an individual and takes shape into more complex mental functioning. It is in virtue of this basic capacity that the unconscious mind can also detect statistical regularities in sequences of signals that it receives incessantly ([109]).

> The mind, made up of unconscious and conscious processes, is indeed an impossibly contrary device that, at the same time, is a remarkably coherent one.

While a number of unconscious mental activities appear to remain confined to the unconscious layer itself (to get expressed as behavior of diverse types), a large number of unconscious processes occur either as adjuncts of conscious ones or with some degree of involvement of the conscious mind. Indeed, it is quite doubtful whether there exists any mental activity that can be identified as entirely unconscious or entirely conscious (indeed, the very concepts referred to as 'unconscious' and 'conscious' are not defined in mutually exclusive terms and are, to some extent, notional).

The unconscious mind carries on the hugely important job of storing experiences into memory and recalling from memory. However, these capacities are confined to unconscious memory systems (in contrast to consciously invoked ones) that can be classified into a number of different types (see [127]). Among these we mention the ones referred to as *procedural* memory and memory associated with *priming* and *perceptual learning*. Procedural memory and, more generally, procedural functions of the mind refer to skills and habits that do not need the overt involvement of conscious processes (unconscious activities in riding a bike or performing a craftsman's trade) while priming denotes the phenomenon of an unconscious pre-disposition generated from perceptual inputs that do not attract conscious attention. Finally, perceptual learning is an implicit learning process based on perceptual inputs, such as detecting statistical regularities and co-variations (see [109], chapters 2,3) among signals received from the environment. In addition, the unconscious mind helps in other cognitive acts as well, such as guessing at mental states of others from facial expressions, and establishing communications by means of affect and emotions,

At a basic level, the capacity of the unconscious mind to form associations and to generate behavior in accordance to these is manifested in simple *conditioning*, i.e., a pairing between stimulus and response and learning processes based on conditioning. All other cognitive capacities of the unconscious can be looked upon as evolved and developed forms of this fundamental capacity (for instance, proto-beliefs are commonly formed by associations of a more complex kind, though rudimentary ones by the standards of the conscious mind).

An important entry in this list of various types of unconscious processes has to be the one relating to the formation of the proto-self, the subsequent evolution of which generates the self-system of an individual. The proto-self, it may be recalled, is entirely of affective origin (refer back to sec. 1.9).

Also of major relevance is the capacity of the unconscious mind to generate a rudimentary sense of *space and time*. From a basic point of view, the sense of space and time originates in special types of *relations between objects and events*. It appears likely that the sense of space and time is the result of *mental models* generated by the activity of neuronal *sequence generators* at specific locations in the brain that get coupled with our experiences relating to navigation in space and intervals between events. Based on this *empirical sense* of space and time there arise, in the conscious mind, the *concepts* of space and time which, at a later stage of development, are further developed so as to lead to *measurements* of spatial separation and temporal duration, and finally, to fundamental *theories* relating to space-time. All along, space and time remain as relational aspects of objects and events of the world, objectified in terms of spatial and temporal *co-ordinates* appearing as book-keeping indexes that may vary from one observer to another.

1. It appears that the mapping and modeling of space is guided by right lobe of the brain, while past-future integration and the mapping of time is done by the left lobe ([27], chapter 14).
2. The representation of space and time in the brain and the mind is incompletely understood. For background, see [19], [20], [137]

We will close this section by mentioning a number of cognitive activities of the unconscious mind where the latter acts in a supporting role in conscious cognition, at times even taking up the lead so as to guide and steer the conscious mind. The making of *judgments, decisions, and inferences* constitute major instances of such 'unconscious cognition' (see chapter 3 for further elaboration). Important instances of the making of unconscious inference is to be found in *intuitions* and *gut reactions* (refer to [57], [58] for background) where a repertoire of *heuristics* lodged in the mind plays an important role. Heuristics and their role in the activities of the mind were introduced in sec. 1.11; for further elaboration, see sec. 2.8.2. Finally, in this list of activities relating to unconscious cognition we mention the vastly important class of activities relating to *creativity* (chapter 4) — so fundamental in art and literature, music, and in the construction of *scientific theories*. Indeed, unconscious cognition covers a wide terrain and has a deep but elusive connect with human *rationality* (see [89] for background).

2.3 From proto-concepts to concepts: how consciousness works

With the resurgence of interest in the workings of the unconscious during the closing decades of the last century, there was a brief period during which the *conscious* mind receded (in relative terms) in the attention it received in the fields of philosophy, psychology and neuroscience. This, however, has been followed by a more recent renewal of interest in the 'problem of consciousness', more or less synchronizing with the turn of the century. The advent of many and varied approaches to the unconscious mind that supplemented (and, to some extent, supplanted) the earlier Freudian approach, enriched the renewed discourse on consciousness by way of comparison and contrast. Two early and influential books — one by Crick [29] and the other by Searle [116] — helped much in setting the terms in which meaningful research could progress, while subsequent work added to the diversity of questions and research techniques brought to bear on the issue of consciousness (see [75] for a brief review of relevant issues). There has been generated a feeling that the 'consciousness problem' is a 'difficult' one, so much so as to be almost 'insolvable'. It has even been suggested that one needs to bring in quantum theory to solve the 'problem of consciousness'. In particular, the term 'hard problem of consciousness' has been used in connection with the explanation of *qualia* — the uniquely subjective feeling generated by perception.

In this book we are not going to explain consciousness or to solve problems in the field of consciousness research. We will address the issue of consciousness in a common-sense approach, treating the workings of the conscious layer of the mind as being *analogous* to those of the unconscious one and adopting the position that what appear as unconscious and conscious are, in part, products of our perception, there being a continuum of categories of mental states ranging from the deep unconscious to the acutely conscious. Indeed, it is this close analogy between implicit and explicit mental functions (the distinction in style notwithstanding) that stood in the way of the discovery and understanding of unconscious capacities till late in the day, since many of these were earlier assumed to be conscious mental activities by default.

Our job will be to look at the conscious layer of the mind principally in the context of the self of an individual. The self has unconscious and conscious aspects to it — the two being, once again, intimately woven into each other. We will try to see how the 'problem' of consciousness relates to the formation of *concepts* that adds a very meaningful dimension to our interaction with the world we live in (refer back to sec. 1.9). In this, we will effectively place the unconscious and the conscious on the same footing, though with discernible differences distinguishing the two. The problem of qualia is to be understood in terms of the unique way that the self perceives the world.

For background to the widely discussed issue of consciousness, refer to [153].

2.3.1 The conscious mind: semantic vs affective

In contrast to the unconscious, the conscious mind operates on *semantic* lines, i.e., on the basis of *meaning* assigned to the representations — ones giving rise to *concepts* — formed of objects and events of the world. The meanings are based on relations between concepts woven into a widely flung network. We recall that the non-conscious mind operates by way of forming proto-concepts where representations are correlated to only a very limited extent and where the meaning is mostly affective and pre-verbal in nature. Associated with this semantic versus pre-semantic distinction between the conscious and the unconscious layers of the mind, there appears the *rational* versus *pre-rational* (commonly referred to as *irrational*) distinction between the styles of working of the two. However, it is a recurrent theme in this book that neither of these two dichotomous distinctions can be taken in an absolute sense, and that there is a continuity and analogy that relates the processes in the non-conscious and conscious minds.

As we have seen, there does not exist a sharp dividing line between a proto-concept and a concept (though the two are distinct sure enough). While a concept is characterized by its relations to other concepts in a widely flung network, a proto-concept is also based on associations — ones that can be interpreted as relations of a rudimentary type. On the other hand, a proto-concepts is commonly generated by affect while, at the same time, it is also common that a concept has affective links in its origin and subsequent evolution. Indeed, concepts evolve from proto-concepts.

As for the rationality of the conscious mind, we will see that it is also not absolutely demarcated from irrationality, which is why it is more appropriate to refer to the latter as pre-rationality or *proto-rationality*. Indeed, we will see that conscious inferential activity is mostly based on *inductive inference* (see section 3.2.1) that makes use of self-linked resources of the mind, which introduces affect-based criteria in inference.

As the conscious layer of the mind grows on the non-conscious layer, the proto-self grows into the self. While confining our considerations in this section to a number of general aspects of the working of the conscious mind, where self- and non-self processes get involved, we will focus more specifically on the operations of the self in sec. 2.7 below.

As in the case of the non-conscious mind, conscious processes (i.e., those that we can be aware of) may be self-based, i.e., may have significant affective links, or may be free of those links to a substantial degree. The latter are the non-self ones, being predominantly based on semantic relations among concepts in the conceptual network. This constitutes the basis of the distinction between the *subjective* and the *objective* functions of the conscious mind. From now on, we will use the terms 'subjective' and 'objective' interchangeably with self-based and non-self based processes of the conscious mind, remembering all the time that the same applies to the non-conscious mind as well, though within circumscribed limits. In the present section we will focus to a greater degree on the objective processes in the conscious mind though, as mentioned repeatedly at various places in this book, these have subjective roots too.

Digression: the problem of qualia. The 'problem of qualia' is elaborated upon in [116] where the subjective aspect of consciousness is considered to be in apparent contrast to what consciousness is supposed to be for: to act as an objective window to the world. However, as Searle states quite clearly, qualia is not something opposed to and apart from consciousness, it *is* consciousness itself. What appears to be a mystery is that consciousness is at once objectve and subjective — it enables an individual to look at the world existing independently of one's mind and, at the same time, it provides the latter with experience that is at core a subjective one. Indeed, this is the fundamental 'problem' of mind itself — one that appears as the continuity between proto-concepts and concepts, between affect and reason (refer to sec. 3.2, chapter 3), between the self of an individual and the group-'self' of social groups that she may belong to (see sec. 2.7 below), and between privately held beliefs and shared ones (sec. 2.8). Stated briefly, every experience has an affective core linked to the self of an individual, where the self depends on her detailed developmental history. When the conscious mind makes an observation and registers an experience, that experience may be a desultory and disinterested one where the non-self part of the the conscious mind acts as the observer, with only a minimal participation of the self — such an observation is only fleetingly subjective. However, as soon as the mind focuses on the situation in question, the self assumes the role of the observer, and the experience gets framed in 'qualia' (refer to sections 2.7.9 and 2.7.10).

2.3.2 Perception interpretation and adaptation

The fundamental job of the conscious mind, as of the mind at large, is to *perceive* the world and to establish *correlations* among perceived entities (objects, events, and processes), where the latter appear as concepts lodged in the mind, based on dynamical excitation patterns in neuronal assemblies in the brain. As for objective (i.e., non-self based) processes in the conscious mind, these correlations take the form of shared beliefs and objectively formulated rules, based on which the mind forms interpretations of the world out of the perceptions that arise primarily as representations of stimuli or signals that it receives incessantly both from without and from within. The mind seeks to form these interpretations with some degree of match with correlations between events in the world that appear independently of the mind itself, i.e., the interpretations are sought to be 'objective' in nature.

In other words, the non-self processes are 'objective' in a dual sense — on the one hand, these are free from links to the self of an individual and, on the other, these seek to reflect 'actual' correlations in the world independently of the mind. In this book we will be concerned with exposing the limits to both these aspects of objectivity while, at the same time, refusing to subscribe to the view that these are vacuous of content.

The interpretations are made use of in the making of decisions and inferences and in the formation of beliefs and *theories*; and all these then generate behavior, determining the way we *act back* on the world. The acid test of everything involved here — the interpretations, the inferences, the decisions, and the beliefs — is that our resulting action is to be *adaptive*, though it is a difficult job to precisely define what 'adaptation' really means. For the time being, let us agree to define 'adaptation' as *bringing us closer to what we want* or, *making us more contented* or even, *increasing our well-being*. This, indeed, is the deepest question ever — in what sense does our behavior take us closer to what we want? Because, what we want has multiple faces and pulls us to multiple directions. We will have occasion to dwell upon this question in this book without, however, arriving at any definitive answer. In any case, we will assume that our mind knows what it wants, in which case we can affirm the statement that the mind acts purposefully or in a goal-directed manner — on receiving stimuli from the world, it forms its response by way of generating our behavior that is, in its own terms, *adaptive*.

1. The goal-directedness of the mind is indeed oriented towards realizing its *preferences*. We recall that, in the ultimate analysis, preferences are formed by *affect system*, which is the pivot on which the self of an individual is assembled. On the other hand, adaptation has to have another aspect too — that of a close match with *reality*. Unless our response to the world fits with the way Nature works, there is no hope for our goals to be realized. In other words, the goal-directed working of the mind has a dualism inherent in it — on the one hand, preferences are generated by the self-related affect system while, on the other, the realization of these preferences depends on the workings of Nature regardless of the way the mind generates its preferences. Generally speaking, it is *cognition* (see 2.3.4 below), based on a large extent on *reason* generated in the conscious mind, that acts as the mediator in bringing together the two sides that are, on the face of it, independent of one another.
2. As we see in the course of our considerations, in generating behavior in a purposeful way, there operates a deep connect (in the form of cognitive processes; see sections 3.1, 3.2) between the self and the non-self — at the end of the day it is affect and the self that sets the mind on to the course of goal-directed activity. It is the job of the *soul* to free the mind from this tie-up with the self, at least partly, and to set new terms in the way the mind operates.

The fact that perception by the conscious mind is based on relational aspects of objects and events making a wide network (in contrast, correlations established in the unconscious mind are of a rudimentary and affective nature) prompts us to introduce a separate designation for it — we will use the term *observation* in order to refer to conscious perception. This is to be distinguished from the use of the same term in the sciences where it is imbued with more precise connotations.

It is important, in this context, to distinguish between observations made by the conscious mind at large and those by the *self* embedded within the conscious mind (refer to sec. 2.7 below). The latter are imbued with very special features owing to affective associations of a more or less direct nature. In particular, we will refer to the problem of *qualia* (see [116]) and its relation to the unique affect-based developmental history of an individual.

2.3.3 Reason and rationality: beliefs as rules

In the following we will, in referring to non-self based processes in the activity relating to the conscious mind, at times use the term ‘neutral’. This term will therefore refer to a quality distinct from ‘affective’ and ‘self-based’ ones.

> However, it is a central theme in this book that no activity of the mind is *completely* neutral. The mind is a complex system, and all descriptions of its processes are bound to be contextual and incomplete. Every description is bound to have loose ends and fuzzy borders. The same applies to our descriptions and theories of ‘reality’ which is a pervasively complex system. *Nothing* in real life is pure and perfect and the really interesting views emerge when this is recognized and the fuzzy borders are examined in details (see [10] for related considerations).

> It will be a principal concern of ours to see how the ‘neutral’ inexorably contains an admixture of affective and self-based ‘impurities’.

In all the activity of the non-self based conscious mind, it is *reason* that plays a stellar role. What is reason ultimately based on? The answer, in a sense, is simple: reason is ultimately based on the relation of *implication*. It is the relation of implication that constitutes the foundation of *logic*.

Implication establishes a correlation between concepts precisely analogous to ones established by beliefs since many of our beliefs share the ‘if-then’ structure of implication. From the point of view of logic, the relation of implication has the feature of being *truth-preserving*.

> *Truth*, of course, is a concept that can be a source of great confusion. In mathematical logic, truth can be defined in an unambiguous and precise manner, though that definition is not easy to state in simple terms. But real life is not mathematical logic. While one can try to apply the logical definition of truth to real life, the result is by no means completely satisfactory. We will dwell upon this unpleasant fact later in this book (see sec. 4.3.10).

In its bid to objectify entities of this world, the conscious mind objectifies our beliefs too and abstracts away from the concrete and limited context of their applicability. For instance the belief *my mother is affectionate* (*if* my mother *then* affectionate) can be abstracted to the general principle *mothers are affectionate* (*if* mother *then* affectionate). From a purely logical point of view, of course this abstraction is invalid because ‘my mother' is a concrete person while 'mother' is an abstraction applicable to all mothers However, in its bid for objectification the conscious mind makes precisely such abstractions. In other words, objectified beliefs act as rules of general validity and can be applied to various different instances, while beliefs of limited validity such as a proto-belief are applicable to one or a few concrete cases only. Among the beliefs lodged in the mind of an individual, the self-linked ones are tied to her developmental history and are strictly limited in their scope while beliefs shared by larger groups of individuals are, in comparison, free of affective ties and enjoy more general applicability, thereby approaching the status of objectified beliefs.

> Affective tags get attached to concrete entities of our experience. On the other hand, looking at an entity in purely relational terms frees it from that affective tag. This is how the transition from the non-conscious to the conscious parallels the transition from the affective and self-based to the neutral (see above for an explanation of the term 'neutral') and the objective.

In summary, the objectification of a belief into an abstract rule ensures that it can be applied not to one single situation but to all situations or, at least to a class of situations.

> As another instance, on working out several multiplication problems by means of repeated addition we find that a certain rule applies to all these individual cases. We then form a general *rule of multiplication* that is now objectified as a tool to be applied to all multiplication problems. This is an instance where the application of the tool guarantees success in all multiplication tasks. The objectification of beliefs proceeds in a similar manner. By observing that a belief is confirmed in numerous situations of interest we convert it into a rule where the rule is an object in itself and can be used to formulate more complex rules. This relates to the problem of *confirmation and justification*, of great relevance to *scientific theories*. However, no process of objectification is ever complete, and all our rules and inferences remain rooted in affect.

Since beliefs span a huge spectrum in respect of their credibility, one can likewise speak of a spectrum of 'rules' made use of by the conscious mind, because rules in real life are truth-preserving only to a limited extent and within strictly limited contexts. *Reason* is based on rules of more or less general validity while the fact that most other rules are *defeasible* to varying extents make it imperative that the term *rationality* be employed instead of reason. For instance, a decision may be rational even if it is later proved to be wrong (it might have been arrived at in the absence of sufficient data).

The distribution of rules — generated from beliefs — across a wide spectrum owes its origin to, precisely, affective and self-based associations that can never be fully distilled away. This, fundamentally, is the way the mind operates.

As a corollary, we refer to the way the mind thinks of *rules inherent in Nature*, i.e., those describing 'reality' at large. The fact that our beliefs have a good degree of fit with reality — a feature that is supposed to make adaptation possible — leads to the *larger belief* that the rules we arrive at on the basis of our beliefs are actually representative of rules *inherent* in Nature. In this book of ours, we will have occasion to examine this larger belief — one that appears in the name of *scientific realism* in philosophy. We will have occasion (see sections 4.3.8, 4.4) to examine this philosophical *meta-belief* of ours.

2.3.4 Social cognition and cognition of physical reality

Speaking of perception and cognition, we draw attention to the distinction between *social cognition* and cognition of the rest of nature — in particular, we will focus on our cognition of *physical reality* (also referred to as *material* reality).

1. The term cognition is used here to include processes of observation, the making of inferences and decisions, and the construction of new concepts, preferences, and beliefs, many of those based on earlier ones. Cognition, in this sense, is a fundamental requisite for adaptation.
2. We do not consider here aspects of cognition relating to plants and animals and to the environmental system as a whole. These aspects of great relevance are beyond the limited concerns of this book.

In contrast to cognition of the physical reality, social cognition, i.e., our cognition relating to people around us, involves a number of distinct and relevant aspects. It is our social reality that primarily shapes our self-system by way of distinguishing between the self and what is perceived as non-self.

While the cognition of the physical reality leads to our scientific quest and will be taken up in later sections of this book, the special aspects of social cognition will be touched upon in the present section.

Social cognition starts in life with the activation of a number of dispositions inherent in the mind, all stemming from the fact that our life is fundamentally based on communication — mostly, communications with fellow human beings. One of the basic dispositions found even in babies is that of developing and invoking a *theory of mind*. Children even a couple of years old have been found to possess the ability of ascribing *mental states* to people around them, where the mental states include beliefs, desires, intentions, emotions, and thoughts. This turns out to be the case even though the child is not necessarily aware of its own mental states and processes, i.e., possessing a theory of mind is in the nature of a largely non-conscious disposition. The child is even capable of identifying deviations from expected behavior that its theory of mind predicts, and correcting its theory of mind in recognition of the deviations — a capacity that develops with advancing age, even though our assessment of others does lead to gaping errors and blunders at every turn of life.

The theory of mind that an individual possesses is, at initial stages, based entirely on affect and emotions, since the only communication with human beings (mostly, the mother or the primary caregiver) that a newborn baby has is pre-verbal and affective in nature. It is on the same affectve basis that the self is assembled initially which implies that our social cognition is primarily self-based, i.e., is primarily a result of self-linked processes.

The same trend continues in later life even after the emergence of consciousness when the self expands to possess a conscious aspect in addition to the one embedded in the non-conscious. However, the emergence of consciousness makes us aware of the non-self in terms of purely relational concepts as well, which enables us to develop social concepts in terms freed from affect and the self. For instance, when one participates in a seminar on social and economic issues, one confines one's deliberations in predominantly conceptual terms, where affective and self-based considerations may remain entirely latent. However, the duality continues to remain and 'reason' continues to be contaminated with affect. As we will see (section 3.2), this intermixing of reason and affect occurs at a deeper level too since inference, which is thought to be an essential product of reason, is fundamentally aided by affect in what can be termed a covert manner. This is a common feature of social cognition and the cognition of physical reality where, among the two, the former is characterized by a greater extent of the overt intrusion of affect on reason. In other words social cognition is distinguished from the cognition of the physical reality in that it is based on affect and self to a much larger extent.

We may as well go so far as to say that our social cognition is predominantly affect-based. This is manifested in our social behavior in the following manner: in our social behavior, whenever we find something to have gone wrong or something to have worked against ourselves, we immediately find someone to *blame* for it and, for this, commonly choose someone for whom we have already developed an antipathy — at some stage we form a lasting *resentment* against some person or group of persons (say, the government), readily making him or them responsible for many of the problems that we face. This is rare in our cognition of physical reality — when a virus causes an epidemic we don't blame the virus, but immediately blame the government for mismanagement, which the government may or may not be culpable of.

One way of expressing this aspect of our social cognition is to say that our theory of mind and the resulting social cognition is heavily *underdetermined* by evidence. Underdetermination is a feature of a theory where the latter is not completely determined by available evidence and may be justified even in the face of some contrary evidence. To be sure, *every* theory that is compatible with available evidence is inderdetermined to some small or large extent. In the case of social cognition, a theory of ours (essentially, a set of beliefs) may go so far as to be compatible with pieces of evidence that may be diametrically opposite ones. For instance, imagine that a sworn enemy (say, A) of a person (say, B) has the option of either attending or not attending church on Sundays. If A takes the latter option (perhaps for reasons entirely his own) B will perhaps call him a godless fiend while, if he takes the former, B may say that A is diabolically concealing his true motives.

In confronting physical reality we incessantly try to find out regularities in it, form inferences whenever we face problems in comprehending reality and whenever our expectations, based on earlier experience and theory, go wrong — we set up specially designed experiments for generating new evidence, revise our beliefs as necessary, and create new *theories* based on rational explanation. However, in all this, the affect system is made inoperative to a considerable extent so that it may not interfere with reason. Assigning blames or allowing resentments to build up do not work. Likewise, undue emotional elation is avoided. This is what is commonly found to be the case in *scientific* investigations.

> However, there exists sufficient scope for affect and emotions to get involved in scientific investigations too, though not in a way directly relating to the investigations themselves. For instance:
>
> *As I feel that I am, at the end of a long spell of sustained and arduous efforts, finally closing in on an important result in my research project, I feel almost like collapsing with excitement*; or, *suspecting that a colleague of mine is close to finding the solution to an intricate problem on which I have been trying in vain to apply my mind, I am filled with an elemental resentment against him, that has the effect of blocking my mental faculties and impeding my own efforts.*
>
> In addition to these and similar other instances of affect and emotions playing a role in scientific investigations that is not directly related to the actual exercise of reason, one has also to refer to the way these are relevant *to the reasoning process itself.* This is the deep connect between reason and affect that we have mentioned at several places in earlier sections, and will dwell upon later in this book (see chapter 3).

Both social cognition and the cognition of physical reality operate to distinguish between the self and the non-self, where the term 'non-self' stands for people in the social environment in the case of social cognition and for objects in the case of cognition of physical reality — in the former, the distinction between the self and the non-self occurs predominantly in affective terms while in the latter, the same distinction appears in neutral, i.e., predominantly relational terms.

Having briefly referred to the distinction between social cognition and the cognition of physical reality, we mention, as a related phenomenon, that it is mainly social cognition that leads to the development of the self. In the case of a newborn, the self emerges in the unconscious process of affective communication with people in the immediate environment. With the subsequent emergence of consciousness, the self-system expands further by means of affect-laden observations while relation-based ('neutral') ones also play their role in the process if the relevant relational aspects happen to have significant association with the extant self, i.e., the self-system that has already been assembled in earlier developmental history.

In sec. 2.7 below we will have a look at the self of an individual and *also* at *group self*, the latter being made up of preferences, beliefs and related dispositions shared and possessed by groups of people to which an individual belongs, where the groups in question may have diverse descriptions (family, peers at workplace, groups sharing political or religious affiliation, and such like; refer back to sec. 1.15).

2.3.5 Consciousness: the neural basis

The content of this section is based on [151], [97], [136], [104], [105].

This book is not focused on the neural basis of mental activity — we have spoken of neuronal correlates from time to time in vague terms, only as a point of reference so as to indicate the physical basis of mental states and processes, the latter being somewhat like emergent qualities.

The term 'emergent quality' is used a lot in the theory of complex systems. However, there is a sense in which the use of this term can be described as more a recognition of complexity than a precise explanation of phenomena. Complex phenomena can seldom be explained in explicit terms — what is left implicit is, at times, described as 'emergent'.

In keeping with the rest of the book, we will indicate briefly and sketchily only a few takeaways — ones having some relevance for our present purpose — from the vast literature on the neuroscience of consciousness.

Consciousness is a unitary phenomenon involving perception of specific elements of reality, — both the reality ‘out there’ and the internal reality made up of mental and bodily processes — their transformation into conceptual categories, the binding together of the concepts into a unified and coherent *experience* immersed in a flow of time, attention to specific details of the experience, and reflection upon various aspects of the experience so as to express these in verbal terms. Along with reflection, some part or the whole of the experience gets stored in memory and, at the same time, the mind makes use of the experience for the purpose of deliberating, reasoning, evaluating, and planning. For our purpose, we will define the conscious process as the one running from perception to reflection, though some specific aspects may be absent in particular cases — for instance, attention (or even reflection) may be involved to only a minimal extent. What is more — and of great relevance too — is that experience gets stamped with affect and emotions that one may be aware of as one reflects.

The literature on consciousness research makes a distinction between consciousness and attention. From a conceptual point of view the two, though distinct, are related in an intimate manner and psychological and neurological studies lend support to this complex relation between the two. Perception begins with stimuli activating the relevant primary sensory centers (we leave out of consideration the perception initiated by mental processes for the sake of simplicity) — mostly towards the posterior region of the cortex — and getting represented in the form of neural excitation patterns. These are then associated together and processed into conceptual categories and bound together in the form of an experience occurring within the flow of time. For this, processes of comparison and integration are brought to play, where the neuronal representations are assumed to be presented to the ‘global neuronal workspace’. The theory based on the idea of the global neuronal workspace ”predicts widespread activation of a cortical workspace network as correlated with phenomenal conscious experience” [151]. The global neurons establish connections between segregated areas of the cortex and are widely distributed. The integration of specific sensory representations into a coherent experience requires that the latter be placed within the flow of time. This, presumably, is achieved by integrating with oscillatory (‘time-keeping’) excitation patterns distributed over diverse cortical regions. The job of integration into a conscious experience requires both sustained, ‘feed-forward’ excitations in the various regions (the output of one feeding into another as input) and ‘re-entrant’ excitations (giving rise to what is referred to as ‘recursive processing’) where neuronal assemblies are reciprocally connected so that they influence each other in some specific manner ([97], [136]). Important re-entrant circuits are not only cortico-cortical but also thalamo-cortical — ”Especially with regard to the thalamus, it has been proposed that this complex structure may play a pivotal role in supporting consciousness .... The numerous nuclei of the thalamus are extensively connected with the cortex, from which they receive feedback projections...” [97]. In particular, thalamic nuclei provide a coherent ‘baseline’ by propagating synchronous oscillations across the brain — temporal aspects of specific conscious experiences are constructed by comparison with this baseline.

This brings us to the role of large scale *oscillations* in distributed brain centers and their *synchronization* for the purpose of binding of separate sensory inputs into a coherent whole in a temporal frame. The possible relevance of synchronous oscillations (especially in the so-called gamma frequency range) in the generation of conscious experience was pointed out in an early and influential work of Crick and Koch (refer to [136]). In this context, it is interesting to note that both high frequency (the ‘beta-gamma’ oscillations, approx. 30-40 Hz) and low frequency (‘theta’ range, approx. 4-8 Hz) oscillations are necessary for the generation of an integrated and unified experience — the former for the integration of activity patterns within specific cortical areas and the latter for the unificaton of individual percepts by processes of integration involving distributed brain areas ([136], [104]).

Interestingly, it appears that there is a ‘job division’ between brain regions (roughly, the posterior, the middle, and the anterior ones) in respect of conscious processes. While the posterior regions are responsible for the production of individual items of conscious experience and the anterior ones for the integration of these for the purpose of reflection and interpretation, the middle region provides for the ‘resting’ or ‘default’ state of consciousness. The default state represents the wakeful condition, not specifically engaged in any conscious processing, that constitutes the essential backdrop against which conscious experiences make their appearance: ”According to the ‘default mode’ paradigm of brain function, a system of extensively interconnected cortical regions located mainly on the medial portion of the hemispheres, which is more active during rest than during perceptual and attentional engagement with the environment, is supposed to be crucial to the maintenance of consciousness” [97]. It is against the backdrop provided by the default mode that the perceptual and the affective and evaluative aspects of experience are integrated (refer to sec. 2.4.1 below; see also sec. 2.7.11).

In this context, we refer to the role of *neurotransmitters* (refer to sec. 2.4.1 below for further elaboration of these chemical information processing systems in the brain) in the generation and modulation of consciousness [105], especially in the broader sense where the term ‘consciousness’ is meant to include an awareness and evaluation of the part of reality one is conscious about. Of particular relevance to the maintenance of the level of consciousness (see below) are GABA and the glutamate transmitters (respectively of inhibitory and excitory types) along with acetylcholine. The monoamine neurotransmitters serotonin, nonadrenaline, adrenaline, and dopamine, along with acetylcholine are responsible for various modulatory functions affecting consciousness, especially ones relating to attention. Nitric oxide (NO), the neurotransmitter having the simplest chemical structure, is thought to be involved in recurrent and feedback neuronal processing. More generally, neurotransmitters are thought to be relevant in the widespread re-entrant neuronal activity that appears to be of central relevance in the generation of conscious experience since such activity requires specific matching between neurons at the synapses.

Finally, we mention the distinction between ‘level of consciousness’ and ‘content of consciousness’ that has been found to be relevant in consciousness research.

The level of consciousness or the degree to which an individual is aware of the world varies over a vast spectrum, ranging from wakeful and mentally active condition, through the default state, various phases of sleep, transient and light unconsciousness, unconsciousness induced by anesthetic drugs, coma, seizure, and various other related conditions. Such variation of level of consciousness corresponds to cessation of global integration between various component systems contributing to consciousness, some of which have been mentioned above. In contrast to non-conscious processes, conscious ones are dependent on widely distributed feed-forward and re-entrant neural circuitry, which need to be switched off to different degrees so as to correspond to the levels of consciousness across the spectrum mentioned above. While this corresponds to what may be looked upon as the presence or absence of global neuronal communications, the content of consciousness at any given point of time refers to the specific items of relevance that the mind is conscious of at that time. For instance, if one has the impression of looking at a red flower and a yellow flower, and if the red flower is removed, then the activity level of a number of specific neuronal assemblies will get reduced that can then be interpreted as the 'neural correlates of consciousness' (NCC) corresponding to the perception of the red flower. The NCC informs us about the way specific neuronal assemblies get activated as we perceive specific entities (or combinations of entities) in our environment.

In the above paragraphs we have, for the sake of simplicity, referred mostly to the conscious perception of signals received from the external world. Consciousness is also routinely generated by means of signals originating in mental processes (*suddenly becoming conscious of an urgent business that I forgot to attend to*), and is sustained by thought processes in the mind (*remembering a missing key, I keep on searching at various possible places, guided by a continuing inference-oriented thought process*). The general features of neural integration over distributed brain regions remain operative in such mentally generated and sustained consciousness.

As mentioned earlier (sec. 1.9) the non-conscious and the conscious processes of the mind are continuous with one another, and operate on analogous principles. The distinction between these two sets of processes lies mainly in the strength of activation of neuronal assemblies and and the degree of integration among the activities of these assemblies. While unconscious processes are characterized by localized and small-scale integration, conscious ones involve integration over distributed regions in the brain (see, for instance, [41]).

2.4 Affect and emotions revisited

Affect and emotions were briefly introduced in sec. 1.5 and sec. 1.6.

In the *reward*-view of affect, the latter can be looked at as an innate response to inputs received from the world, marking the response with either positive or negative valence, associated with a strength (the 'level of arousal'). Emotions, in contrast, are more complex markers of situations and experiences, some of which may be assumed to be more basic than others. These markers assume relatively more complex texture with accumulating experience as, in the course of ontogenic brain development, there arise mixed emotions of various descriptions generated by the activation of diverse neuronal circuits. Affect, along with a few basic emotions, are of evolutionary origin, while emotions of greater complexity get defined during the developmental process of an individual.

Nathanson, in [98], identifies *shame* and *pride* as a pair of affects of fundamental relevance in that the self-system is built around the axis made up of these two. We will dwell upon this at greater length in a later section (sec. 2.7.4).

Affect and emotions serve the fundamental function of *classifying* the world around us. The most basic classification is in terms of 'positive' and 'negative', – desirable and undesirable, based on the affective valence resulting from an object or an event.

This basic classification, however, is not an entirely objective one since it is based as much on the objects and events of experience as on a large number of factors depending on the mental ingredients of an individual accumulated in the past and on social factors influencing her mind (refer to section 5.2.2).

Higher and more complex levels of classification are provided by the associated set of emotions. The sight of mother approaching her child is, fundamentally, a pleasing one to her (the child) and, described more fully, produces joy, excitement and, at the same time, anxiety that the mother's visit may not last long (perhaps gathered from mother's attire and demeanor). Here the last named emotion is associated with a mild negative affect while the first two have strong positive valence. The basic affects ('positive'-'negative') are characterized by the feature that *these can be mapped onto a single value-dimension (often referred to as a 'common currency') so that they can produce a net resultant valuation.* Thus: *caught under a spell of conflicting emotions, the child eventually leaps to her mother in joy.* Here is another instance: *a deer standing by a stream in a forest is happy to find drinking water close by and, at the same time, scents danger in the possible presence of a predator — it flees the spot, based on the resultant valuation.* Classifications — in particular, the dichotomous one in terms of perceptions either congenial or inimical — have great adaptive value and are of fundamental significance in life. Affective communication between individuals consists of pre-linguistic transfer of information, based on affective valuation, where emotions routinely get involved.

Classifications more refined and complex than the basic 'positive'-'negative' one are made possible by emotions acting as complex markers of experience. As mentioned, all this classification is fundamentally *implicit* or *pre-linguistic* in nature, typical of the unconscious mind. The conscious mind, on the other hand, is capable of much more explicit classification that is often precise and sophisticated, but the implicit classification is more effective in a range of circumstances where quick results are important, and analyses based on fine-tuned conceptual correlations are not of essential relevance.

We now mention a feature of emotions and affect (in the above narrowed down sense) that play a seminal role in a wide range of psychological processes. We note, to begin with, that the state of mind of an individual is in a constant process of evolution along multifarious dimensions. In this ongoing complex evolution, emotions (along with their affective core) play the role of either *amplifying* or *moderating* factors or, put differently, as factors that lead to *instability* or *stability* of mental processes.

While reading a book on a quiet evening in my living room, I am reminded by a certain passage in it of a happy occurrence that happened long back in my teenage days, and my reading gets disrupted when I set out to relive those bygone days in a nostalgic mood.

This is an instance of a smooth process being de-stabilized by an emotion with a positive valence and a new course of thought being pursued. The fleeting activation of an unpleasant thought, on the other hand, would have forced my mind back to the book. Our psychological life, in other words is, to a large extent dominated by the effects of emotions that channel our thoughts and activities along very many courses that would otherwise not be accessible under ordinary circumstances. We are relieved from anxiety, boredom, and feelings of depression by the operation of affect and emotions. It is emotions that lend color to our life while, at the same time, emotions may be instrumental in causing despondency and despair — emotions, in other words toss us back and forth between darkness and bliss. Our mental life never runs in a smooth manner. It proceeds through swings in various directions under the influence of affect and emotions. This is of great value in keeping us on course in the face of a constantly fluctuating environment though, at times, it leads to mental disruption too.

As mentioned earlier, the instability or the stability alluded to above is of a *local* nature. At any given moment, our mind evolves along very many dimensions, many of those beneath our awareness by means of associations — often under affective guidance — and also by means of more complex correlations between ideas. The activation of an emotion or an affect with a positive or negative valence relates to only one or a few of those, and only those few strands of thought (often correlated with one another) get affected by emotion, making the mind stray into distinct channels. This feature of local stability and instability of our thought processes can be further elaborated in terms of a mathematical analogy, as we will see later (sections 2.4.2, 2.7.3, 5.2.3).

On the whole, affect and emotions are of great adaptive value in that these make possible quick and effective response to a complex and uncertain environment that, moreover, keeps on evolving along multifarious dimensions. Of major importance, in particular, is the *social* environment that requires exquisite emotional regulation to be exercised by the mind in order that an individual can navigate through it without suffering setbacks.

In the complementary *seeking*-view, affect is once again looked upon as the innate response system of an organism developed in the evolutionary process, with at least seven basic affect-based mechanisms generating behavior in a complex manner. The latter operates at three levels that can be distinguished from a neurological point of view — the primary level of innate emotional response, the secondary level of learning processes based on affect, and the tertiary top-down level of consciously controlled emotional-cognitive behavior. The seeking view, in other words, attempts at explaining behavior in terms of a set of drives and dispositions that the reward view associates with complex interactions of the basic emotions and their affective core. The primary-level emotions, for instance, involve a valenced core that lead to secondary-level learning through a reward-punishment type feedback in the associated neuronal system [37].

For backgound on the seeking-based view of affect, refer to [150], [37], [103]; in this context see also [17].

A huge literature exists on the *neural* underpinnings of affects and emotions. In particular, the *reward-punishment* network or, in brief, the *affect network* continues to be of great interest to neuropsychologists in that it possesses immense significance in almost all unconscious and conscious thought that one can have, especially the self-related ones. In particular, it provides the basis for the affect theory of *the making of decisions* (section 3.1). Later, we will outline a case that the role of affect in the making of decisions extends to the making of *inferences* as well (section 3.2). Notably, the affect network is instrumental in the *establishment of novel correlations in the conceptual space* and, in the ultimate analysis, in *creativity*, where the setting up of extensive correlations results in a *restructuring* of the conceptual space (section 4.5.4). In all these processes there is an apparent lack of determinism involved, analogous to a similar feature in the exercise of *free will* (refer to section 5.1). We will argue that this can, to a certai extent, be explained in terms of the self-linked psychological ingredients of an individual, i.e., those mentioned above in connection with the adaptive and the repressed self.

It is of interest to note that concepts and the conceptual space result from our conscious thought process while unconscious decision making pertains to a substratum of the conscious mind. In reality, unconscious psychological processes eventually find expression in conscious thought and behavior.

The idea of the conceptual space has been mooted in sec. 1.10; for background, see [14], [53], [54]. In this book, we make use of the idea in a qualitative sense, emphasizing on the enormously complex, convoluted, and nested structure of this 'space'. An alternative visualization of the vast set of concepts lodged in the mind, along with various types of correlations among those, is in terms of a complex *network*.

In this context, it is of importance to note that affect establishes correlations between concepts fundamentally through *analogy* which, however, works mostly beneath the level of awareness — in other words, analogy and affect go hand in hand. Reason, on the other hand, operates by locating *similarities* between concepts. Analogy establishes far-flung (and often wild) correlations in the conceptual space (in contrast, reason establishes sequential but local correlations) — a process that is often computationally intractable. It is this role of affect and analogy that underlies inference-making and creativity. We note in passing that emotions are generated in the mind basically as analogies, marking and classifying our complex experiences in the world.

How are preferences generated in the psychological value system? This question will be addressed in section 5.2.2 in chapter 5.

2.4.1 Affect and emotions: neural underpinnings

Affective neuroscience is a discipline of relatively recent origin that studies the neural basis underlying the psychological phenomena relating to emotion and mood.

Mood can be interpreted as a sustained emotional state, with the cause underlying the generation of the emotion remaining unresolved or, at least, perceived to continue. In particular, mood involves the persistence of negative affect in many individuals and is often found to swing rapidly under a change of context.

I feel humiliated on being ignored by a lady I secretly adore since, in spite of my best efforts I have not been able to draw her attention — this results in a sulky mood where I find fault with everything and everybody. One evening she accompanies me to tea (maybe for reasons of her own) — *suddenly my mood changes to one of elation.*

The understanding of diverse phenomena involving affect and emotion has progressed along with advancing knowledge of the underlying neuronal circuitry [36]. In virtue of increasing recognition of the fact that a great bulk of our cognitive processes occurs beneath the level of awareness, emotional phenomena are explored in neuroscience and psychology by addressing the unconscious layer of the mind to an increasing extent.

While the mechanisms underlying the affect system operate unconsciously, the generation of emotions produces a feeling in the conscious mind. As mentioned above, this raises the question as to how far the emotions represent a genuinely unconscious phenomenon, in respect of which opposing viewpoints have been put forward. In this context, see [148].

As mentioned earlier, emotion is always valenced [77] — either pleasant or unpleasant. The generation and activation of emotions depend on a distributed neuronal system leading to a host of affect-related functioning of the mind. The basic affect network is referred to as the reward system (alternatively the reward-punishment system) or the pleasure system (alternatively, the hedonic system) — in this book we will use the term 'psychological value system' (the *value system* in brief) to denote a neuronal system (and its psychological aspect) that has far-ranging implications in human behavior and in foundational issues relating to human existence as such.

In the *seeking view* of affect, neuronal conglomerates are again identified [150], which overlap with but are distinct from the ones thought to be relevant in the reward view.

However, we will not enter into the vast literature on the neuroscience of affect and emotions (which is an impossible task anyway in a book like the present one, not counting the level of competence of the author in such an undertaking), and will only mention in general terms a few regions of the brain involved in significant affective and emotional activity. While emotions and their affective core cannot be clearly demarcated in terms of neuronal origin and activation, we will refer to the affect system first.

2.4.1.1 Neuroscience of the value system

We include here, for the sake of reference, a few sketchy remarks on the neuronal basis of the affect system or, what will at times be referred to as the *value system* in this book because it is the affect system that is responsible for the generation of *psychological value* that the mind assigns to perceptual inputs of diverse types.

The affect system is located in several regions, including the ones referred to as the prefrontal cortex (PFC), the anterior cingulate cortex (ACC), the hippocampus, and the amygdala ([36], chapter 2). Of these the first two are in the outer cortical layer of the brain — of late evolutionary origin — while the last two are sub-cortical regions of earlier origin. Affective experience and the resulting behavior can be broken up into a number of phases. For instance, speaking of positive affect, there correspond the phases of wanting (seeking pleasure), consummation (actual feeling of pleasure), and learning (storing in memory, planning for the future). Among this, seeking pleasure and the associated motivational drive is largely due to chemical (e.g., dopaminergic) modulation while learning involves higher forms of activity such as integration with past experience and resolution of competing drives due to other extant preferences.

> Among the various types of pleasurable experience, *social* pleasure is to be included as one of major adaptive value ([78], chapter 12).

It is generally the case that the generation of an affective experience takes place in sub-cortical centers, while learning and planning occurs in cortical regions. The learning and memorizing may be of the explicit (episodic or declarative) type, though an implicit component is always present.

Among the brain regions mentioned above, the ventromedial prefrontal cortex (vmPFC, belonging to the PFC; a closely associated and overlapping region is the orbitofrontal cortex, OFC) is of major relevance ([111], [63]). The affective reponse to an environmental stimulus does not originate in the vmPFC, but signals generated in the subcortical centers undergo a lot of processing here by being integrated with diverse sensory and cognitive information so as to result in learning, planning, and the generation of affective *meaning*.

Indeed, the vmPFC has been described as a 'system of systems' since it bridges *conceptual* and affective processes, and thus functions as a hub that links concepts with brainstem systems capable of coordinating organism-wide emotional behavior. Information from various sources and brain centers are integrated "into a gestalt representation of *how an organism is situated in its environment*, which then drives predictions about future events. A straightforward term for such processes is affective meaning, *a sense of the significance of events for an organism's well being and future prospects*" ([111]; emphasis ours). Fundamentally of the nature of an unconscious process, the generation of affective meaning results in a conscious appraisal of the affect in terms of experience, current perception, and planned pursuit.

The integration of affect and conceptual-cognitive functions along with emotion regulation, for which the vmPFC is responsible to a large extent, has a far-reaching — almost unfathomable — consequence as regards the human condition and foundational issues in human existence. In particular, it is of relevance in the formation and ceaseless process of ontogenic evolution of the *self* of an individual, a major component of which relates to *social cognition.* The vmPFC is implicated in the activation of 'theory of mind', i.e., the ability of an individual to attribute 'mental states' (made up of such ingredients as beliefs, preferences, desires, intents, emotions and knowledge) to self and others, and has a significant role to play in moral judgments and in providing support to self-relevant processes. In respect of the processing of self-relevant information, the vmPFC has significant interaction with the so-called *default-mode network* (DMN), a network of brain regions including those referred to as the dorso-medial PFC, the posterior cingulate cortex, and the precuneus.

2.4.1.2 Towards a neuroscience of emotions

Emotions constitute a highly complex aspect of the working of our mind. Emotions are routinely referred to in everyday discourse of ours ('*He had a fit of anger last evening*'), and are addressed in more well-founded terms in psychology, while *neurobiological* investigations on emotions have also made some progress. However, there is a lack of consensus on even a few basic questions relating to emotions in the scientific literature (already a prohibitively vast one), and numerous concepts in emotion theory may undergo substantial modification in the future. We present here a brief and sketchy outline of a number of aspects of emotion theory (refer back, for a beginning, to sec. 1.5) as we understand it, in elementary terms.

We will be concerned more with emotions in general than with specific emotions, referring to those only in passing. Three types of questions are relevant: (a) what are emotions constituted of, (b) what mental and other physiological changes are caused by emotions and how these are related to behavioral changes, and (c) what types of perceptual stimuli lead to emotions of various descriptions.

In this book, we will broadly address the first of the above three types of questions. Generally speaking, emotions are mental states corresponding to neuronal excitation patterns in the brain, associated with the distribution and activation of chemical modulators of neuronal activity. A great deal is now known relating to the brain regions associated with the generation of emotional response, where some of the specific emotion types have been found to be correlated with a multitude of brain regions. However, the actual correspondence between neurochemical configurations and emotions is known only vaguely and incompletely. There do exist ways of decoding, to some extent, well-defined emotional states (pronounced fear, for instance) of an individual by analyzing the spatio-temporal excitation patterns associated with these, but it is still with some justification that one can state that a complete decoding is out of the question because of the *complexity* of the underlying systems (refer back to sec. 1.13; further considerations on complexity will be found in chapter 4). Indeed it seems that a definition of emotion in neurobiological terms is more far-fetched an undertaking than one in *functional* terms where the causal role of emotional states in generating mental, physiological, and behavioral response is accorded primary consideration.

One fundamental question in the field of emotion neuroscience relates to the one of 'basic' or *primary* emotions — where the term is commonly meant to include fear, anger, joy, sadness, surprise, and disgust. To what extent are emotions hardwired in the brain — are there a set of primary emotions that have emerged in the evolutionary process — ones that can be associated with genetically inherited and dedicated neuronal structures, and how are these related to more complex emotions that emerge during the developmental history of an individual? Nathanson [98] has used the term 'firmwired' in respect of a number of basic psychological emotions (nine in number — seven consisting of oppositely valenced pairs, and two unpaired ones) that have a strong hereditary component but, at the same time, depend to some extent on the early developmental history specific to a person — at least, that seems to be the way Nathanson means it in his use of the term.

We have, in earlier paragraphs, adopted the position that affective valence is innate and intrinsic, having an early evolutionary origin. Such clearly defined origin, however, cannot be attributed to the primary emotions, though these can perhaps be explained in terms of innately structured neuronal assemblies with mutual interactions of a complex nature (for background, see [23]). An inclusive proposition admitting of a broad range of theoretical approaches and experimental observations (in animals and humans) seems to be the following:

There exists a spectrum *of mental states answering to the description of emotional ones, to one end of which there corresponds affect — evolutionary in origin and hereditary in nature — innately hardwired in a network of neuronal assemblies, while the other end corresponds to complex emotions arising by means of social interactions modulating innate neuronal structures, generated in the course of the developmental history of an individual.*

Put differently, emotions are psychological states of mind generated in successive layers where the core layer is innate and inherited while subsequent layers are generated in succession during the developmental history of a person, with one layer making essential use of the previously generated ones in marking and classifying more and more complex experiences in social interactions and in experience in general.

Statements such as the ones proposed above are aimed at clarifying the nature of various different *categories* of emotion (such as affect, primary emotions and more complex ones) but do not help us understand what the term ‘emotion’ actually refers to. For this, one needs a *functional* definition — an approach that pays dividends in the case of numerous other types of psychological states as well. For instance, a partial definition that appears to be a pointer in the right direction would state that emotions are functional states that are typically caused by sensory inputs, that typically cause behavioral outputs, and that also cause changes in, and can be caused by, other mental states such as perceptions, memories, and attention ([2], chapter 2).

Another way of characterizing emotions (refer back to sec. 1.5) is to say that emotions classify the complex world of ours in a pre-verbal mode and that emotions serve to amplify and inhibit the effect of certain perturbations on the course along which the mind runs in its incessant encounter with reality. Finally, emotions (along with affect) constitute a great means of *social communication*. While the role of emotions in survival and adaptation in humans seems to be less compared to that of affect (Darwin emphasized the value of emotion in evolutionary adaptation), the emotion system has evolved to serve the purpose of *implicit* communication in social interactions ([2], chapter 2; a useful summary of the properties and functional features of emotions are to be found in chapter 3 of this important book).

Compared to the complex organizaton of neuronal assemblies generating affect and affective meaning, emotions involve a yet more distributed network of such assemblies, as a result of which the neurobiological investigation of emotions, while yielding a stupendous amount of relevant data, have not led to any precise theoretical formulation of sufficient breadth in this field of vast relevance.

Emotions arise from the activation of specialized neuronal assemblies in several parts of the cerebral cortex, notably the ones referred to as the anterior cingulate cotex, insula, ventromedial prefrontal cortex, and also in subcortical structures, such as the amygdala, ventral striatum, putamen, caudate nucleus, and ventral tegmental area (for background, including the important and central role played by the amygdala, refer to [119]). However, the question of the complex communication among these assemblies in the generation of emotional experiences is outside the scope of this book.

> When we turn to the question of how the functional architecture of an emotion — with the focus on specific emotional states — is actually implemented in the nervous system of an individual person, and in turn what clues about the functional architecture the neural mechanisms can give us, we immediately grasp the importance of referring to large-scale neuronal *assemblies* rather than individual neurons, as has been repeatedly underlined in this book. This, indeed, is the tacitly acknowledged approach in describing and explaining

mental states of individuals.

As in the case of affect, the generation of an emotional experience is associated with complex processes of learning and memory storage (especially, the storage of *episodic* memory) — ones that make possible long term planning and motivational drive and leads to conceptual enrichment. The emotional marker on an experience stored in episodic memory makes possible the recreation of a past experience with rich associations. In this context, we reiterate that it does not make much sense to conceptually dissociate emotion effects from affect.

Is emotional experience fundamentally a conscious or an unconscious one? The position we adopt in this book is that, emotions, like affect, originate in the unconscious layer of the mind, while rich and varied conscious effects arise from resulting processes of association and integration (see [2], chapters 9, 10, for a more cautious and informed appraisal; see also [148]). In particular, the feeling of an emotion having been activated and the associated bodily responses are all complex aspects of one and the same basic process that originates in unconscious depths of the mind. Fundamentally, emotions result from the integration of unconscious affective processes with past experiences stored in memory (again,the episodic memory in particular) along with their own emotional tags.

For a readable and informative account of the neural architecture relating to affect and emotions, especially with reference to the three-level (primary, secondary, and tertiary, refer to sec. 2.4) processing in the 'seeking view' of affect, see, for instance, [103]

2.4.1.3 Neurotransmitters in affect and emotions

Affect and emotions, and their effect on cognition, self-regulation, and integration of neuronal activity across brain areas, are generated not solely by the neuronal architecture and organization in the brain, but equally so by the action of *neurotransmitters*. The latter are chemical species that modulate the transmission of electrical signals across synaptic junctions between nerve cells. The two together (interacting neuronal assemblies and neurotransmitter systems) lead to an incredibly rich and complex activity of the brain and the mind, including an exquisitely entangled mix of affect, emotions, and the ‘higher’ cognitive functionalities (‘higher’ in the sense of phylogenic and ontogenic history).

Briefly, the principal neurotransmitters (refer to [146]; see also [105]) involved in affect and emotions are dopamine, gamma-aminobutyric acid (GABA), serotonin, and noradrenaline. Among these, Dopamine can act as both an excitatory and an inhibitory neurotransmitter, depending on the particular synaptic site it binds to. Of the rest, noradrenaline is excitatory in nature, while Serotonin and GABA are inhibitory neurotransmitters.

Dopamine is involved in the brain’s reward system, and can alter our motivation, and emotional states. Dopamine levels are increased when we partake in activities that bring us satisfaction and pleasure, but too high a level can cause us to become hypersensitive to our environment, and can result in disrupted thought process. A shortage of dopamine in the cortex mixed with an excess of dopamine in the limbic system, can possibly lead to anxiety and a hypersensitive mood, narrowing and intensifying our focus, generating the possibility of emotional instability. Overall, dopamine is of great relevance in the generation and control of our emotional states.

Noradrenaline is regularly associated with the body’s fight or flight response to stress. It has the capacity to influence our physical responses, create a sense of urgency overpowering fear, and to cause an increase in heart rate and blood pressure. This neurotransmitter level in the brain is related to both depression and anxiety, with a role that appears to uphold the balance between depression and nervousness. Almost all anxiety disorders involve an elevation of noradrenaline.

The neurotransmitter serotonin was first isolated in 1933, with its role in various psychiatric disorders discovered over the years. It is a critical regulator of bodily functions such as libido level, sleep, body temperature, and is often associated with depression, social anxiety, eating disorders, and obsessive-compulsive disorder (OCD). When levels of this neurotransmitter are low, an individual would experience problems with attention, become unorganised, and lack concentration. As serotonin levels decrease further, simple everyday tasks become overwhelming due to the lack of ability to plan. Consequently, this causes an individual to experience negative emotional states.

Finally, GABA is an inhibitory neurotransmitter, essentially reducing the communication between neurons. Fundamentally, GABA is involved in our level of excitement, which is essential in brain areas involving emotional states. When levels of this neurotransmitter are low in the brain, an individual will be overstimulated, causing communication to become uncontrollable. On the other hand, when levels are high in the brain, an individual experiences impaired reactions, is prone to become excessively relaxed, and even becomes sedated. Additionally, when levels of this neurotransmitter in the brain are within normal or stable ranges, the individual would not be anxious or excessively aroused, with suitable reactions to the surrounding environmental situation. Principally, GABA acts as the body's communication speed controller, ensuring that brain communications operate at the appropriate intensity and speed, so as to suit the situational environment.

2.4.2 Affect and emotions: overview

Affect and emotions are themes of central relevance in this book. We refer here to a number of the more important aspects of the way affect and emotions operate in the mind. This will involve repetitions but will constitute an overview for ready reference.

First, a word on nomenclature. We use the term 'affect' in two senses — once as the core valence (desirable or not, positive or negative) associated with an experience, along with a strength representing the 'arousal level' of affect, and again in a more inclusive sense so as to include emotions that impart complex coloration to affect. We assume that the reader will have no difficulty in picking up the sense of usage from the context. The term 'affect system' will, at times, refer to neural assemblies in the brain where affect and emotions are generated so as to be integrated with other aspects of experiences across diverse other similar assemblies and, at others, to refer to activities of the mind that associate a valence and a set of emotions with an experience.

Affect and emotions are generated in neuronal assemblies across various regions in the brain. As regards the valence-based affect, one can specify with relatively less uncertainty a number of brain areas where this process of generation takes place, while in the case of emotions the brain regions are not so specific since these are experience-dependent to a greater extent.

Valence-based affect serves to generate a fundamental classification of the world that is of the binary type. Every experience is mapped on to a single value axis with a sign and a magnitude, which is of immense relevance in the processes of decision-making and inference-making — two adaptive processes of foundational significance in our mental life (see chapter 3 for more detailed considerations).

Emotions, on the other hand, constitute a more complex classification of the world by means of analogy, and serve the purpose of implicit appraisal of the social and physical reality, making possible a hugely effective non-semantic (non-linguistic) communication between individuals. Emotions are the vehicles by which the mind makes sense of the world without overt reference to relations within the infinitely extended and complex conceptual network. The latter constitutes the basis on which consciousness works, making possible a semantic and *reason*-based appraisal of the world we find ourselves in. Affect (in the sense of valence *and* emotions), is incessantly engaged in a strange and complex interaction with reason, so much so that the two appear to be blended into an inseparable unity.

Affect, in addition to effecting a classification of our experiences and constituting a means of communication between men, acts as an amplifying and moderating influence in the activity of the mind that can be likened to an incessant dynamic in a ‘phase space’ of vast dimensions. At any given point of time, there take place processes described, in the main, as trajectories covering a number of dimensions, some of which get destabilized by the amplifying action of the emotions, initiating a dynamical evolution along new dimensions. However, the instabilities are mostly local in nature and are moderated by means of further emotional mechanisms so that the overall coherence and integrity of the mind is preserved. In other words, emotions play both an amplifying and a moderating role in the activities of the mind, analogous to the effect of positive and negative Lyapunov exponents (see, for instance, [51]) in the motion of a dynamical system in the phase space where, moreover, the exponents vary from one region to another in it. In other words, a de-stabilizing influence in one region of the phase space may be balanced by a stabilizing influence in some other region.

Affect generates preferences in the mind in a complex process that cannot be captured in terms of a shared 'logic' (see section 5.2.2 for more detailed considerations). Indeed, it is the huge set of preferences lodged in the mind of an individual that goes a long way to constitute his self, where the latter depends on the thousand and one details of his developmental history. Affect guides us at every turn of our life and determines our social behavior by modulating our judgments of our fellow-men. As we see in chapter 3, affect plays a crucial role in our decisions and inferences, acting like an 'internal monitor', steering the inferential process to a conclusion that proves to be useful in subsequent inferential activity.

Affect regulates our mood and such aspects of mental states as motivation, commitment, and drive. It is of direct relevance to fundamental existential issues in our mental life such as anxiety, boredom, aggression, and restlessness, and determines what we choose to be.

Finally, affect is vital to the operation of the *soul* (refer back to section 1.14; more detailed considerations on the relevance of the soul in human existence will be found in chapter 5). On the one hand, the soul stands for consciousness and reason guiding and modulating affect while, on the other, it is affect itself that generates the resolve and commitment for reason to regulate affect by intent. The soul is primarily a mechanism for the conscious regulation of emotions. While the detailed mechanism of how this can be made possible is far from known, it is nevertheless within the realm of possibility and is "a huge topic in human psychology" ([118], chapter 2; see chapter 3 of Siegel's remarkable book for a number of indications in this regard). Stated briefly, the role of affect in making consciousness examine the self and restructure it, is strange and intriguing (see [118], chapter 7, for the central role of emotions in self-regulation).

> Detailed considerations regarding the nature, origin, and multifarious role of affect in our mental life are to be found in [2].

2.5 The brain and the mind: developmental processes

The subject of developmental biology of the brain is a vast one. Our focus in this book, however, is on the mind. Developmental processes in the brain result in the dynamical evolution of the mind over various distinct time scales and, eventually, in the generation and alteration of behavior. We include below a brief sketch of the evolution of structure and function in the brain within the limited context of the present book.

A major part of brain development occurs in the prenatal and immediate postnatal periods, and continues to infancy. This phase is mostly controlled by genetic factors inherited from evolution, resulting in the structural divisions in the brain, the production of neurons, and the establishment of major neuronal pathways. A different type of developmental process begins in early infancy where *experience-dependent brain plasticity* provides a means to the emergence of numerous mental functions of fundamental relevance. Foremost among these is affect-laden communication between the newborn and her mother (or the principal caregiver), along with other persons in close interaction with her (the baby). This crucial phase, which witnesses the emergence of the nascent self has been found to lay the foundation for a great many psychological traits of a person, to be manifested in later life.

Experience-dependent brain plasticity continues even to relatively late stages in adult life and consists of establishment of synaptic connections among neurons, along with the process of *pruning* of redundant junctions. It includes the laying of neuronal circuits, mostly localized within regions of limited extent in the brain. These processes are complemented by the adaptation of the neurotransmitter systems to the plastic brain.

A third type of process of great relevance in the mental life of an individual consists of *epigenetic* modification of the genomic material in neurons, modulated by social experience and environmental factors. The epigenetic modifications results in altered gene expressions in various brain regions, leading to modified mental functioning and behavior. At a fundamental level, epigenesis is responsible for the diversity of cellular structure and function in various different parts of the body. However, it is of particular relevance in the neurons in the brain where there takes place a profusion of epigenetic modifications in different brain regions even up to late stages of adult life, among which *experience-dependent* factors are of substantial relevance. While epigenesis follows a more or less rigidly fixed pattern in the other parts of the body, it is a much more flexible process in the brain, where the exact nature and role of this flexibility is being revealed in rapidly progressing research, though a reasonably complete picture is yet to emerge. Of course, epigenetic processes do not occur in constant abundance and, in reality, occupy an intermediate position between programmed differentiation among the various brain regions and the experience-dependent plasticity of the brain, exerting a relatively long-lasting modification of mental functioning in an individual. It now appears that much of the variability among individual behavior patterns in a given social environment may be due to experience-dependent epigenetic processes.

Stress-induced epigenetic processes are of particular relevance in this context. Mental and physical stress and trauma causes epigenetic modifications, some of which may even be handed down from one generation to the next. Stress factors are known to play a major role in the parental care of newborns, where parental stress may produce significant epigenetic changes in the baby, so as to have lasting effects later in her life.

Analogous to the case of stress, an *acute mental experience* appears to be likely to result in epigenetic modifications and a resulting alteration in the mental life of an individual. Such acute experience is accompanied by emotional surge, especially like the one experienced in an event of dire *humiliation*. As we will see, shame and pride constitute a couple of emotions that provide an axis around which the self-system of an individual is assembled (sec. 2.7.4), making possible the distinction between the self and the non-self among entities of the world. Overwhelming emotional experiences of embarrassment, humiliation or shame are known to lead to changes in the personality structure of an individual, that may possibly be traced back to epigenetic neuronal modifications.

> Here is a short list of sample resources on the subject of brain plasticity and epigenetics that may be found to be of relevance in the context of the present book.
>
> A highly recommended source providing general background on neuroplasticity and epigenetics (among many other topics of interest for this book) is [118]. The role of early experience in modifying gene expression that results in altered psychological meake-up and behavior is pointed out in [28]. A more general and detailed summary of epigenetics in brain development and behavior is to be found in [74]. The prevalence of epigenetic modifications in the brain is indicated in [134]. The role of epigenesis in experience-dependent plasticity, including the response of the brain to acute experience is outlined in [25]. Epigenetic inheritance is discussed in [81]. Epigenetics is important for the development of the affect system; see, for instance, [38]. See also [73]. The role of the affect system in early infancy is to be found in [114].
>
> For a detailed general background to the subject of brain plasticity, see [11]. An introduction to the developmental biology of the brain is to be found in [132].

The evolution of brain connectivity is a highly complex process where evolutionary and developmental factors interact in an intricate manner. An optimum evolution of brain connectivity has to take into account a large number of factors — a task that is computationally intractable. Thus the brain evolves in a kind of piecemeal way in which various contrary factors are attended to at different points of time as they attain importance. This leads to enormous variability in brain architecture and, coupled with experience-dependent plasticity, generates almost infinite variability in the working of the mind of an individual.

It is a recurrent theme in this book that the freshly received or freshly generated signals in the brain are integrated with existing excitation patterns so as to lead to *interpretations* of these signals, depending on the existing neuronal architecture. That architecture itself gets modified continuously by experience-dependent plasticity of the brain. As a result, the activity of the mind is the result of an intriguing and complex combination of endogenous and exogenous factors. The mind, in brief is a *non-determinable* dynamical entity as a whole (refer, in this context, to [126], chapters 7, 8).

2.6 Memory: mind's storehouse of history

The *memory* constitutes a remarkable capacity of the mind. It makes the activity of the mind truly complex by making mental processes *history dependent*. As we will see in sec. 4.2.6 history dependent processes are commonly observed in complex systems (a brief introduction to complexity will be given in sec 4.2), in virtue of which such systems acquire an element of non-predictability.

> Generally speaking, the history-dependence of a complex system arises when one attempts to describe the behavior of some part of it in *reducecd* terms, i.e., without overt reference to the time evolution of the rest of the system. In the case of the brain and the mind, the memory introduces an aspect of explicit history dependence over and above the implicit one that appears whenever a reduced description is resorted to.

The memory is not a single or homogeneous entity, and is made up of a number of different kinds of storage systems, which are supported by different brain set-ups. One major distinction can be drawn between working memory and long-term memory. Long-term memory, in turn, can be separated into declarative (explicit) memory and a collection of nondeclarative (implicit) forms of memory that include habits, skills, priming, and simple forms of conditioning. These memory systems depend variously on brain systems such as the hippocampus and related structures in the parahippocampal gyrus, the amygdala, the striatum, cerebellum, and the neocortex (see [127]).

The storage and retrieval of non-declarative memory occurs mostly automatically, within the unconscious layer of the mind. On the other hand, explicit or declarative memory and working memory operate by conscious storage and retrieval, though with the participation of underlying unconscious processes.

The memory systems act as repositories of all the ingredients of the mind including the self-linked and non-self linked ones. Among the non-self linked ingredients are concepts, items of knowledge (facts) and shared beliefs in declarative memory, and procedural skills and primed perceptions in non-declarative memory. *Events* are recorded in what is referred to as episodic memory, among which the biographical events are of special relevance since these are constituents of the *self*-system. It seems likely that preferences (generated by the affect system) and the emotion-laden beliefs are stored as non-declarative memory. In accordance with our fundamental premise, all these items are stored in memory in the form of dynamical excitation patterns that get regenerated on memory recall.

In recording an item of experience (this includes mentally generated constructs) the memory stores the composite picture of it as originally experienced — of course, innumerable details of the experience are ignored and only those aspects that are salient with reference to the state of the mind at any given point of time are retained. In this there takes place a good deal of *interpretation* by associated cognitive systems so that the record is tagged with similar records stored in the past and some appropriate system of *cataloging* is invoked for subsequent recall.

The exact mechanisms involved in the storage and retrieval of memory are not known, and a few theories have been proposed. One plausible point of view is that an experience generating a set of dynamical excitation patterns in various neuronal assemblies is stored in the *same* set of locations in which these excitations were created in the first place, with a set of *links* created in specialized cataloging zones that keep a record as to how the various excitations specific to an experience are to be integrated with one another — this includes, in particular, their temporal relations and their relative salience. On storage, some kind of switch turns off the excitations to below-threshold limit while, on recall, the switch is turned on from the cataloging zone and the details of integration of the various aspects into a single whole are supplied back so as to regenerate the experience (refer to [32], chapter 6, for an analogous point of view). The act of memory recall is initiated by some cue from the current state of the mind, including the current set of perceptual inputs that activates the cataloging zone. Looked at from this perspective, a point of distinction between unconscious and conscious memory records relates to whether the cue is activated unconsciously or by conscious means. Likewise, the storage of self-linked and non-self linked ingredients of the mind differ in the way as to whether or not affect-based inputs are linked to the cataloging zone. The various different memory systems would differ in the locations of the cataloging zones and in manner in which they operate.

2.7 The self

2.7.1 What constitutes the self?

The self is made up of a special class of states, processes and dispositions of the mind, all generated from perceptions of the world. We begin by briefly explaining the term 'disposition'.

> What does the term 'disposition' mean?
>
> A disposition stands for a tendency or a propensity of the mind to respond in a particular way on being presented with a certain class of stimuli. In this book we adopt the foundational position that all states, processes, and dispositions of the mind are represented by means of dynamical excitation patterns in neuronal assemblies having diverse spatio-temporal structures. This, however, is in the nature of a big assumption, supposed to provide a *possible* basis to all our considerations in this book. The space of all possible spatio-temporal excitation pattern is then supposed to represent the 'phase space' of the mind (refer back to sec. 1.16), where the phase space is also nothing more than a notional construct here. In terms of the metaphor provided by the notions of the dynamical excitation patterns in neuronal assemblies and the phase space of the mind, a 'state' of the mind or a disposition possibly corresponds to an excitation pattern in the nature of a *stationary wave*, while a process corresponds to a *traveling wave*. This once again, is nothing more than a convenient way to visualize the states, dispositions, and processes of the mind.
>
> Incidentally, mental configurations commonly referred to as 'states of the mind' can, from a broader point of view, be looked upon as dispositions since a state of mind also leads to a certain response when presented with certain

stimuli. Incidentally, when we speak of a configuration of the mind, or of a mental state, we actually refer to the configuration in some particular set of neuronal assemblies. Each such neuronal assembly can be thought to have a phase space of its own, and the overall phase space characterizing the mind may be looked upon as one composed of all these component spaces.

In this book we place special emphasis on *preferences* and *beliefs* which are nothing but dispositions of particular types. Other than these, repressed tendencies such as desires, drives, fantasies, yearnings, resentments, hostilities — all constitute dispositions of the mind. These are the ones we mean when we make liberal use of the phrase ‘preferences, beliefs, and dispositions’ in this book — at times, we simply use ‘preferences’ or ‘dispositions’ or, say, ‘beliefs’, for the sake of brevity.

In this context, the term ‘intent’ also deserves mention. While a disposition is often an unconsciously generated inclination of the mind, an intent is a inclination of the conscious mind in some specified context. Intent is typically the way the conscious mind responds to specific sets of circumstances it is confronted with.

Fundamentally, the self is built up on the basis of affect and emotions, and is constituted of an enormous number of *preferences* and *aversions*, and emotional *associations*. In particular, the emerging self of a newborn is entirely affective in nature, when the differentiation between the self and non-self in the mental world of the baby is yet to take shape. Other than the set of preferences and emotional associations, the self of an individual is made up of emotion-laden cores of concepts (i.e., proto-concepts), beliefs with emotional associations, along with unconsciously held proto-beliefs, and a large number of emotionally driven dispositions such as desires, fantasies, deeply rooted aversions, deep resentments, anxieties, aggressiveness, and similar other mental attitudes. Among the dispositions of the type mentioned here, are the suppressed and latent ones that constitute the repressed self (refer back to sec. 1.14). All these constituents of the self are generated from current perceptions and, additionally, are associated with stored or recalled *memories*, by means of which past experience is integrated with current percepts.

For a readable introduction and background to the self and its affective core, see [79]. Aspects of Kristjánsson's views on the self are of relevance to what we put forward in the present book. Detailed background material on the topic of self is to be found in [84], [52], [117].

However, the self is not made up exclusively of dispositions dominated by affect and emotion. The conscious component of self also includes concepts, beliefs, intents and dispositions (i.e., ones that have been referred to as 'neutral' ones (see sec. 2.3.3)). Put differently, perceptions and memories involving non-self entities can also be assimilated to the self, provided these have some association of a conceptual nature (i.e., one not necessarily affective) with the currently constituted affect-laden self. For instance:

Sitting among the audience while attending a seminar, I find a person on the dais among the invited speakers, whom I recall as a friend I used to play with in a shabby courtyard some fifty years back. This perception gets included as an accretion to my self-system even though the person on the dais does not generate discernible emotional feeling in me. However, *when, during his speech in the seminar, he does a bad job of communicating with his audience, I feel vaguely disturbed* because of that linkage to self.

1. The self is tied to the concept and the sense of *one's own person* — the physical and biological entity that one is. It is the physical and biological being that serves as the basis of the distinction between the self and the non-self, where the term 'biological' is to be interpreted as including the *psychological* collection of preferences, beliefs, and dispositions, along with the repertoire of heuristics lodged in the mind. The self, in other words, refers to a physical, biological, and psychological entity.
2. The self of an individual is known to others by her behavior pattern — her response to situations she faces. It is her behavior that is determined by her dispositions.

2.7.2 Social interactions and the self: the self corresponds to a unique developmental history

The self of an individual is largely generated by means of social interactions since such interactions predominantly involve affect and emotions.

> Exceptions to the above statement are not rare. For instance: *On looking at a scenic view of breathtaking beauty, I am flooded with a strange feeling of joy that I am certain will stay with me for the rest of my life.* However, even here, the sense of beauty generated in me has a socially generated component involving my developmental history.

In countless events of interaction with other persons in society we generate a vast set of likes, dislikes, attachments, resentments, reasoned assessments of people mixed with complex feelings, and a large number of other similar perceptions, that get assimilated into our self, and define the distinction between the self and non-self. It is through such associations that we develop our habits, ideas, ideals, commitments, modes of reasoning, our beliefs and desires, our behavior pattern, our attitude to life — all that constitutes our individual personalities. The self is a product of *specific experiences* that we happen to have throughout our life, i.e., of the unique developmental history of an individual. Since no two persons can have identical developmental histories, *each of us possesses a uniquely constituted self.*

The mind is in ceaseless interaction with the environment — social and physical. This is an interaction between two exquisitely complex systems, carrying their own internal dynamical processes, themselves enormously complex. The developmental history of an individual can be looked upon as a trajectory in a 'phase space' of vast dimensions, resulting from this interaction along with the two associated sets of internal dynamical factors. One realizes that the developmental history of an individual is so unique that it bears no comparison with the developmental histories of other similarly endowed individuals even under similar environmental conditions.

2.7.3 Self: the special role of affect and emotions

This uniqueness is augmented even more when one takes into account the role of affect and emotions that generally play an *amplifying* role in mental processes. Such amplification can be compared with the role of positive Lyapunov exponents ([51]) in the phase space of a dynamical system. In the presence of positive Lyapunov exponents, the dynamics of a system is endowed with the feature of *sensitivity to initial conditions* — small perturbations in the state of the system (say, due to external influences) result in a relatively large effect on the trajectory causing it to deviate significantly from what it would be in the absence of the perturbations. However, the deviations do not grow indefinitely since, in general, counterbalancing tendencies are brought to play in the case of complex systems. Put differently, the deviations remain confined to some relatively small subspace of the phase space, giving rise to a new mode of behavior while the system as a whole retains its integrity.

As an instance of how emotions can lead to extreme sensitivity in the evolution of the mind, consider the following:

Two children of nearly the same age, growing within the common environment of a family, have similar endowments and receive similar training and instructions. One day the two are asked to solve a number of arithmetic exercises when one of them receives a bit of extra praise and appreciation from a number of elders in the family on her performance while the other, by default, felt slighted in some measure. This seemingly insignificant event triggers a course of development of far-reaching consequences, when the first child engages feverishly in mathematically oriented activities and self-training, eventually developing great mathematical skill, while the second child shies away from mathematics and becomes indifferent to exercising his mental faculties.

The sensitivity to perturbations resulting from the amplifying effect of affect and emotions endows the developmental history an individual with a degree of uniqueness that defies comparison.

2.7.4 Shame and pride as the axis around which the self is assembled

While affect and emotions are generally associated with the self, it appears to be the case that the emotions of *shame* (humiliation, embarrassment) and *pride* (being acknowledged, praised, or glorified) are two emotions that have a very special role in respect of the self-system of an individual [98] — these two provide the axis around which the self develops and grows. Indeed, shame and pride prove to be instrumental in distinguishing between the self and the non-self aspects of the mind since the self is exquisitely sensitive to these two complex social emotions.

A baby likes nothing more than to be appreciated by her mother, and learns things with remarkable rapidity when appreciated appropriately. On the other hand, a feeling of shame, humiliation, and embarrassment deflates and deactivates the self and the mind, in a sense, loses the services of its driving engine. The power of the shame-pride axis in modulating the activities of the self-system becomes apparent when one looks probingly at the *self-image* of an individual — a person invariably tries to project to himself that he has a lofty nature and stands for human dignity. For instance, when a beggar approaches him for a paltry donation, he may send him away with the admonition that he (the beggar) should value his dignity and engage in physical labor rather than beg from strangers. When he feels that someone is criticizing him or making fun of him, he either starts sulking or attempts to 'even the account' by criticizing that other person on some count or other — one may even discern a 'balance sheet' maintained with meticulous care where every incident of real or imagined slight is 'answered' or responded to in some appropriate way. It is self-pride, commonly referred to as 'ego', that is of supreme value to a person, where he wants to be held in admiration not only by others but by himself even though in actual fact he may be engaged in unworthy activity.

In real life, much of our self-system gets built by trying to draw recognition and admiration and to avoid humiliation where it is *perception* rather than reality that has the last say — for instance, one has to perceive others as acknowledging his supremacy while in reality those others may secretly make fun of him for his hubris.

2.7.5 Self: emulation and social empathy

Of equal importance in the matter of building up of the self is our tendency to emulate those we hold in esteem or, on the other hand, to behave oppositely to ones we disparage (at times, we unconsciously emulate a person we strongly dislike, out of awe or a sense of insufficiency). Emulation is closely linked with *empathy* that modulates much of our social interaction and our self-system. Empathy is fundamentally an affect-guided process though there may be admixture of an affect-free component too in our empathic behavior. A heightened form of empathy is observed in *resonance* (see [39] for background) — whereas empathy may be one-sided, resonance is a two-way interaction in which two persons share intimate and privately held thoughts, preferences, and values, again, primarily by affect-based communication.

The neuronal basis of empathy and resonance has been explored to some depth, and neuronal systems promoting empathy have been identified. As for emulation, a neuronal system of considerable relevance that human beings share with other primates is the so-called *mirror neurons* ([27], chapter 14; [143]) that is likely to have affective links in leading to imitation at an unconscious level.

The *genetic* aspects of empathy have been outlined in [40], chapter 14.

2.7.6 The self and the non-self: connected by group perceptions

We have, at earlier places in this book (see, for instance, sections 1.9, 1.10), contrasted the self with the non-self, the two constituting distinct systems in the mind where the latter refers to mental ingredients and processes defined in relational terms and having little to do with affect and emotions. While the self is based on the unique developmental history of an individual, the non-self is made up of *shared* entities in the mind.

However, the mind is too complex a system to make definitive and clear-cut statements about. Speaking of shared ingredients such as preferences, beliefs and dispositions, one always has to qualify as to the *degree* of sharing. Only very few mental resources of an individual can be said to be *universally* shared, and most instances of sharing are restricted to various *social groups* the individual may belong to — the family, friends at school, colleagues at workplace, people having common political affiliations, adherents of a common religious faith, cultural groups, and so on. The interactions of an individual with persons in each of these social groups causes a number of specific features to appear in his developmental history, in consequence of which his self acquires, so to speak, a new layer in its evolving structure.

> In other words, the self of an individual may be thought to have a layered structure, with a layer being added by means of his interactions with people in some particular social group or other.

> We raise ourselves from the confines of the proto-self by looking at the wide world around us — but in reality, we do this only by adopting *group perceptions* characterized by various degrees of generality. As an infant, we identify ourselves with mother. As we grow, various other layers of identity are assembled by means of group perceptions of various degrees of depth and breadth, gradually extending the scope of the non-self in the mind, and expanding and enriching our network of concepts — affect gets blended with 'neutral' processes based on relations between entities of the world. The *implicit* and intrinsic gets blended with the *explicit* and the extrinsic.

As the self grows in social interactions, the *non-self* part of our mind acquires new structure too, with new concepts and enhanced correlations appearing in the conceptual network. This enrichment of the non-self part of the mind arises by means of new concepts, new preferences, new beliefs, and new dispositions being acquired in virtue of group affiliation. For instance:

My younger brother was a shy and bookish guy when in school. As he was admitted to the history department of a big university he became aware of political affairs, turned into an adherent of socialism, started spending endless hours with friends, and became extremely sure of himself in most matters of life. He read voraciously, but had time to *attend political meetings and to set out on long journeys, and absorbed a great deal of history, science and social matters. His world underwent a remarkable transformation and expanded rapidly.*

2.7.7 The 'group-self'

Speaking of specific sets of preferences, beliefs, and dispositions acquired by an individual in virtue of his membership in a group, one observes the converse phenomenon too, where specific sets of preferences, beliefs, and dispositions are found to constitute the identity or *'self' of that group.*

Indeed, as indicated in sec. 1.15 (see also section 5.2.8 for further elaboration), a group of people, having a common aim and a common way of functioning, can be looked upon as a collection having a 'mind' of its own, where the 'neurons' giving rise to the mind are, precisely, the individuals constituting the group, while dominant sub-groups play the role of neuronal assemblies . This analogy is far from being a vacuous one, in that the interactions among the individuals and among the sub-groups in a given environment in which the group finds itself, give rise to processes that can be termed the 'response' of the group, where these processes occur with some 'purpose' designed to satisfy some of the preferences that the group comes to possess. Conflicts within the group, caused by diverging preferences of the individuals and subgroups, are moderated by *power relations* within it and by means of 'administrative measures', in the face of which there arise latent resentments and propensities, akin to the repressed self, within the group. Indeed a sufficiently large group of individuals such as an entire society with its own culture, administration, and economy can in itself be looked at as a complex system, much like the brain of an individual, and the analogy between the mind of that individual and the 'mind' of the group (the 'social psychology') becomes even more apparent.

We will not pursue here the diverse aspects of social psychology (see [27] for considerations on a number of aspects) and take for granted the existence of 'non-conscious' and 'conscious' layers in it (refer, once again, to section 5.2.8 for further considerations), as well as the affect-based and the concept-based aspects of the perceptions of the group. The non-conscious part of social psychology involves processes that run autonomously in it such as the day-to-day activities in the large number of agricultural firms in a country that run independently and in parallel, while the conscious part corresponds to co-ordination among the units running in parallel and various purposeful activities such as the ones relating to storing, marketing, and exporting of the products, where specific *rules* and regulations are to be invoked. The relevance of these considerations will be apparent later when we consider the *soul* of an individual as the capacity of his conscious mind to look at his own self and to reconstruct it from time to time — the ultimate instance of *regulation* that can operate in his mind. That same process of self-regulation is also possible, and is indeed essential, for an entire society of humans to flourish by the process of having a sincere look at its own latent beliefs and dispositions impeding its onward journey in this world.

> The self of an individual and the 'self' of a social group to which she belongs are *mutually* determined. For numerous situations of practical interest, however, we can provisionally distinguish between the two. This helps in matters of description and explanation provided that the significance of the underlying links between the two are duly recognized. This, indeed, is the idea behind the concept of *decomposability* between entities constituting a complex system (refer to sec. 4.2.1 in chapter 4).

> Thus, the preferences and dispositions of the individual are generated by induced effect of the social group while group preference (i.e., the set of preferences held by the group as a whole) is produced by dominant individuals and sub-groups within it.

2.7.8 From proto-self to group-induced shared self

For now, we revert to the consideration of the self and the non-self components in the mind of an individual, and begin by clarifying a few terms. When we speak of 'self' as such, we refer to an individual. In contrast, the term 'group-self' refers to the mind-like identity of a social group made up of individuals, the latter acting as 'neurons' of the group. On the other hand, the preferences, beliefs, and dispositions of an individual that she shares with a social group will be referred to as her 'group-induced self' or 'shared self'. There may be numerous layers of shared self, depending on the various social groups that a person may belong to. All the layers of shared self eventually build up the non-self component in the mind of an individual.

> Group-induced self is, in part, unique to an individual and in part shared by others in some social group or other. It links the self associated with the unique developmental history of an individual with the *non-self*.

Put differently, the self of an individual has a complex structure made up of layers of shared self growing on the proto-self where the preferences, beliefs, and dispositions arising from any particular layer cease to be unique to the individual under consideration, and there exists an entire spectrum depending on the degree and extent of sharing, with the proto-self at one end of the spectrum and *universally shared* beliefs and dispositions at the other.

> We will later have occasion to refer to the question of 'universally shared rules' and their significance. Recalling that beliefs act as rules establishing correlations between concepts (and between other beliefs too), a likely candidate for universally shared beliefs would be the rules of mathematical logic, while the theories of natural science constitute a close approximation. However, there still remains the question of existence of universal preferences and dispositions, if any, of *mankind*. This is where humanity has to grope for an answer, and grope for ever.

In other words, the preferences, beliefs, and dispositions (we will, at times, use the term 'belief' in an inclusive sense, referring to all the three taken together; at times, the term 'disposition' is used in this inclusive sense) of an individual can be grouped into two classes — one of these consists of beliefs unique to her, not shared with any larger social group, and the other of those that are shared. Among the latter group of beliefs are to be distinguished various classes corresponding to the various different social groups to which the individual may belong. For instance, there may be a group of five or six intimate friends with whom she may share a few beliefs (say, about the probable involvement of a common acquaintance in a love affair), or, on the other hand, there may be a number of culturally defined dispositions shared with a much larger group of people (faith in the powers of a certain deity, for instance). The self-linked beliefs of an individual unique to her will from now on be understood as the 'uniquely personal' ones. Other beliefs, corresponding to the 'self' induced by various social groups of persons will be termed 'shared' beliefs, where the degree and extent of sharing may differ for beliefs belonging to this class. In this context, it is important to note that *heuristics* (refer back to sec. 1.11; see 2.8.2 below for further considerations) are to be included among the beliefs of an individual.

Recalling that beliefs are, generally speaking, of the nature of rules correlating concepts and other beliefs, one can now see how rules of varying degrees of generality are involved in defining our self, which includes various layers of group-induced self. The uniquely personal beliefs of an individual can be looked upon as rules of the least possible generality, while other beliefs, shared by progressively larger groups of people constitute rules of progressively higher degrees of generality — ones we inclusively refer to as *shared* beliefs or rules. As we will see in chapter 3, this distinction regarding our beliefs assumes great relevance in the making of *decisions* and *inferences*. Shared beliefs and ideas enter into the constitution of the 'non-self'. In other words, the self is connected continuously to the non-self by shared components of an over-all world-view.

2.7.9 The self and the non-self: self as the observer

While keeping in mind the layered structure of the self referred to above, we are not to lose sight of the distinction between the self and the non-self that the mind is stamped with, though, at the same time, the two are continuously connected. The unconscious mind does make a distinction between the self and the non-self though we are, naturally, not aware of it — the distinction being created by affect-linked processes. It is affect that continues to generate the distinction between the non-self and self in the conscious mind too but now the non-self consists of objects of awareness in virtue of widely flung relational links between concepts that assumes basic relevance in processes of the mind.

It is apparent that we have to resign ourselves to some degree of ambiguity and contrariness here, as in every other question relating to the mind. The self and the non-self are not absolutely distinct compartments of the mind. What we have called the uniquely personal actually lies at one end of a spectrum spanning the uniquely personal and the universally general. While the uniquely personal self (actualy, the proto-self in the strict sense) is made up of strictly limited conceptual relations and affectively tied preferences and beliefs, the non-self is made up of shared mental resources with successive layers of sharing characterized by progressively widening relational links and progressively loosening affective ties. In other words, the non-self is, in reality, nothing but *shared self* with various degrees of sharing with social groups. At the distal end of the spectrum is the non-self made up of concepts, ideas, preferences, and beliefs shared by humanity at large — more an ideal than reality.

In order that the non-self may be assembled in our mind, we have to look outward at the world. That outward view, however, is possible only through the eyes of our fellowmen, mostly collected in social groups of various descriptions.

We recall that the mind is fundamentally a huge and complex *process* made up of a vast number of component processes. These processes are represented by dynamical excitation patterns in neuronal assemblies where the spatio-temporal structure of a pattern evolves over time. Concepts, in this scheme, are also represented by such patterns, generated through processes of perception and interpretation by means of interactions between neuronal assemblies, where the patterns representing the concepts themselves evolve slowly over time since the concepts get enriched by acquiring new relational links.

Perception of the external and the internal worlds is the fundamental job of the mind, where perception is accompanied with interpretation and response. Earlier, in sec. 2.3.2 we have used the term 'observation' in order to refer to perception by the conscious mind.

Recall, incidentally, that the conscious mind can even make a partial map of the self (the ontological self, that is), whereby the self appears as an object, generating the *concept* of the self. In a manner of speaking, and somewhat paradoxically, the self acquires membership of the repertoire of non-self objects that can be reflected and deliberated upon by the conscious mind.

A special class of observations made by the conscious mind are *observations by the self*, these being ones that are affect linked, though 'neutral' observations (i.e., ones relatively free of affect; refer back to sec. 2.3.3) can also be included in this category provided these are linked to the current self-system of the individual under consideration. These self-linked observations are very special where the *self appears as the observer, interpreter, and knower* — the objective 'me' assumes the role of the subjective 'I': in brief, *the self confronts the world* now as a subject. While observation relates to the basic act of perception, interpretation and acquisition of knowledge are acts of cognition that the self engages in, pursuant to the act of observation. In each of these, the self appears in a dual role — on the one hand, as the brain-based arena where all the relevant neuronal processes occur and, on the other, as perceptions by a special subsystem of the conscious mind, namely, the self.

Observations by the self add a definite flavor to observations by the conscious mind, not to speak of perceptions of the world in general. This is because of the link to affect and to the unique developmental history of the self. Thus, consider:

I stand idly at roadside and observe a bus moving by; I get a momentary glimpse of a person seated in the bus whom I had met some time before on a business I had some interest in, that association having been an unpleasant one. In this case, this perception of the bus moving by gets associated with my self-system and it is the subjective self ('I') that acquires the status of an observer having had a view of the moving bus.

There is another way that the self turns into an observer, namely, when there is some *plan* of one in making an observation. For instance:

I stand at roadside taking count of the frequency of buses going through at a particular hour of the day, this being a job entrusted to me by the local civic body. Here again, I become the observer, as against the conscious mind observing things without links to my ontological self. In the present instance, the link is not overtly an affective one, though, ultimately, an affective tie continues to be there through the plan that I intend to carry out — it must have been generated by some affective means (*receiving compensation for my time*).

In other words, as in the case of the unconscious mind, the perceptions made by the conscious mind may or may not be self-linked where, in the latter case the mind becomes the ‘disinterested’ or ‘neutral’ observer while, in the case of links to the self being established, it is the self that assumes the status of the observer (or the reasoner or knower, as the case may be). In the case of the self being the observer, the mind is focused on some entity or some event in the world as compared to the ‘disinterested’ observation which is often in the nature of a relatively desultory one. A thought process or an observation may be based predominantly on relational and non-self aspects but still may have a remote connect with the self through some commitment or plan, when it is attended by a heightened awareness compared to desultory mental processes. This is the case, for instance, when one is focused on a job one is committed to (*working on an experimental investigation as part of a research project*).

2.7.10 Attention is related to self

This leads us to the observation that *attention is related to self* (refer back to sec. 2.3.5 where attention was distinguished from consciousness in general), i.e., the transition from conscious observation without attention (this we refer to as ‘neutral’ or ‘disinterested’ observation — one basically desultory in nature) to one involving attention is essentially a transition where the self assumes the role of the observer. Another way to get at this idea is to note that attention is, in some sense, related to *interest*, which in turn is related to affect and to self. For instance, one becomes attentive of a situation when she is startled by it. Here the emotion of being startled brings in the self that rises to the role of the observer. Consider, for instance, the following:

I am going through a sheaf of experimental data gathered by a student of mine when I suddenly come across a set of data that seems to connect to a conjecture that I had made earlier and had almost forgotten about. I begin to look through the mass of data in feverish expectation of other regularities in it that may lend support to my conjecture.

2.7.11 The self, the non-self, and the default mode

The borderline between neutral (i.e., solely relation-based) observations and observations made by the self (ones possibly attended by interest and attention) corresponds to the *default mode* of the mind, where the default mode network (DMN; refer back to sec. 2.4.1.1 where the term was introduced) takes over. In this mode there is a continual switching between a range of idly performed operations of the mind based on affective and relational terms. The default mode, in other words, provides *a meeting ground between the self and the non-self* (refer to [152],).

It is the default mode that points to the possibility of the conscious mind and the subjective 'I' accessing the objective 'me' (in this context, see [35], [34] [15]). This relates notably to the activity of the *soul* that we turn to in chapter 5.

Finally, it is the role of the self, based on the unique developmental history of an individual, that can explain the unique subjective experience permeating perceptions of ours, made up of aspects referred to as *qualia*. This however, does not specifically belong to the range of notions of direct relevance in this book.

In chapter 3 we will turn our attention to how the mind makes decisions and inferences — processes where rules of various degrees of generality play their role. These processes of immense significance in our life are modulated in a complex manner by the self, with all its complex structure arising from layers of preferences, beliefs, and dispositions accruing from the membership of an individual in larger social groups. At the same time, the self will also be seen to be implicated in a big way in the exercise of *free will* (chapter 5) where the complex relation between *determinism* and *determinability* will be found to be of fundamental relevance.

2.8 Revisiting the world of beliefs

In this section we repeat and clarify the notions introduced in sec. 1.11.

Beliefs, like preferences, are in the nature of dispositions that make us behave in certain ways when presented with stimuli of various specific types. For instance, the belief that a certain person is dishonest makes us avoid being engaged in financial transactions with him. In sec. 1.11 we found that beliefs have important and pervasive psychological role in our life. We use our web of beliefs as a map with the help of which we navigate the world, incessantly invoking these in inferential and decision-making acts, when these generate new beliefs that get incorporated in the currently existing web of beliefs lodged in the mind.

Beliefs themselves can be looked upon as concepts having complex structures, since these can be objectified and reflected upon. One has to distinguish between two things here: the content of a belief acting as a conviction of some sorts in the mind of a person, generating a disposition in her: '*My neighbor is a liar*' — here the content of the belief has the psychological role of orienting the believer against the person alluded to. On the other hand, a belief can be an object of contemplation: '*I don't think that I am wrong in my belief that he is a liar*' — here I am alluding to my belief (a certain person being a liar) as an object of reference about which I am positing another belief of mine (to the effect that the belief referred to is a justified one). In the present book we will refer to beliefs in either of the above two aspects — the intended use will have to be picked up from the context.

The big question that arises in this context is, how justified are our beliefs? We have noted that the beliefs span a wide spectrum, with fully confirmed and justified beliefs, converted into items of knowledge, at one end and patently false ones at the other. In between, there exist beliefs with various degrees of justification against evidence, where *emotions* provide the ground for their continued presence in the great web of beliefs in our mind. In other words, our beliefs form a coherent system where the coherence is not based exclusively on their logical consistency with one another, but on logic *along with* emotions. Beliefs of a completely affective nature lodged in the unconscious mind have been referred to as proto-beliefs.

'Logic', as it is referred to, is extraneous to the mind of an individual, and is based on rules of a broad generality, arrived at by a process of distillation and abstraction where the applicability of a rule is freed from specific situations to all situations within a given context. In contrast to universally valid rules, there exist rules of varying degrees of generality. Indeed, beliefs themselves constitute instances of such rules of less than universal generality.

Affect and emotions, on the other hand, constitute a 'logic' intrinsic to the mind. An inference may be formed solely in virtue of the fact that it agrees with a preference set up by the affect system.

Finally we have also noted that beliefs blended with emotions act as potent forces that spur us into action. Our response to signals and stimuli received from the world (including the world within) is generated in virtue of our beliefs, where it is to be mentioned that the term 'action' includes mental ones in the form of initiation of trains of thought — a stimulus generated in one part of the underlying neuronal assembly may lead to a thought process of some consequence by way of some belief having been invoked.

The action-promoting role of beliefs is essentially due to their tie-up with emotions and to the *amplifying* action of emotions in ongoing mental processes (refer back to sec. 2.7.3).

Justified beliefs that have been substantially freed from emotional baggage also lead to action, this time by means of *reason*.

In the context of the present book, the principal feature of interest is the extent to which a belief is linked to the self of an individual or, alternatively, is shared by larger groups of people to which that individual may belong.

The lesson we have learned from sec. 2.7 is that *pure* self without any involvement of shared dispositions is an idealized concept that applies only to the proto-self of a new-born baby. Generally speaking, self-linked and shared aspects of beliefs are blended with one another in virtue of layers of group-induced self over the proto-self, with a corresponding mix of intrinsic and extrinsic 'logic' in these.

Here are instances of shared beliefs: An individual belonging to a certain community may entertain a belief that the number thirteen carries a bad portent and may avoid making important business appointments on the thirteenth day of a month. On the other hand, a member of some other community may offer homage to a certain deity before embarking on a long journey, again based on a shared belief.

Beliefs shared with other individuals in virtue of membership in a social group act as reason having the general *if-then* structure (*if* homage *then* auspicious journey).

As mentioned above, shared beliefs are to be contrasted with personal ones that depend on the unique developmental history of an individual, and are affect-based. For instance, *my mother did not like respectable women to laugh loudly in the company of strangers; in consequence, my thought and action at times make me avoid* (irrationally so, to be sure) *the company of ladies who have the habit of laughing loudly.*

2.8.1 Belief revision.

The important thing to note about beliefs is their dual nature — most of our beliefs are tied with emotions, but to varying degrees, and they are guided by reason to a complementary extent. They are resistant to revision under the impact of new evidence, but again, to varying degrees. Beliefs do change, but change in a strange and complex manner. At times, a highly recalcitrant belief that we stubbornly shield from being confronted with evidence gets revised dramatically in the face of some apparently insignificant event. The tension between the force of reason backed by evidence on one hand and emotion-based resistance on the other pervades the entire span of the widely flung web of our beliefs, owing to which there arise local instabilities in the web, and a dramatic switch-over to a new stable configuration by the appearance of new beliefs and a re-alignment among earlier ones. For instance:

I was of the strong opinion that children were to to be dealt with sternly. One evening I found my next-door neighbor harshly admonishing his little daughter for a minor offense, and within a few hours the child had a severe bout of illness, perhaps coincidentally. That single event, followed by a phase of rethinking, has changed my outlook and I am now a staunch believer in soft and tender dealing with children.

Such precipitous change is analogous to a tsunami or an earthquake at some place of the globe due to a small impact on some fault-line deep down the surface of the earth, that generates a meta-stable configuration to appear in some region of limited extent — globally, the geological configuration remains stable. This, once again, is an instance of 'sensitivity to small perturbations' typical of complex systems, and typical of the role of emotions in the mind. Some other beliefs, however, get revised more smoothly by the weight of evidence, and eventually attain the status of knowledge, having been adequately justified.

On the whole, the revision of beliefs is a complex affair, and is riddled with contrary aspects. Of major relevance in this context is the *multi-layered* structure (see chapter 4 for further considerations) of the vast conceptual network of ours, in which beliefs act as links establishing correlations among concepts. In such a network, there exists multiple layers of links between nodes (i.e., the concepts or other beliefs in the present context), and the restructuring of one or a number of links results in a new configuration where there appears both a continuity and a disconnect with respect to the earlier one. In the context of revision of beliefs and of our scientific theories, this is referred to as *incommensurability* , and was deliberated upon by Polanyi, Kuhn, and Feyrabend, among others (in this context, see [68]).

> We note, incidentally, that beliefs themselves are in the nature of concepts having a complex structure. What is more, with the development of consciousness, beliefs become objectified and themselves serve as nodes in a highly convoluted network of concepts, in addition to serving as links between concepts in the network.

As one set of beliefs gets modified or built afresh, there is brought about a change in one or a few of the multiple layers of links in the conceptual network while the other layers are left largely intact. As a result, the new set of conceptual correlations differs significantly from the previous one but, with reference to the previous configuration, there remains a continuity whereby an earlier set of beliefs retains a certain measure of validity and there remains a common area of discourse where the concepts within the earlier configuration can be understood in terms of those in the newly emerging one while, at the same time, the landscape of beliefs undergoes a change. Consider, for instance, the following:

I was of the opinion that my friend, Mr. A, was a highly educated and an honest and *genuine person. I have now been told that he has had to serve time in a penitentiary for a couple of years. I am now in doubt if I will continue to keep in touch with him.*

In this example, the entire scenario of beliefs seems to have undergone a change, but the identity of Mr. A as an educated person and his possible status as an acquaintance continue to remain partly intact, though my new set of beliefs regarding the person are clearly incommensurate with the earlier one that has now turned out to have been defeasible. One notes that the change in the belief structure cannot be described as an increase in the 'truth content' of the earlier set of beliefs. For the sake of comparison, consider:

I was of the opinion that my friend, Mr. A, was a highly educated and an honest and genuine person. I have heard someone say that he has perhaps been appointed to a government job of high responsibility, and am convinced that Mr. A will very shortly rise to a position of high stature in society.

Here there is no question of my earlier beliefs being called into question and, if anything, the 'truth content' of those beliefs has increased.

In summary, belief revision is often a complex process, owing to local tensions (between affect and reason) in the wide web of beliefs lodged in our mind. This frequently gives rise to a precipitous revision of the beliefs under apparently small impacts of evidence, though more smooth processes of revision also occur. Belief revision often results in the new configuration of beliefs being incommensurate with the earlier one, where new evidence does not necessarily increase the 'truth content' of our set of beliefs about 'reality'.

Here we have used scare quotes on the terms 'truth content' and 'reality' since these need a bit of explanation. Both reality and truth are slippery concepts. While there certainly exists a reality independently of our beliefs, that reality cannot be described in ultimate terms. Nor can it be approached indefinitely close through a smooth process of revision of beliefs since some subset of our beliefs is always at complete odds with aspects of reality. The term 'truth content' refers to a measure of truth inherent in a set of beliefs, for which Karl Popper coined the term 'verisimilitude' — even so the truth content seems to elude all 'logically' precise definition. Incidentally, we occasionally use quotes on the term 'logic' since logic is seldom of universal validity, the case of mathematical logic being the single exception, and that too within the context of mathematical derivations. Generally speaking, the logic we employ in various different contexts remains valid within those limited contexts only, being dependent on beliefs of varying degrees of validity.

2.8.2 Heuristics: rules of thumb

Heuristics were introduced in sec. 1.11 as makeshift beliefs that often come in handy in the making of decisions and inferences. Heuristics may be lodged in the non-conscious or conscious layer of the mind, and may be self-linked in a large or small measure. Generally speaking, heuristics are more easily discarded if found to go against evidence as compared with emotion-tied beliefs and are often in the nature of short-cuts to justified rules or principles.

In physical problems, corollaries to *symmetry principles* often constitute a prolific source of heuristics. If a problem in dynamics, for instance, has some inherent symmetry built into it then such corollaries can provide clues to the solution to the problem without the necessity of having to work out a detailed solution by integration of the relevant equations of motion. For instance, the result that the motion of a particle in a central field of force remains confined to a plane can be straightaway concluded from symmetry principles without solving for the equations of motion. In mathematics, one often uses lemmas to prove theorems, where the lemmas are in the nature of mid-way posts to theorems and act as heuristics. In these examples, heuristics are true statements that are used as intermediate constructs in arriving at a result.

Generally speaking, most of our inferential activities are conducted in a hierarchical manner, where results found to be useful in one set of inferential processes are made use of as intermediate rules in a broader class of inferences. In other words, inferences that have been found to be more or less successful in one stage of the hierarchy are used as heuristics in the next stage. In games of chess, for instance, one uses distinct types of game-plans in various stages of the game, where the game-plans are invoked hierarchically — only those classes of moves in the opening game are tried that produce an advantage in the mid-game, and so on from the mid-game to the end-game and finally, expert players make use of separate sets of heuristics in the end-game itself.

The same applies to innumerable small and large inferential activities that we undertake in our daily life where, once again, hierarchical processes are of common occurrence. Heuristics made use of in these inferences are variously known as 'common sense', 'wisdom', 'rules of thumb', 'gut feeling', 'lay-man's approach', 'experience', and so on. In choosing a good career for our children, we commonly inquire as to how our neighbors have chosen for their kids. Having identified two or three promising alternatives, we seek more specialized knowledge regarding the institutions where one can have training for these careers, and finally, we look into available statistics relating to the job prospects that trainees in these institutions have had in recent times. On finally making a choice, we pass on these bits of information as heuristics to others making similar queries later. All the while we remain aware that the information and the intermediate inferences arrived at by such means may not be fully reliable but still we often find it convenient to go by these.

In other words, heuristics are in the nature of beliefs that have proved their worth in some situations, being retained for use as intermediate resources for inferential activities in a broader terrain, not unlike semi-cooked product being used in the final stage of cooking a delicacy, even though that semi-cooked product may not be fully appropriate for achieving the best possible culinary effect.

Heuristics, in other words, are beliefs generally enjoying a good degree of justification, even to the extent of being fully justified in some situations. In those cases in which a heuristic has some degree of misfit with evidence, it is more readily modified in conformity with the evidence than is generally the case for an emotion-linked belief. On the other hand, a heuristic may be quite poorly justified in a real life situation such as a lay person's advise to a patient, where that person thinks himself to be eminently competent in the role of adviser. If the advise proves to be useful, the patient puts much faith in it and passes it on to other sufferers even though the advice may later turn out to be quite useless.

What is important to note is that, even when a heuristic is found to be useful in some practical situation, it may be a *self-linked* rather than a shared one. For instance, one can have a large number of heuristics lodged in one's mind for arithmetic calculations. Every child 'discovers' sort-cuts for arithmetic problems such as those of addition and multiplication — some trick that works well (for instance, *in multiplying 19 with 13, we first multiply 19 with 10, then with 3, and then add*). However, a mathematical prodigy has a far greater number of useful tricks lodged in her head compared to others less gifted in the field of mathematics, most of which she (the prodigy) works out in her mind at leisure or while working out problems. A grown up mathematician of high stature (say, a Ramanujan or a Gauss) possesses a vast number of heuristics in his mind, most of which may not even be dreamt of by other mathematicians. Some (or even many) of these may, of course, be shared by these other mathematicians, but it is precisely the enormous number of other gems of heuristics lodged in the mind of a Ramanujan or a Gauss that makes him so special as a mathematician. Some specific heuristic developed by an individual can be shared by a peer but the way that particular heuristic is invoked and made use of in association with other heuristics lodged in the mind may be self-linked, even associated with affect to a good extent.

The same goes for the game of chess where an expert player ceaselessly, at every waking moment of her life, keeps on looking for good moves in all kinds of imagined situations on the chess-board, and it is the vast number of exquisitely tricky moves stored in her mind as heuristics that makes her so very remarkable a player.

While a large number of heuristics lodged in the mind of an individual may be self-linked and may, therefore, depend on her developmental history, including her nature of specialized training, if any, her intensity of practice in some given area of activity (cooking, for instance), her level of involvement with that activity, and such matters specific to her, many of those may have a higher degree of affective and emotional associations too. For instance, a mathematician may be particularly proud to have worked out in her mind a lemma by invoking an especially tricky and elegant approach, which she may jealously preserve as her private asset, not intending others to learn of it. Indeed, a number of mathematicians have been known to be particularly jealous and self-centered in protecting their discoveries (especially, the relatively less spectacular ones that they use as heuristics) from the knowledge of their peers — some are even known to have gone to great lengths in making a secret of their derivations of difficult theorems that have defied other mathematicians.

Finally, heuristics can be lodged in the unconscious mind too and, once again, may or may not have affective associations. A little child keeps on generating and making use of a great many unconscious and affectively tied heuristics such as ones relating to anticipated behavior of her mother and other dear ones. On the other hand, a craftsman's apprentice, in learning a skill, unconsciously generates in her mind heuristics that need not have affective associations — how to hold a tool, how to make the best use of it, and so on. Likewise, a surgeon possesses a vast store of unconscious heuristics of an objective nature regarding the internal anatomy of a patient and the skillful use of his surgical tools in relation to the anatomy, many of which may have only minimal associations of an affective nature..

In summary, heuristics are beliefs of varying degrees of justification that we constantly work out in our mind in our daily inferential experience as well as in specialized areas, many of which are constantly in the process of being discarded and modified in the face of evidence. Heuristics are best made use of in a hierarchical approach, when beliefs that have found justification in some stage of an inference are used in the next stage and the process is continued, with the repertoire of useful heuristics stored for repeated use in subsequent inferential processes. Heuristics can be used consciously or unconsciously, and may or may not be affect-laden, but they do have some link with the self, even if weak ones. They are of great use in all fields of specialized skill and performance.

Heuristics are of great relevance in *artificial intelligence* (AI) where they form, in a manner of speaking, the backbone of learning algorithms. This, however, is not of direct interest in respect of the framework that we aim to put together in the present book.

2.9 The bigger picture: the brain, the mind, and the society

This is where we look at the bigger picture in which the brain, the mind, and the society make up an integrated whole. Within this bigger frame, integrative and regulatory processes run side by side so as to produce a coherent wholeness where, at the same time, disruptions and instabilities appear incessantly, at times to be repaired at a relatively early stage and at other times, to break out in major turmoils, shaking the very foundations of human existence.

This book is not the place to engage with neurobiology on the one hand, and with bigger social and philosophical issues on the other. Our job in it is to focus on the mind while pointing to the essential links between the mind of an individual, the underlying neuronal processes, and the overarching social reality modulating the mind.

The mind is an exquisitely complex system ('system' not in the sense of a material body) where integrative and regulatory activities go on side by side to maintain its coherence. It emerges as a spectrum of properties of the brain and, in turn, is modulated by *interpersonal* relationships — one brought to existence in virtue of interactions among individuals belonging to social groups of diverse descriptions.

The coherence of the mind as a whole is brought about by large scale integration of the activities in distributed neuronal assembles, effected by electrical and chemical means. The chemical control of the mind is exerted principally by means of neurotransmitters that modulate our affect and emotions while the electrical control is exerted by means of electrical excitation patterns along neuronal circuits and pathways connecting the various brain regions. The excitation patterns appear mostly as large scale oscillations across brain regions, covering a wide span of time scales, some of which are observed as brain waves recorded in electro-encephalography. These chemical and electrical activities are indicative of the fact that mental processes are emergent aspects of brain activity, constituting an instance of emergent properties of complex systems (refer to sec. 4.2.8 in chapter 4).

For background on the vast significance of oscillatory rhythms in brain activity, see [18].

The integration of brain activities of diverse types does not occur in vacuum — it occurs within a complex matrix provided by *experience*. And experience is mostly *interpersonal* in nature. It is experience that modulates the connectivity among neurons by the process of *neuroplasticity* (refer to sec. 2.5). Of equal importance is the fact that the existing store of spatio-temporal excitation patterns in the diverse brain regions make up a huge *library*, with reference to which the representations formed out of the incoming signals from the world are *interpreted* so as to acquire *meaning* in the mind.

It is in this context that the emerging subjects referred to as 'interpersonal neurobiology' (IPNB, [118]) and 'social neuroscience' ([27]) have gained in importance. In this little book of ours, these two constitute the backdrop against which we focus on the activities of the mind, which includes the self and the soul as subsystems. In this, we have also taken into consideration how the mind engages in comprehending and explaining the physical reality where theories are constructed to explain natural phenomena and processes. The two together — the social reality and the physical reality, along with the deep and fathomless reality of our own *self* — make up reality as a whole, with which we as individuals are in incessant interaction in our ongoing journey in life. — it is in this journey that the soul proves to be of central relevance.

Chapter 3

Reason and affect holding hands

In this chapter we make use of the notions introduced and explained in chapters 1, 2 in looking at processes of *decision* and *inference* — cognitive-adaptive activities of major relevance in our life. We will see how each of the two processes involves a complex interweaving of *reason* and *affect*. By 'reason' we mean processes and approaches based on relations between concepts in the conceptual network in which shared rules of some generality are made use of — relations of implication based on these shared rules are commonly referred to as rules of 'logic'. Thus 'reason' and 'logic' are commonly taken to be co-extensive. On the other hand, affect is a psychological ingredient that depends in an essential way on the *self* of an individual with its unique developmental history that contrasts essentially with shared mental ingredients like reason and logic. In other words, the 'logic' of mental processes based on affect differs essentially from the one based on shared rules, the latter being extrinsic to the self. Put differently, the contrast between affect and reason is nothing but the contrast between the *intrinsic* and the *extrinsic* when looked at from the point of view of an individual. It is within this context that the inferential processes of an individual is often diagnosed as *irrational* when compared with 'norms of rationality'.

The contrast between the extrinsic and the intrinsic is related to that between 'necessary' and 'contingent', between the 'abstract' and the 'concrete', or between the explicit and the implicit. It may be mentioned that the most pervasive theme running through the entire gamut of considerations in this book is that *real life is no respecter of dichotomous distinctions*. If anything, real life makes a tangled mess of precisely stated definitions and notions. The processes that take place without a moment's pause in the mind of an individual are all an essential synthesis of what appears to us as the *intrinsic* and the *extrinsic* since there is no such thing as 'pure' self and 'pure' non-self — the self and the non-self blend together. The 'pure' categories are useful only in simplifying the great complexities inherent in life where, however, even a notional simplification often takes away all that is of essential relevance. The simplified categories serve, though, as points of reference from which we find it convenient to construct our concepts of real-life entities — an approach that can take us only so far, and no further.

In this book, the dichotomous approach will be followed to some extent, if only because that is the dominant paradigm of discourse in current literature. and I have been brought up within the culture of that discourse. However, the complex reality is a vast tangle, fundamentally based on the principle of *unity of opposites* — opposites are inextricably united, and can never be prized apart.

Traditionally, two brain systems are distinguished — the 'cognitive' and the 'emotional', each associated with its distinctive sets of mental processes (see, e.g., [102]). Within this framework, the term 'cognitive' can be interpreted to include both unconscious and conscious processes. Unconscious cognition has been briefly referred to in section 2.2 and covers a wide range of adaptive processes. In the present chapter we will be concerned to a greater extent with conscious processes where cognition is supposedly based on 'reason', i.e., shared rules of various degrees of generality that constitute a 'logic' extrinsic to the self of an individual. We will, however, see that reason and affect (the basic emotional system) are strangely and intricately intertwined in the making of decisions and inferences — acts so essential for our survival and onward journey in life. From a neurobiological point of view too, the traditional distinction between the cognitive and the emotional brain systems appears to be in need of revision as suggested in [102]:

"This work demonstrates that emotion and cognition are deeply interwoven in the fabric of the brain, suggesting that widely held beliefs about the key constituents of 'the emotional brain' and 'the cognitive brain' are fundamentally flawed."

As mentioned earlier several times in this book, affect operates fundamentally at an unconscious level. Thus, emotion-cognition interaction also occurs mostly at unconscious depths of the mind. For general background in this context, refer to [89].

3.1 The making of decisions: how affect permeates reason

Suppose you are going to have to make the following momentous decision:

Whether to let your son follow the dictates of his heart so as to take up a lifetime career of social work, or to have a showdown and force him to enter the medical profession, which has been your own lost dream in life.

How do you decide? You have, before you, a choice between *disparate alternatives*. Most of our choices in life, even the most mundane ones, are like this: *with limited money in my purse, do I purchase a set of crockery sorely needed for family, or a remote-controlled toy robot to make my daughter smile with delight*?

Beliefs, emotions, and affect provide for a 'simple' but effective strategy that the mind adopts in solving the impasse. In expanding on this statement, we recall from section 2.4 the examples of the happy child and the terrified deer, involving instances of how a real-life situation may raise more than one emotions in our mind and how the resultant effect of these emotions can be *represented in terms of a single value with a sign and a magnitude*. In either of these two instances, each of two contrary emotions is associated with affect, generating a positive or negative valence with some magnitude, i.e., an arousal level. In the resulting situation of push and pull, the valence with the dominant value prevails.

> In this context, one has to face the question as to what determines the arousal level in the event of some particular emotion having been generated. This is, like so many other things in our mental life, a complex question because it relates to untold details of past experience, to the complexity of current perception with all its multiple aspects, their relative salience, and with the appeal to the unconscious of *sub-liminal* components of perception. The arousal level of an emotion is determined by the complex interaction of all these factors. It is highly unlikely that the determination of the arousal level can be reproduced by any externally generated algorithm — in other words, this may be one more of the *intractable* aspects of our mental activity.

What is important to note here is that, in a situation of potential conflict and indeterminacy, the affect system is capable of producing a valuation along *one single value dimension* in terms of a sign — positive or negative, good or bad, pleasant or aversive — and a magnitude, i.e., how strongly positive or how strongly negative (refer back to sec 2.4.2). This is not unlike the way a computer algorithm is designed to produce, at certain junctures in going through the execution of a program, two (or more) signed numbers, each with a magnitude, on the basis of which the next step in the program execution is to be chosen (by identifying the number with the largest magnitude or, possibly, by the simple addition of the numbers involved). It is the valuation along a single value dimension that can be compared with a number with a sign and a magnitude.

> Of course, the mind does not literally generate a number since all it is capable of is to produce a neural excitation pattern. What the affect system does is to produce a pattern in some neuronal assembly that, in some sense, can be interpreted as a number with a sign and a magnitude. As has been stated several times in this book, the correspondence between our concepts and the space-time dependent excitation patterns is likely to remain fundamentally undecipherable.

With this background in place, we indicate — in abstract but simple terms — a possible strategy that the mind can adopt in making a choice between disparate alternatives. Suppose that I have to make a choice between alternatives 'X' and 'Y'. We assume that X may raise in me a number of emotions (say, 'A', 'B'), possibly linked with beliefs, each with its own valence. The affective cores of the emotions activate the affect network to a degree corresponding to the net affective value, say, x (recall from section 2.4 the examples of the happy child and the terrified deer as to how the generation of multiple emotions can produce a net affective valuation along a single value dimension, having a sign and and a magnitude). In a similar manner, the emotions (say, 'C', 'D') raised by 'Y' generates an affective value, say, y. The value with the larger magnitude would then determine my choice .

It goes without saying that the above account can only be a very vague hint for the actual neural processes – akin to a *computational* procedure – that take place in the complex affect-emotion network in the human mind. Nevertheless the following emerge as general features of the process of decision-making.

1. In the making of a decision, where one has to choose between disparate alternatives, the mind makes heavy use of *self-linked* psychological resources, namely, beliefs, emotions, and affect. In other words, the computational procedure does not follow a *fixed* algorithm determined by Nature – the algorithm (if we call it that) varies from person to person, though the general scheme followed remains the same and can be assumed to be the product of biological evolution. What is of importance is to note that the computational procedure resulting in a decision depends on the *developmental history* of an individual in that the self-linked psychological resources evolve throughout her life in a complex manner.

2. Generally speaking, the self-linked resources operate beneath the level of conscious awareness of the person concerned ([43], [125], [140]; see, however, [99]). Though the process leading to the making of decisions has to be anchored in her neurobiological set-up and hence is not random, it remains, at the same time, fundamentally *unpredictable* to an observer and, perhaps, even to herself. While the general *scheme* that the process follows can be unearthed bit by bit in days to come (the above paragraphs are in the nature of a suggestion), the decision itself cannot be mechanically reproduced.

3. The basic strategy by which the mind makes the decision problem tractable is the one of making use of the activity of the affect network, owing to which the response to a perceived situation (calling for a choice to be adopted), based on the affective cores of all the various emotions and mental responses triggered by it, ultimately reduces to an evaluation that maps onto *a single value dimension*, analogous to the case of a computation whose output is a number having a magnitude and a sign. It is the putative value output that is to be eventually used in adopting a choice among alternatives in the execution of a program. Among the candidate alternatives, the one generating the maximum value is chosen. In other words, the affect network reduces the perception of various different situations to a *common*

currency [77].

4. While the affect-driven process indicated above, is essentially 'irrational' from the point of view of 'norms of rationality' it is, at the same time, perfectly possible that, instead of a number of emotions being activated in a perceived situation, one adopts a reasoned computation based on some particular set of epistemically shared rules and the computation again produces an evaluation that can be represented by a single number (with a magnitude and a sign) for each of the alternatives involved. In that case too it would, in the final analysis, be the affect network that would pick on the numbers and lead to the choice made – *this time in a predictable manner.* The predictability may, more generally, be probabilistic in nature.

   In other words, a decision can also be arrived at in certain circumstances by following a rationally determined procedure based on epistemically shared rules . However, as we have repeatedly observed, *shared rules are not necessarily universal ones*, and there exists a wide spectrum of beliefs acting as effective rules that are tied to the self of an individual *as well as* to dispositions of larger groups to which that individual belongs. In any case, the decision-making process involves predictable and unpredictable aspects woven into it.

Decision theory, in the last few decades, has taken several turns. Until recent years, the making of decisions was assumed to be free of emotional involvement, and was sought to be described in terms of 'maximization of utility', looked at as a *rational* process. This phase of rationality-based explanation was followed by the 'heuristics-and-biases' approach propounded by Tversky and Kahneman (see, for instance, [60]) where the role of hitherto unexplored features of the human cognitive process, including those relating to beliefs and biases, was recognized. This approach ushered in a new era in which the focus shifted somewhat from the account based principally on rational optimization procedures. However, even the heuristics-and-biases approach stopped half-way in giving due recognition to the workings of human psychology, and has more recently been replaced with the affect-based theory of the making of decision, of which an outline – as I understand it – has been given above.

The role of emotions and the reward-network in decision-making has been discussed in more concrete terms in numerous publications, including [45], [87], [86], [7] (in which other relevant papers are cited), and [111]; [60] provides a detailed survey of the emerging subject of *Neuroeconomics*. The role of emotions in Decision-making has also been discussed briefly in [2].

Having said this, it is important to note that the affect theory can also be viewed as an incarnation of the utility theory where the term 'utility' is to be interpreted in a new context – the one of affect-based valuation. In the earlier rationality-based theory, utility was introduced as a feature that reduced the response (to a perceived situation) to the much-needed common denominator (often referred to as a 'common currency'), thereby eliminating disparities among the alternatives from which a choice was to be made; however, it was left unspecified as to what exactly the utility function was constituted of. Having introduced the utility function by fiat, it was assumed that the making of a decision consisted of a rational optimization procedure (often intractable from a mathematical and computational point of view) in a manner independent of the psychological vagaries of an individual. In the affect-based theory, one has a formal analogy with this rationality-based one in that the 'utility' is now to be interpreted as the valuation resulting from the activation of the affect network and the 'optimization' is now a psychological procedure where the self-linked resources such as the beliefs and emotions are involved. The latter is now no longer 'rational' in the sense of being determined independently of the psychology of the individual. In other words, unlike the utility-based theory, the affect-based one tells us that the 'optimization' can no longer claim to be a *normative* approach.

3.2 Inference: the strange interaction between reason and affect

Inference is a process in which we engage all the while in our day-to-day activity and in more specialized inquiries such as in scientific research. When we receive an input generated from the external environment or from internal mental processes, and that input poses some kind of challenge (say, an out-of-the-ordinary observation) that does not find a ready-made explanation at hand, we have to engage in an inferential act, based on external cues, internally stored beliefs and rules generated in past experience, and — *affect and emotions* too.

Apart from acts of inference that can be identified as such, we keep on making 'microscopic' inferences — ones that are scarcely noticed in order to be identified. For instance, while driving a car, one has to make inferences almost every moment so as to avoid undesirable encounters and to keep on the right course. Some of these inferences are in the nature of predictions — what is likely to happen the very next moment — and some others emerge as implications of unconsciously followed rules (see, for instance, [9], chapter 4).

We keep on making inferences implicitly and unconsciously all the while, most of which are inductive in nature (see sec. 3.2.1 below). Such implicit inferences are common in non-human species as well, the capacity for which is inherited in the process of biological evolution (see, e.g., [109]).

The conclusions arrived at in an inferential act are subsequently compared with evidence (once, again, the act of comparison can be an unconscious one). Successful inferences generate positive affect — either implicitly or explicitly — in the mind (likewise, wrong ones generate a negative affect) and help in subsequent acts of inference. Successful conclusions generated in an inference often serve as heuristics that come in handy in subsequent inferential acts.

3.2.1 Inference: inductive and deductive

Inference is commonly classified as being either *inductive* or *deductive*. However, while the two differ noticeably in the *form* of the problems required to be solved, their distinction is not so clear when the actual processes taking place in the mind are considered in details.

Consider (a) *All human beings are mortal / Mr. A is a human being / Mr. A is mortal*, and (b) *all ravens observed till now have been found to be black / all ravens are black*. Clearly, the two inferences are of distinct types — one is from the general to a particular, while the other is from particulars to the general. However the psychological processes (refer to [48], chapter 1) involved in inferential acts cannot be demarcated into two clearly distinct classes, since these generally involve an inseparable admixture of *reason* and *affect*.

An inductive inference is one where the conclusion does not follow uniquely and invariably from the premises, and one needs to make a *choice* or a *decision* from among possible alternatives at one or more junctures in the inferential process. Deductive inference, on the other hand, admits of no such uncertainty and produces a unique conclusion from a given set of premises. Clearly, no inference in real-life situations can be purely deductive, and some inductive element *always* remains in the inferential process, introducing uncertainty and defeasibility in the conclusion arrived at. The necessity of making a choice in inference makes the latter appear as essentially a *guess-work* — though an *informed* one— or an exercise in *logical leap*. In the following the term 'inference' will be used to refer to any inferential process that has some involvement of choice, decision, or guess-work in it, while the term 'deductive' will be used when the inferential process is overwhelmingly choice-free and is based on a set of pre-determined rules of wide acceptance. Strictly speaking, only mathematical derivation can claim to be truly deductive in this sense, being based on *universally valid rules of mathematics*.

> However, even mathematical reasoning can have an essential involvement of inductive processes in it. For instance:
>
> *"It is most unlikely that more than a tiny minority of mathematical theorems were ever in fact arrived at, "discovered", merely by the exercise of deductive reasoning. Most of them entered the mind by processes of the kind vaguely called "intuitive"; deduction or logical derivation came later, to justify or falsify what was in the first place an "inspiration" or an intuitive belief."* [94]

In trying to set up the derivation of a mathematical theorem, a mathematician, regardless of her level of expertise, comes across junctures where she has to decide as to which axioms, lemmas, rules of inference, or other known theorems to use so as to be on the right track and to eventually converge on the proof. Had there been no question of a choice or a decision, there would not have been anything challenging in the act of derivation — the challenge is met precisely by someone whose self-linked repertoire of heuristics is sufficiently enriched to suggest some special trick that does the job.

Inductive inference spans a stupendous spectrum of processes in our daily life and in specialized inquiry. Even a child has to engage in inductive activity frequently and ceaselessly and in this, she proves herself to be no less 'original' and 'clever' than the scientist who makes a great discovery — the latter draws our attention and admiration (as the child is admired by her mother who goes ecstatic over her daughter's marvelous feats) partly because she (the scientist) has had to engage in deductive activity too that requires a distinct and specialized expertise.

No *theory* of induction, however, has been developed in spite of centuries of efforts on the part of philosophers and scientists since, unlike deduction, there is no universally accepted 'logic' of induction. It has eventually been named the *scandal of philosophy*, though its seminal role in scientific discovery has also earned it the nickname of *glory of science* [16].

As we are going to see, inductive inference can have no universal logic simply because it has to depend on prior *experience* of individuals and of larger social groups, i.e., on their developmental history that cannot be formulated in general and abstract terms. From the point of view of an individual, the inferential process involves in an essential way her self-linked psychological resources as well as those specific to larger groups to which she may belong. What is special about scientific investigation is that *univerasally* valid logical resources are added to this inseparable mix of self-linked and shared resources in a process of justification and confirmation. However, *no ultimate justification* can be found even for scientific inferences that, in the end, remain *necessarily* defeasible.

3.2.2 Reason and affect in inference

3.2.2.1 Inference: the intrinsic and the extrinsic.

Every act of inference, in reaching from a set of premises to one or more conclusions, has to make use of a set of rules by means of which one can successively generate intermediate stages of the inference. Here we distinguish between the individual engaged in the act of making the inference and a second person that tries to understand or to decipher the rules made use of by the inference-maker. For the sake of concreteness, consider a psychologist trying to see if a subject (someone asked to engage in some act of inference, with adequate explanations as to what the inference is to be about) invokes rules of inference of an intrinsic or extrinsic nature (with reference to him, i.e., the subject in the experiment).

> Countless experiments of this type have been reported in the literature where the psychologist tries to ascertain whether the subject performs in accordance with standard 'norms of rationality'. Such experiments were launched in a big way in the early sixties of the last century, beginning with *Wason's selection task* ([90], [129]) and similar other tasks designed for the purpose.

It is found that, generally speaking, the subject makes the inference by making use of a set of *intrinsic* rules even as the psychologist tries to fit the response of the subject with a set of *extrinsic ones* that can be formulated and described in precise terms. These extrinsic rules are ones shared by members of some sufficiently large group or *community* of people (say, the community of psychologists or mathematicians) and the subject is asked to engage in the inferential job to check whether his (the subject's) mental processes answer to these shared rules. Here the shared rules are assumed to be 'rational' ones but unfortunately, in most rationality-testing experiments, it turned out that the subjects failed the test of rationality (see, for instance, [129], [90]).

3.2.2.2 The use of self-linked rules in inference

With all the background so far built up in earlier sections of this book, the general explanation of this phenomenon — the 'irrationality' of inferential processes of the mind — almost suggests itself: *inference makes use of self-linked psychological resources of an individual.*

However, as we saw in sec. 2.7.6, the self — continuously connected to the non-self — is assembled in 'layers', corresponding to preferences, beliefs, and dispositions of the various social groups that one belongs to in various different capacities. The layer that the membership of a group causes to insert into the self of an individual is, in some sense, an image of the 'self' of the group itself, made up of its (i.e., the group's) sets of preferences, beliefs, and dispositions as distinct from the corresponding selves of other groups. Indeed, as outlined in sections 1.15, 2.7.7 (see also sec. 5.2.8 for further considerations), there runs a deep analogy between a group and an individual in that both are complex systems with constituents interacting with one another where the behavior of the system (the group or the individual as the case may be) is determined by 'mental' processes generated by interactions among macroscopic assemblies of the constituents.

Analogous to the self of an individual, a group also possesses a 'self' that determines the group behavior predominantly through the detailed developmental history of the group, without heed to neutral relational aspects among entities in the world and to broader rules of inference shared by *larger* groups of people that envelope the group in question.

As a simple example, *a sovereign state frequently allows its foreign policy to be driven by its very own economic and political interests, and by the desire to recapture its past glory of domination over other states, without heed to voices calling for global co-operation.*

We have also observed that beliefs and heuristics act as rules of inference, establishing correlations among concepts and beliefs, and generating enriched concepts by way of inference. The above considerations then tell us that there exists a wide spectrum of rules of inference depending on the extent to which the rules are shared in virtue of membership to various social groups. In this context, the following instance is relevant too:

The mother of a little child works as a nurse in a hospital. As her child falls ill, she meticulously observes all rules followed in the hospital that she, as a nurse, is obliged to comply with in the interest of sound patient management. On the other hand, on her husband's advice, she calls in a doctor not qualified to diagnose the child's disease which, in fact, turns out to be a complicated one.

Here the lady in question invokes rules followed by the group of her peers — the nurses in the hospital — but, not being a member of the medical fraternity, goes by rules shared by her husband in the matter of calling in a physician. With reference to both sets of rules (one shared by her peers and the other by her husband), she is likely to have had some affective involvement too (confidence upon the peers, knowing that they are following hospital mandate, and affectionate faith on husband, mixed with the urge to follow authority), where privately held beliefs and shared ones —affect and reason — merge into a complex whole.

3.2.2.3 Affect as an internal monitor and guide

Affect tells us what is desirable and what is not, where the terms 'desirable' and 'undesirable' cover a wide terrain. For instance, it is affect that tells us, in times of indecision mixed with urgency, whether a particular course we adopt is 'correct' or not. This, of course, brings us back to the role of affect in the making of decisions (refer back to sec. 3.1), though in a somewhat different context — making a decision in the course of an act of inference, possibly consisting of multiple intermediate stages. We consider a simple example first.

As I am in the process of sending my daughter off to school, I discover that she is running a slight fever. Daughter tells me she can't stay back because she is going to have an important class test scheduled for the day. But I hesitate, knowing that her father would not like her to attend school with fever. On an impulse, I order her to stay home.

Here, my motherly instinct and emotions may have been responsible in tilting the balance in favor of my decision in an inferential process, putting the stamp of approval on it.

The predicament of *Bouridan's donkey* is a case in point. When unable to make a 'rational' decision in a given situation, we have to have recourse to affect by way of invoking self-linked preferences and beliefs. Affect pays no heed to externally justified logical principles, and is based on purely private preferences, aversions, wishes, fantasies, whatever. If 'principles of rationality' fail to show us the way, the *self* takes it upon itself to break the impasse and point to the 'correct' course. In a broader sense, it may actually be the rational thing to heed to the dictates of the self when extrinsic reason fails us. It may be noted in this context that preferences too are built upon past experiences and are frequently based on memories of successes and failures in past inferential acts.

It is in this manner that affect plays the role of an internal monitor and guide, either approving or disapproving our decisions in the course of an inference being made.

In a deeply insightful essay on induction and intuition in scientific inquiry, Peter B. Medawar, the pioneer in the field of tissue transplantation, writes:

”In real life, of course, just as the crudest inductive observations will always be limited by some unspoken criterion of relevance, so also the hypotheses that enter our minds will as a rule be plausible and not, as in theory they could be, idiotic. But this implies the existence of some internal censorship which restricts hypotheses to those that are not absurd, and the internal circuitry of this process is quite unknown. The critical process in scientific reasoning is not therefore wholly logical in character, though it can be made to appear so when we look back upon a completed episode of thought.” [94]

While Medawar writes of this ‘internal censorship’ in connection with the generation of new hypotheses in science (see sec. 4.5.4 for an account of an ‘internal monitor’ in *creativity*), it is fundamentally in the context of inductive inference that affect plays the role of an internal censor or a monitor, issuing the stamp of approval in all our decision-based acts of inference. Though Medawar speaks of the workings of the affect system as ‘quite unknown’ as indeed it was at his time, there has since grown a vast literature on the mechanisms of the affect system (see, for instance, [12]), commonly referred to as the ‘reward system’ or the ‘reward-punishment system’ in neuropsychology (a complementary approach, initiated by Panksepp, speaks of the ‘seeking system’ — see sections 1.6, 2.4). In this book, we use the term ‘affect’ to denote the valence of an emotional experience — the psychological core that tells us whether the experience is a desirable or an undesirable one, while the terms ‘affective’ and ‘affect system’ are, at times, also meant inclusively to refer to the emotional flavor of an experience. We also use the terms ‘psychological *value* system’ for this purpose, since the affective valence results in an assessment of an experience — even a complex one — along a single value dimension whereby it is marked ‘positive’ or ‘negative’, along with a certain strength or ‘arousal level’ (refer back to sec. 3.1).

With this background in place, we will now enter into the description of an inferential process in somewhat greater details.

3.2.2.4 Inferential processes make use of rules and beliefs of varying degrees of generality

As in everything else in real life, inferential processes span a spectrum with purely deductive inferences at one end and highly original and creative inferences at the other. We take the case of a deductive inference first.

Suppose that a mathematics graduate is asked to establish an algebraic identity. More likely than not, she will take only a moment to write out the derivation without having to think twice. In this, she makes use of *universally* accepted rules of mathematical derivation. Though the knowledge of these rules is confined to people with a mathematical background, anyone can go through the process of mastering these rules (at least the ones needed to establish a simple algebraic identity) and checking that these retain validity in all mathematical derivations, and are consistent with the whole of mathematics (we mention, though, that consistency check is not a trivial matter).

> However, if the same person were asked to establish a really difficult identity or a theorem, her thought process would be quite different. Instead of sequentially making use of a set of shared rules that guarantees a successful derivation, the mathematician would now have to fall back upon a set of heuristics and lemmas stored in her mind that may not be shared by many of her peers. In the process of derivation, she would have to guess, at certain junctures, as to which rules of derivation would lead her to success. Her success would depend on her command over mathematical principles, i.e., eventually, on her past training, her level of commitment, and her store of heuristics, all of which, even though truth-based, would be linked to her self system. In this case, the affect-laden self would play a role, not by making necessary the use of beliefs without adequate justification, but by certain requirements on her developmental history that would be shared only by a limited few among her peers. As she succeeds in her assigned task, she would be one among a few among her peers who would be capable of similar success.

However, not all rules — not even the ones made use of in the sciences — are *universally* shared. The degree to which a rule is shared and accepted by others depends on how these rules come into existence. Mostly, rules of inference used in our daily life and our social interaction appear in the form of beliefs formed in the course of experience of individuals and social groups. For instance:

I know a number of family members of the principal of our local school, and have found them to be very nice people. I should not hesitate to concur with your decision to send your daughter to that school.

In this instance, the speaker seems to base her conclusion on a set of factors largely irrelevant to the question of acceptability of a school meant for children. What puts the stamp of approval on her inference is not a set of 'objectively' formulated criteria — we will dwell upon the question of objectivity in a later section (sec. 3.2.2.7) — but her feeling that the headmistress must be a 'nice' person. But even this conclusion based on seemingly irrelevant factors may, in the end, turn out to be correct because the personality of the headmistress *does* have a bearing in the matter of choosing a school for one's child. In other words, beliefs based on past experience do provide us with cues of *some* relevance in numerous situations arising in our daily life. We will enter into the question of beliefs being relevant in our scientific inquiries in chapter 4.

3.2.2.5 Inference through a sequence of intermediate stages and logical leaps

Inference is seldom a one-shot process. In most case, it proceeds through stages and in a hierarchical manner, with the conclusion obtained in one stage acting as an input in the next stage. What is more, inference often proceeds along several directions in parallel, and intermediate conclusions along the various branches are often compared so as to terminate some of those and to continue others, depending on the results of such comparisons. It is likely that the comparisons are effected by means of 'error signals' generated in interactions among neuronal assemblies. In addition, the mind makes guesses in advance as to what may prove likely to lead to the eventual solution to the problem at hand (the one in respect of which the inference is being made) and keeps on making comparisons of the intermediate inferences with these guessed 'postulates' generated in *logical leaps* — a logical leap being out of bounds of semantic reasoning. In other words, an inferential act is seldom a simple progression of a succession of inferential stages where the conclusion of one stage acts as the input for the next stage — such a simple progression will be referred to as a *linear* inferential sequence.

An actual inferential process is, generally speaking, composed of a parallel combination of branches, where each branch in itself is a linear sequence and where there continually occurs comparisons between intermediate 'outputs' of various branches.

> Often, all this complex mix of processes proceeds below our level of awareness.

Among the branches running in parallel are those originating in guessed postulates, somewhat like counterfactual reasoning ('*if it were the case that .....*'). In other words, an inferential process consists of simultaneous activity in 'forward' and 'reverse' reasoning, the latter being analogous to the process of *abduction*, i.e., reverse guessing where postulates or hypotheses are produced so as to explain evidence.

There is a fundamental difference between a linear inferential sequence and a logical leap (i.e., the initiation of a guessed postulate) that captures the essential contrast between deduction and induction — an actual real-life inference being, in general, a complex combination of the two types of processes. Imagining a linear sequence as a succession of 'nodes' (intermediate premises) connected by 'links' (operation of a rule of inference), the latter are essentially in the nature of *necessary* connections between the former (with reference to the individual making the inference). On the other hand the link between a premise and another obtained by a logical leap is not a necessary but a *contingent* one.

Before expounding on this last statement, we *summarize* by stating that an inference is made by starting from an initial set of observations, concepts and beliefs (and *heuristics* too — we look upon heuristics as beliefs of a specific type), when the mind is challenged with a 'problem' whose solution would constitute the end-point of the inferential act. The 'solution' arises in the form of one or more new beliefs that establish correlations between concepts over and above those that were there at the initiation of the inference. The inferential act consists of setting up a sequence of intermediate stages of inference where the passage from one intermediate stage to another occurs by way of *rules* being made use of, these rules being nothing but beliefs of various shades of generality and reliability. What is more, the intermediate stages of inference may be arrived at by a simple, linear succession, or by branching out into various 'directions' in parallel, and by the use of reverse reasoning where possible intermediate inferences are set up that may lead to the required end-point of inference. In all this, a number of intermediate inferences are arrived at as 'guessed postulates', by means of 'logical leaps'. This leads us to compare two fundamental processes of setting up inferential links in succession — one where the link in question is a *necessary* one, with little uncertainty associated with it, and the other of a *contingent* and *unpredictable* nature.

Imagine two physicians with similar background visiting in succession a patient afflicted with major problems, with each making a diagnosis. Only one of the two succeeds in making a correct diagnosis and prescribing medication that results in dramatic improvement in the patient's condition. Here the inferential process followed by either of the two professionals involves shared and self-linked resources, the shared resources being more or less the same for the two of them since they have had similar medical education and training. So, the difference between the two must have been caused by factors specific to the developmental history of each, where it is the unique developmental history that generates the self-linked psychological resources — preferences, beliefs, attitudes, and other psychological dispositions — of an individual.

In this present instance, the relevant self-linked resources reduce to the store of medical heuristics and the level of rigor in training, commitment, mental acuity developed by relentless practice, sensitivity to minute details in studying the patient, and a host of similar factors depending on attitude and specific features of training of the successful physician. The 'medical heuristics' in question relate to myriads of small items of knowledge and belief (for instance, recalled memory of case histories and anecdotes recounted by teachers and mentors) that a medical professional keeps in her mind for ready recall. It is possible that each item of knowledge in itself can be shared by other members of the profession, but the entire store of heuristics, along with the facility with which an individual can make use of those is specific to her. The size of the store of heuristics and the way these can be connected with particulars of a perceived situation is ultimately linked with affect and belongs to the self of an individual.

Thus, while the known medical facts brought to bear upon a problem (say, the results of pathological tests and their implications for a diagnosis) are shared resources and their application leads to predictable or *necessary* implications, the difference between one physician and another in the context of a highly insightful diagnosis lies in the stock of medical heuristics and in the level of commitment and attitude, possibly involving unconscious traits, the combination of which is specific to an individual and is *unpredictable* (one does not know beforehand whether some particular physician will be able to make a successful diagnosis) and *contingent* upon a host of factors that cannot be specified with completeness and in advance.

In other words, while *shared* beliefs and rules of inference lead to necessary implications, self-linked ones are made use of in setting up guessed postulates in an inferential act where the link between one intermediate inference and the next in a succession of inferential steps can be referred to as a 'logical leap' since it cannot be described in terms of a logical rule of some degree of generality.

Overview of the inferential process: Generally speaking, an inferential act involves a mix of such 'necessary' and 'contingent' links spread over 'forward' and 'reverse' 'branches' — all running in parallel — making up a tangled process where there occurs frequent comparisons of the intermediate inferences by means of 'error signals' and ceaseless monitoring by the affect system, acting as an 'internal censor'.

1. We have made liberal use of scare quotes in the above paragraph since one cannot expect the above description of an inferential process to be literally true — it is to be taken as indicating metaphorically a *possible* way in which an inferential process could progress, involving the use of a mix of necessary and contingent implications. The 'forward' branches referred to are ones where the successive inferential links flow from the initial premises of the inference towards the concluding premises — the

latter are not known beforehand, which is why the 'reverse' branches are initiated by way of hypothesis-building ('abduction'), where an assumed conclusion is adopted as an initial premise and an inferential chain is directed towards the actual initial premises of the entire process.

2. At every intermediate stage of the process, comparisons are made between the inferences currently available to check whether the progress is 'on the right track' where, once again, the affect system makes liberal use of beliefs acquired during past experience. As a simple but instructive metaphor, imagine a traveler setting out for a distant city who finds a bifurcation on his way, and takes the road on his left, basing his decision on a number of circumstantial cues. Subsequently he is forced to take a number of other decisions too regarding the choice of route, and he eventually chooses to be guided by the position of a number of stars in the heavens and comparing those with information acquired earlier regarding the general direction in which his destination lies. As he perceives a deviation from this direction his mind generates a warning signal for him to correct his course. This is akin to the role of the affect system as an 'internal' sensor or monitor referred to earlier.

Thus, beliefs and heuristics have a dual role to play in inference: on the one hand, these act as rules of implication (necessary or contingent) and, on the other, these are made use of in telling the affect system as to whether the flow of the inferential process is in the right direction so that a signal carrying a positive or negative valence (with, perhaps, a non-conscious arousal level) can act as a monitoring device — based on the signal, the flow can continue or can be made to change course so as to reach a desired conclusion. Of course, there is no sharp line of demarcation between the two modes of use of beliefs and heuristics, and one does not even know the exact way an inferential act proceeds, but what we are interested in here is to build up a possible logical outline that may help in understanding the inferential process in more concrete and precise terms.

3.2.2.6 Inference and decision: the dual-process theory

The *dual-process* theory (see, for instance, [47], [129]; an early exponent of the dual-process theory was William James) is in the nature of a conjecture regarding the way information processing is conducted in the human mind and is, in fact, a statement on how affect-linked ingredients possibly interact with reason in our cognitive activities. This is a view based on the assumption, consistent with a large body of literature, that human cognitive process involves two distinct streams — *process-one* and *process-two*. Of the two, the former is a fast, non-deliberative, massively parallel and autonomous process, while the latter is a slow and deliberative one, predominantly based on sequential processing that makes copious use of working memory. Significantly, process-two is perceived as being capable of acting as the control of the cognitive process as a whole.

> The question of control is a complex and non-trivial one, though. Commonly, the controlling action of cognitive processes is assumed to be exercised by the conscious mind. However, it is more likely that much of the controlling action is, in fact, executed without overt awareness [144], and is distributed over localized regions in the brain.

Process-one operates on the basis of self-linked resources of the mind and are, in a large measure, affect-based. Process-two, on the other hand makes use of shared mental resources (beliefs held in common, reasons of general validity, items of information known to be true in particular contexts).

What is of great significance is the manner in which these two process types *interact* in the course of an inference. This, of course, is a matter of speculation though, once again, one that may prove to be useful and suggestive in some considerable measure. It seems likely that both the processes run simultaneously, though process-two has periods of relative latency while, at the same time, there occurs a continual transfer of control to it when it evaluates, by means of comparatively secure rules of implication, some intermediate idea, concept, or belief produced in the course of activity of process-one, and passes on the ‘result’ of that evaluation to the latter which is then used as an ingredient in the further processing of information carried out autonomously of process-two. In this, process-two is likely to contribute its own quota of information processing (apart from evaluating and *justifying* what process-one has handed over to it), handing back a more 'finished output' to process-one.

> Medawar speaks of "a rapid reciprocation between an imaginative and a critical process, between imaginative conjecture and critical evaluation" ([94], p 44) in the context of the so-called hypothetico-deductive method of science, championed by Popper. The hypothetico-deductive method presupposes a distinction between phases of 'discovery' and 'justification' in scientific exploration. The two, however, can be distinguished only notionally since they are but two aspects of the same inferential process. The dual-process theory speaks of a more intimate mix of the 'imaginative' and the 'critical' processes where it need not even be necessary that there be two separate processes involved. A more likely scenario is that the inferential process is a single complex one involving 'imaginative' leaps and rule-driven 'critical' sequences. The latter picture of the inferential process is consistent with the one we attempt to put together in this book.

It is in the transfer of process control from process-one to process-two and back (the "restless to-and-fro motion of thought" spoken of by Medawar in [94], p 48) where one is likely to find the clue to the way the reasoning capacity of an individual evolves in the course of his or her developmental history, and to the remarkable difference in the reasoning capacities of individuals, including the difference in *modes* of reasoning of individuals and across cultures.

To be more specific, there are, in all likelihood, *two* things involved here, one dependent on the other. First, the quantity and quality of the vast repertoire of beliefs and heuristics that one can draw from in the course of an inference. And, secondly, the manner and the frequency of transfer of control from process-one to process-two as mentioned above or, speaking in more general terms, the manner of *interaction* between the two processes. When one stops to think of it, there can be an infinite range of variation in the mode and efficacy of the inferential process, depending on the interplay of these two factors. But one cannot say more since one has to draw a line here before indulging in dangerously unfounded speculation.

> *Emotion-driven* processes play a great role in leading to remarkable variations in the reasoning capacities among individuals, and to the progressive development of the reasoning capacity of a single individual. Success in an inferential act is a potent emotional factor that inspires enhanced performance in subsequent acts of inference of related kinds. Success elicits recognition and acknowledgment from peers, and public approval as well — all of these intoxicating indeed. No less intoxicating is the *internally generated affect* relating to success, and the resulting sense of confidence, of which a likely consequence is a heightened ability to explore possible alternatives in an inferential process — picking up cues in profusion from environmental inputs and making remarkable use of heuristics made available by the cognitive unconscious under the amplifying action of favorable emotions. In other words, emotions have an amplifying psychological role that may produce dramatic effects in the development of cognitive skills.

Numerous exponents of the dual-process theory converge on the opinion that process-one is of more ancient evolutionary origin than process-two, and there also appears to be a widespread acceptance of the view that the cognitive system, especially the one involved in process-one inference, comes equipped with a considerably developed *tool-kit* for adaptive thinking that has emerged in the course of biological evolution. While the 'tool-kit' is of an innate nature, it gets *expressed* [30] in the course of the individual developmental process — a process greatly influenced by the cultural environment of the individual. However, it is no less likely that the innate inferential capacity of evolutionary vintage is supplemented with capacities not originating in biological evolution but ones emerging in a secondary developmental process in the course of the inferential history of the individual, though this is a point where we desist from further speculation.

On the whole, the dual-process theory seems to be a pointer in the right direction, with which the viewpoint we attempt to develop in the present book is consistent.

3.2.2.7 How can self-linked resources produce correct inference

If inference involves self-linked psychological ingredients in intermediate stages, based on affective links being established between concepts, how is it that a correct conclusion can ever be reached in inference? Affect is essentially a subjective ingredient of the mind depending on the developmental history of an individual and often acts unconsciously.

On the other hand, it seems to be a general requirement that one needs to arrive at an 'objectively correct' conclusion in an inferential act. This is an issue of especial relevance in scientific reasoning where subjective factors play a considerable role in inference-making, hypothesis building and scientific creativity while, at the same time, one needs to comprehend and explain phenomena and events in an objectively existing world, i.e., one existing independently of our mental processes.

Even as the question appears to be a clear-cut one, it doesn't have an equally clear and precise answer. To start with, one needs to avoid philosophical cross-talk. For instance, the terms 'subjective' and 'objective' do not refer to two sharply distinct worlds. While it is true that reality exists independently of our mind and that, on the other hand, the mind is but one part of reality, it is *also* to be admitted that our phenomenal reality is a derivative of the 'real' or 'noumenal' reality, the latter being the truly independent and 'objective' one. This will be explained at greater length in chapter 4 (see section 4.3.1; refer back to section 1.12 for a brief introduction) and is more a point of view than a statement justified beyond doubt by evidence. What is of relevance here is that the proximate world of our perception is generated from the noumenal world in a process of *interpretation* by the mind. Which is how reality, the phenomenal world, and the mind interpenetrate one another.

Take, for instance, the person standing in front of me, gesticulating and trying to impress upon me the importance of adopting a certain course of action. As I watch him and listen to what he has to say, I keep on making a series of inferences as to the course that I should choose to adopt. How much of my inferences is objectively valid and how much is subjective? Or again, take the case of the scientist who has just come up with an idea that he *feels* is going to unify gravitation with strong interaction — he is ridiculed by most of his fellow scientists as being completely off his mind while he himself is convinced that his newfound idea is going to revolutionize physics.

These instances tell us that inference is an entangled mess made up of a mix of the subjective and the objective, and that the subjective is not always the fatal contamination on the objective that it appears to be. For, my inference of what the man in front is trying to communicate to me or the aged scientist's idea about the unification of forces *may just prove to be right*, or at least, may not be far off the track.

We will broadly divide our inferences into two classes — inferences relating to social reality, and those pertaining to the physical reality, such as those in the sciences, and will take up the case of the physical reality first.

> Of course, the division of inferential acts into two distinct classes is not to be taken literally because more often than not the two are mixed intimately — as in the case of, say, the administrative and scientific efforts at containing a severe pandemic.

In the case of inferences in scientific pursuit, the self-linked mental resources are mostly distinct from the beliefs of dubious validity that we often entertain in our head. What is of much greater relevance here is the set of heuristics lodged in the mind, most of which have been found to be effective in past inferential experience. Even though each single item in the store of heuristics may be accepted as true by a larger epistemic community, these are collectively as much an individual possession as, say the collection of dresses in the wardrobe of a person. Most of the heuristics are generated in the mind of a person in the course of his or her training and self-enrichment by processes specific to him and reflect the specific mode and manner in which he engages his mind in inferential activity.

More generally, the specific mode and manner in which an individual makes use of affect and reason (relating to process-one and process-two respectively) and the way the inferential process switches back and forth between the predominantly affect-based and the predominantly reason-oriented streams of thought, decides whether and to what extent the conclusion arrived at in an inferential act ultimately proves to be correct within a given context and conforms to evidence collected within that context.

> Of course, an inference can agree with evidence only within some given context, and may get modified drastically as the context changes. For instance: *based on the fact that my office assistant was absent from office on three consecutive days without notice, I concluded that he was guilty of gross negligence and dereliction of duty;*

> *however, it transpired later that his son was critically ill and had to be hospitalized on an emergency basis, when I had to change my opinion of him.*

Thus, it may well be that in generating a novel idea on the unification of the strong and gravitational forces, the veteran scientist is indeed on the right track in virtue of having stumbled across a hidden symmetry principle. That symmetry principle may turn out to be a perfectly valid mathematical one intelligible to many of his fellow scientists, but still it may well be true that the relevance of that principle was apparent only to him among his fellow-men in virtue of the way he keeps on combining and recombining some particular set of mathematical principles in his mind and feverishly tries to apply those to some specific set of physical situations.

> In other words, the self-linked resources of an individual that she may make use of in an inferential process (say, one in scientific research) may mostly be acceptable ones on grounds of objective validity in any given context. Even those of more dubious validity may possibly be of some relevance in the inferential process since these are always generated on the basis of past experience and generally have some correspondence with reality, even though tenuous.

In the case of inference-making in the social context, on the other hand, self-linked preferences and beliefs of more questionable worth play a much more important role. But here again, along with beliefs that have little to do with corroboration by evidence, there get involved in an inferential act, ones that have proved their worth at least in some measure in past experience, such beliefs being of overwhelming preponderance when judged by their sheer number. The standard of effectiveness of belief-based rules in social inference is much less stringent when compared with inference-making in respect of physical reality, since controlled experiments are often not on the cards in the social context — indeed, real life itself is the most effective laboratory in social cognition so that, when an inference fails to produce expected results, it is always possible to alter a number of beliefs and assumptions and go on to the next inferential act. Put differently, inference-making in the social context does not require standards of consistency and efficacy of beliefs as stringent as in the the case of inferences relating to the physical reality. In other words, beliefs of even partial efficacy may suffice in leading to 'correct' conclusions in the social context since standards of correctness are quite different here, being affect-ridden themselves.

> Even in the scientific context, emotion-laden beliefs may be of relevance in the generation of useful hypotheses (see section 4.5.10 in chapter 4), since inferences are always subjected to a stringent process of justification in terms of logical consistency and agreement with evidence, often obtained in controlled experiments. Affect is always prolific in the production of hypotheses, many (indeed, most) of which actually prove to be useless in subsequent reality check. Reason, on the other hand, is parsimonious in generating hypotheses in necessary profusion. When inference is robbed of affect, the fountain-head of hypotheses gets dried, making inferences timid and anemic.

3.2.2.8 Inference in creativity: the role of analogy

Inference is fundamentally involved in the process of scientific creativity. Put differently, creativity is essentially an act of inference on a grand scale, where a novel hypothesis is generated by the process of *abduction*, the latter being a variant of inductive inference.

As we will elaborate in greater details in chapter 4 (sec. 4.5), scientific creativity involves a process of 'exploration' of the conceptual space, where new correlations are established between *remote* conceptual domains. This, fundamentally, is the process of inference where correlations are successively established between concepts, though the concepts so correlated need not ordinarily belong to remote domains in the conceptual space. In other words, ordinary inferences establish *local* correlations while acts of creativity involve a more global *restructuring* of the conceptual space.

We will see in chapter 4 that, in the context of creativity, *analogies* are of great use in establishing links between remote domains in the conceptual space. Equally important is the role of analogies in establishing local correlations as well. In other words, analogies are relevant in the restructuring of the conceptual space *at all scales*.

The ability to invoke analogies is a deeply personal one, since analogies are closely associated with emotions. Indeed, emotions are generated as analogies between aspects of experience as the mind attempts to classify the infinitely varied and chaotic stream of events that we encounter in life. In this, the analogies possess strong implicit components that we make explicit by means of semantic relations between concepts in a language-based description. However, while an analogy can be understood in linguistic terms, it is generated entirely in implicit mental processes.

1. Analogies may even be generated by means of privately held beliefs (such as, say, mythological ones). The more constraint-free the mind is, the more prolific is the source of analogies (in this context see section 4.5.5).
2. The deeply personal nature of the capacity to make analogies is seen in the phenomenon of *deja vu*, where one gets a feeling of 're-visiting' a situation or an experience. Such a feeling possibly involves an analogy being unconsciously set up in the mind between a current experience and a an earlier one, whose details have got lost.

Analogies are often involved in the making of 'logical leaps' mentioned earlier in sec. 3.2.2.5. These leaps are made by affect-based implicit processes of exploration of remote domains of the conceptual space and, at the same time, occur within some particular context set by the external world as well, depending on the problem at hand that one needs to address.

Thus, a logical leap needs affect-based processes occurring within an objectively specified context that reason can comprehend so as to access shared rules valid within some specific conceptual domain. The meeting place between shared knowledge of the world and privately held affect-laden views (the so-called 'egocentric' and 'allocentric' perspectives held by the mind or, to put differently, between affect and reason) is provided by the *default mode* of the mind (see, for instance, [108]) that generates fruitful analogies for the logical leaps to be made possible. As recent investigations show, the default mode of the brain is incessantly active as the springboard of a number of major activities of the mind.

In summary, analogies are instrumental in the making of logical leaps in the process of exploration of the conceptual space in inferential acts. Such exploration, when conducted in large regions of the conceptual space covering sparsely connected domains tends to result in scientific creativity (sec. 4.5), which can be described as inference on a grand scale. Both inference in general and creativity in particular fundamentally require the activity of the default mode network (DMN) of the brain.

The exploration of a certain domain of the conceptual space in an inferential act has been referred to as movement in a 'problem space' ([89], chapter 3), analogous to a walk in a labyrinth.

3.3 Reason, affect, and social interactions

We interact with the world (and with ourselves too) by means of *affect* and *reason*. But affect and reason are not ingrained in us, only the *capacities* for these are. Our capacity for responding to the world by affect and reason have been handed down to us by a protracted process of evolution. The capacities are *realized* in our *social interactions* (in this context, see [135]).

Social interactions are initiated right after birth through the persons of the mother (or other principal caregivers) and the close ones attending a newborn. To begin with, the distinction between self and non-self is vague and fuzzy, and almost all impressions in the baby's mind are assembled in the form of its self-system. But the differentiation between the self and the non-self also begins right away and all the early experience of the child get funneled though the society of people around it. Every experience carries dual significance for a child as it is impressed in its mind, and this continues in later life too. On the one hand, it generates a valence-based valuation produced by its affect system, a valuation that gets colored by emotions generated in a complex interaction between impressions of earlier experiences and fresh inputs generated in current social interactions (*a happy news received at a time of bereavement in the family*) and, on the other, it generates a sense of interconnection between concepts from which reason is formed as *necessary* relations of implication.

Both affect and reason are generated in processes involving distributed neuronal assemblies in the brain where these processes occur as parts of a complex and *integrated* dynamics, and where the integration is effected in the form of interactions between diverse brain regions, among which the frontal cortex plays a major role. The operational distinction between affect and reason is that the former is associated with the unique developmental history of an individual, while the latter is shared with larger social groups to which the individual may belong. As the individual engages in interactions with other persons in diverse social groups, some small and some large, with one group included within some other, her mind gets endowed with more and more complex interrelations among concepts (generated in a process of objectification of percepts) many of which appear as *necessary* relations of implication. This process gives rise to reasons of ever greater generality in the mind of the individual, forming a stock of *shared* mental ingredients.

In other words, socially generated experience produces, on the one hand, an accumulation of preferences, personal beliefs, and dispositions that lead to the growth and evolution of the self-system of an individual, with affect constituting its axis and, on the other, an expansion and enrichment of the conceptual network, with new links between concepts, along with the formation of an expanded set of shared beliefs and shared rules that themselves appear as objectified entities in the conceptual network. The two sets of mental entities, namely, the self-linked and shared ones — the former affect-laden and the latter reason-based, are inextricably linked so as to be involved together in all our mental processes, endowing the mind with capacities of operation in two distinct *styles*.

> The conceptual network briefly introduced in section 1.7, 2.3.1 (see section 4.3.12 for the analogous idea of the conceptual space) refers to a network of concepts lodged in our mind, where the latter are linked with one another by correlations of multifarious types. The network is infinitely extended since the number of concepts that can be linked with one another is unlimited. However, in reality, the conceptual network is accessed by the mind of an individual only in finite chunks — in this sense the extent of the conceptual network is only notionally or potentially infinite. As an individual accumulates experience in stages during her journey in life, more and more concepts get linked, thereby progressively widening the network. The extension takes place along a multitude of dimensions, some of which pertain to social relations, some to the biological world, and some others to the physical reality, with a great deal of overlap between these. A great majority of our concepts are induced in us by means of social interactions. In a sense, we look at reality through the eyes of

our fellowmen.

Chapter 4

Complexity reality and scientific realism

We recall the statement, made in the opening section of this book, to the effect that the mind and Nature are two incredibly complex systems confronting each other, and begin the present chapter with an account of *complexity*. The mind is incessantly engaged in its efforts at comprehending reality. But the pervasive complexity inherent in Nature and in the mind stamps the process with a unique unevenness. In this chapter we will see what to make of this process.

4.1 Complexity and reality

Complexity is all around us all the time but is something that has come under the lens of focused scientific inquiry in relatively recent times, in contrast to systematic and sustained investigations on *idealized* or *simplified* systems. Simplification is the name of the game, and quite understandably so.

When faced with complex systems that defy our understanding and baffle us, we seek out relatively simple parts of those, whose behavior can be probed and explained with material and intellectual means within our reach. Even this requires stupendous and prodigious efforts of intellect that mankind has not been found to be lacking in. However, there is no hard and fast demarcation line between 'simple' and 'complex' systems, and the human mind has, throughout history, taken up challenges from both the simple and the complex — the former in more precise and rigorous terms and the latter more qualitatively and phenomenologically.

To start with, a simple system is one imagined to be isolated from the rest of the world so that the mechanism underlying its behavior can be specified unambiguously. The influence of the rest of the world is then postulated, again in simplified and unambiguous terms, which means that only certain special classes of influence can be included in the theory. The resulting behavior is then determined, but this time only in approximate terms, since an exact determination is often beyond reach. This is essentially how a *model* of some part of the world is built, where simplification and idealization rule the show.

But the world around us is fundamentally complex, and every time mankind has triumphed in formulating exquisitely devised theories about systems, complexity has raised its head, making necessary a fresh look at things so as to account for incongruities arising from phenomena that were left unexplored at the earlier stages. This goes on and on .... like a recurrent process that never approaches conclusion.

Of course, one can be philosophical and say — in tune with a number of great minds in history — that the task of science is to seek out simplicity out of apparent complexity, and that the world is fundamentally simple, ruled by an all-pervading harmony. The position we are going to adopt in the present book is at odds with the last part of the above proposition. On the other hand, it is indeed true that much of our scientific endeavor consists of an effort at locating relatively simple features in complex systems, in terms of which one gains leverage in understanding and explaining the latter. The simple features provide us with foothold for surveying the complex landscape that lies around and for planning for the next phase of journey, looking for more of such footholds. Simplification and idealization is a necessary strategy in this complex world of ours, made up of complex parts. Indeed, the world uncovers its complexity to us in stages, all the while deceiving us into thinking that the ultimate secret — simple and beautiful — lies round the corner. This is something that will demand our attention in a later section (sec. 4.3.8).

The *theory of complexity* has had a long history but is still an emerging one. We will not go into the history here, nor shall we go through the various aspects of that theory in technical terms. Our job in the next section (sec. 4.2) is to briefly look at a number of major features considered to be common to numerous complex systems of interest. These will then be made use of as a backdrop in addressing issues regarding *reality* and *scientific realism* in section 4.3. This will include a brief look at the *complexity of the human mind* (sec. 4.3.11). We sum up these considerations in sec. 4.4. The chapter will conclude by way of briefly addressing the issue of *creativity* (we will mostly be concerned with *scientific* creativity) — how the mind opens up new horizons in its ongoing query on the way Nature works.

4.2 Complexity: a brief outline

4.2.1 Complex systems: decomposability

Generally speaking, a complex system is made of a large number of components, or *subsystems*, where the subsystems interact with one another, and where *each subsystem is often a complex system* in its own right. Thus, one finds nested *levels* of complexity forming a hierarchy, where each level corresponds to a complex system, with levels below and above it (the terms 'below' and 'above' are for convenience of reference only) making up the hierarchy and where, generally speaking, a system at any specified level constitutes a subsystem of the one above.

The systems and subsystems need not be ones located in our familiar three dimensional space. For instance, every individual possesses a vast set of *beliefs* that interact with and influence one another while a belief itself is made up of a relatively large number of *concepts*. A concept in turn involves a number of other concepts, some of which may relate to *objects*, while the concept of an object is generally made up of a number of *attributes*. All these may be located in some abstract *conceptual space* (refer to sections 1.10, 2.4), whose relation with our familiar *physical* space is non-trivial. An alternative way of looking at the vastly complex interrelations between concepts is in terms of a conceptual *network* (refer to sec. 1.10; the conceptual network has been referred to repeatedly in chapter 2; however, we talk of the conceptual network and the conceptual space only in intuitive and qualitative terms).

> We recall that the beliefs lodged in our mind can be objectified, making them appear as concepts themselves, i.e., as nodes in the conceptual network. This makes the latter (and the conceptual space) acquire an exquisitely complex and convoluted character.

The interactions among the subsystems of a complex system may be rich and varied. For instance, in a big business organization, there may be a hierarchy of management levels, each level being made up of a number of departments. The relations between the various departments belonging to any given level may differ widely, depending on their type. Thus, the exchanges between an accounts department and a purchase department are way different as compared with those between the former and a personnel department. It is not uncommon to encounter a complex system whose subsystems have a *wide spectrum* of interactions.

However, despite these rich and varied interactions, it often makes sense to refer to the subsystems individually and severally, i.e., in other words, the subsystems retain a measure of identity. This is expressed by saying that a complex system is *decomposable* ([122]). On the other hand, this decomposability is only approximate and need not mean that the interactions among the subsystems are not of much consequence. Indeed, it is precisely these interactions that gives a complex system its very own identity, including the way it evolves in time, passes through phases of stability and instability, and generates new levels of organization. Looking at a growing fetus, its course of development into an adult depends on the interactions among its rudimentary bodily organs and the myriads of cells making up its body, along with the genetic and epigenetic interactions at the molecular level within the cells — in addition, its interactions with surrounding systems, such as sources of nourishment, are also of great relevance. All these multiple interactions make the fetus go though an exquisite succession of developmental regimes characterized by more and more novel forms of organization.

> A subsystem within a complex system is often referred to as an independent entity, but one with certain *properties* — it is the set of properties that is indicative of the fact that the subsystem in question is not really independent, but interacts with the rest of the subsystems. The interactions between a subsystem with other subsystems determine how the subsystem in question behaves in the company of of its peers. Strictly speaking, the repertoire of possible behaviors is infinite, and analogously, the interactions in a complex system cannot be encoded in simple terms either. Indeed, the correspondence between the interactions and the behavior patterns is anything but simple.

In respect of the *theory* describing a complex system (say, ‘S’), the particular level at which its subsystems are located (with levels above and below it often forming a nested hierarchy) is pertinent to that theory. Such a theory depends crucially on the correct formulation of the interactions between the subsystems in the level located below the one in which ‘S’ is located (let us say that ‘S’ is located in level ‘L’), while other systems located in level ‘L’ and the ones above constitute the *context* of that theory. For instance, the properties of a solid are determined by the interactions among the atoms or molecules making it up, while these interactions, in turn, are determined by the disposition of the electrons and the nuclei within the atoms. The various macroscopic systems with which the solid exchanges energy and matter, such as large heat reservoirs (the atmosphere, for instance) set the context in which the properties of the solid (such as its thermal expansion and contraction) are expressed. Once again, this distinction between the essential ingredients and the context in respect of a theory is, to some extent, arbitrary but is useful nonetheless.

In summary, a complex system is generally made up of a large number of subsystems, with the latter interacting with one another in intricate ways. The system, in turn, interacts with other complex systems forming its environment, while all these complex systems (the system under consideration and those forming its environment) act as subsystems of a complex system at a still higher level. Depending on the context, the hierarchy of systems may be assumed to be terminated at some specified stage, with the levels higher up or lower down not being relevant in respect of the behavior of the system under consideration in the given context.

There is an important qualification here: decomposability has a wide variation in gradation across systems and their subsystems. Some complex systems — ones whose constituents interact, in some sense, relatively weakly among themselves — are characterized by a considerable degree of decomposability. On the other hand, ones in which the interactions are relatively strong, are to be described as decomposable only notionally. In reality, most complex systems are characterized by a wide spectrum of interactions among their subsystems and one finds a gradation of decomposability across various different chunks of it.

As we will see, *reality* at large is a vastly complex system — it includes *all* the complexity there is. At the same time, the human mind — a part of reality — is also enormously complex in its own right. Interestingly, large parts of reality — more specifically, many of the systems studied in the physical sciences — belong to the category of systems with a relatively high degree of decomposability and their behavior can be described with reference to suitably idealized 'simple' systems. The *mind*, on the other hand, is a system that is much less decomposable. This makes it much more difficult to build *theories* that describe and explain the human mind and the behavior pattern of human beings that the mind generates. Such theories are useful only within very limited contexts and need radical revisions as the context changes.

4.2.2 Behavior patterns of complex systems: CPS and CAS

The interactions among the subsystems of a complex system and those with various other systems (ones constituting the level higher up in the hierarchy) are generally of the *nonlinear* type, for which the rule *the whole is different from the sum of the parts* applies. This is not a very precise statement but appropriately sums up a number of features observed for actual systems.

For instance, consider a system made up of three subsystems, say, ‘A’, ‘B’, and ‘C’. If the behavior of the combination of ‘A’ and ‘B’ is known, along with the behavior of each of the combinations ‘A’,‘C’ and ‘B’,‘C’ (at times, the three combinations behave in analogous manners), then one cannot infer the behavior of the combination of all three taken together from the properties of the pairwise combinations and from those of the individual systems under consideration. In other words, the presence of additional systems makes a notable difference. Suppose ‘A, ‘B, and ‘C’ are three persons of known temperament and mental disposition, and also suppose that the behavior of ‘A’ in the presence of each of ‘B’ and ‘C’ is known. This may prove to be utterly inadequate in explaining the behavior of ‘A’ in the presence of ‘B and ‘C’ taken together (‘A’ may exhibit friendly or neutral behavior toward ‘B’ but may show loving considerations toward ‘C’; on the other hand, ‘A’ may be found to be seething with suppressed emotions in the presence of both ‘B’ and ‘C’, and may even exhibit some degree of belligerence towards ‘B’ because of ‘C’ apparently ignoring the presence of ‘A’).

> In the case of three particles, this may sound like ‘three-body interactions’ dominating over ‘two-body’ ones. However, even in the absence of three-body interactions, the behavior of a system of three particles may be quite intractable, looked at in terms of the interactions considered pairwise.

Described in general terms, the behavior of a complex system made up of numerous subsystems turns out to be *non-trivial in a major way*. In this context, one distinguishes between *complex physical systems* (CPS) and *complex adaptive systems* (CAS), as highlighted in [66].

A CPS is made up of elements or subsystems that have fixed properties — the molecules of a gas, the spins in a magnetic lattice, or the parts of an automobile. A subsystem in this case can be in any one of a fixed set of states, where a state can change under the interaction with other subsystems belonging to the CPS — often the ones that, in some sense, are ‘close’ to the subsystem under consideration. For instance, the position and momentum of any particular molecule in a gas get modified by interaction with other molecules in its close vicinity, while the effects of distant ones are usually small.

A note on 'remote causes'.

Considering any specified molecule in a gas, distant molecules exerting a negligible effect on it constitute instances of what may be referred to as *remote causes*. The hall-mark of a complex system is that, such remote causes become relevant beyond some characteristic time scale, i.e., *remote causes cannot be ignored for long.* For instance, molecules lying at a large distance from some specified molecule in a gas will ultimately come close to it and influence its motion, i.e., remote causes are relevant in the behavior of the gas as a whole. *Complexity*, in other words, is, to a large extent, generated by the operation of remote causes that assume relevance in the long run.

In contrast to CPS, the properties of elements making up a CAS get changed in the presence of other elements and of other systems interacting with these. For instance, the ability of a gene to express itself as a sequence of amino acids may change under the influence of reactants around it. The elements of such a system — commonly referred to as *agents* — 'learn' or 'adapt' themselves as they interact with other agents.

The ability of the elements of a CAS to adapt themselves leads to quite amazing behavior exhibited by such systems — often in the nature of *goal-directed* processes, such as the self-replication of genes, or the making of *decisions* by the human mind. To be sure, a CPS may also behave in a 'purposeful' manner, such as a cellular automaton devised in early days by von Neumann that could be made to replicate itself, and a vast number of cellular automata designed subsequently. The difference between such CPS with strange behavior and CAS with adaptive elements often lies in the way these systems are generated — while the purposiveness of a CPS is given to it by some kind of human intervention (a 'programming'), a CAS usually evolves in virtue of its own dynamical characteristics where, at some level deep down the hierarchy, CPS elements (complex molecules, for instance) may be found to play a crucial role. In other words, the learning or adaptive abilities of a CAS may appear as *emergent properties* of assemblies of CPS (example: biological evolution emerging from pre-biotic evolution). The important issue of emergent properties in complex systems — considered to be problematic one by numerous scientists and philosophers — will be taken up in sec. 4.2.8 later in this chapter.

> In the present book, we will not refer in any major way to the fascinating behavior patterns of complex adaptive systems (CAS), and will mostly confine ourselves to examples and illustrations relating to complex physical systems (CPS). In particular, our considerations in sec. 4.3 below will mostly focus on CPS in order to explain the nature of scientific theories in relation to our experienced physical reality.

It is difficult to exhaustively categorize — item by item — the extremely rich and diverse behavior patterns of complex systems. Even the more notable ones like the appearance of emergent properties become somewhat elusive when one attempts to pin these down to precise formulation. This does not mean that the various behavior patterns themselves are figments of imagination — the very complexity of the systems prevents an unambiguous and universally valid characterization of these behavior patterns. We will first take up the case of apparently simple systems whose time evolution is described by means of *differential equations*.

The rich and intricate behavior patterns of a complex system often appear in the form of impenetrable perplexities in the cause-effect relationship that it exhibits. A ‘small’ or insignificant ‘cause’ often leads to quite dramatic ‘effect’ as observed in bifurcations and the ‘butterfly effect’ (sensitive dependence on initial conditions) in nonlinear systems. Likewise, as an instance relating to a CAS, a ‘small’ change in environmental conditions leads to the eventual emergence of a new species in biological evolution. Commonly, a small or ‘negligible’ cause is found to lead to notable effects because of the role of factors hidden in the depths of complexity of the system under consideration or of context effects (erroneously) assumed to be of no consequence. Thus, a few grains of sand added to a sand-pile may cause the latter to collapse because of the fact that it was close to *criticality* to start with. Analogous intricacies and puzzles are met with in respect of *emergent properties* of complex systems, as outlined in sec. 4.2.8.

4.2.3 Complex systems: nonlinear dynamics

Numerous complex physical systems (CPS; recall that the subsystems making up a CPS are not adaptive in nature) are described in terms of *differential equations* where these equations are, generally speaking, of the *nonlinear* variety (linear systems are, in a sense, exceptional though these are familiar, well-studied, and useful too).

The differential equations may describe the time evolution of a finite number (say, N) of variables making up a N-dimensional *phase space*, where these belong to the class of *ordinary* differential equations. More generally, a system may be described by a number of *fields*, such as the velocity field of a fluid, in which case the behavior of the system is described by a set of *partial differential equations*, representing the time evolution in an *infinite* dimensional phase space. Once again, the partial differential equations are generally speaking, nonlinear ones, though linear partial differential equations, on which a vast literature (in physics and mathematics) exists, are also of great relevance. The *Navier-Stokes* equations describing the flow of a fluid constitute a well-known example of a set of nonlinear partial differential equations in physics.

A set of ordinary or partial differential equations in a (finite- or infinite-dimensional) phase space is said to describe a *flow* in that space. Alternatively, one can consider a *mapping* in the phase space describing the evolution in *discrete* time-steps. In the context of the present book we will refer mostly to nonlinear ordinary differential equations in phase spaces of relatively low dimensions ($N = 2, 3, \cdots$).

Let $x_1, x_2, \cdots, x_N$ be the variables describing the *state* of a system at any given instant of time. Such a state is represented by a point in its (N-dimensional) phase space. With the passage of time, the state *evolves* in accordance with the set of differential equations under consideration, describing a *trajectory* in the phase space. While such a trajectory is a continuous one, the time-evolution of a system described by a mapping is represented by a discrete succession of points, representing the succession of states at discretely spaced time instants.

While nonlinear differential equations (mappings will also be occasionally referred to) will be seen below to generate complex behavior, linear systems are, typically, *simple* ones obtained as limiting cases from nonlinear systems. The term 'simple', however, need not imply that the behavior of such systems can be described trivially and without effort — these being of enormous relevance in *models* studied in all branches of natural science. An area in which a set of linear partial differential equations describe a real-life system (and not simply an idealized model) is *electromagnetic theory* where the *Maxwell equations* describe the space-time dependence of an electromagnetic field. Great mathematical difficulties are encountered (and often dealt with by invoking *approximation schemes*) in virtue of *boundary conditions* of various types, relevant to specific problems. The boundary conditions may, in a sense, be looked upon as the effect of a set of 'remote causes'.

In contrast to systems represented by nonlinear differential equations that serve as mathematical models of complexity, ones represented by linear differential equations can be described as 'simple', though not in the sense of 'trivial' or 'easy to analyze'.

Nonlinear equations do not conform to the *principle of superposition*, and serve as illustrations of the rule expressed qualitatively as 'the whole is different from the sum of parts'. No general principles exist for the construction of solutions of nonlinear differential equations, and the infinite diversity and variety in the time evolution of systems described by these equations remains largely unexplored. Nonetheless, deep insights have been developed regarding various *types* of behavior that these systems follow. The *qualitative theory* of nonlinear systems was developed by Poincare and other great mathematicians in the first quarter of the last century. Their investigations were carried forward in large strides by others during the second half of the century, resulting in a highly developed theory that is far beyond the scope of the present book.

In describing the various types of behavior of a system represented by a set of nonlinear differential equations, one generally looks at the *large time* regime, i.e., the one in which the *transient* behavior, if any, is not of relevance, when the system exhibits a behavior pattern that is termed 'asymptotic'. Speaking schematically (i.e., not entering into a precise classification, which is fraught with difficulties anyway), this long-term or asymptotic pattern may correspond to a time-invariant state, an oscillatory state, a quasi-periodic state, or to *chaotic* behavior.

> A quasi-periodic state is a generalization of a periodically varying one, where the time-dependence of the relevant state variables involves several frequencies, incommensurate with one another.

There exist several quantitative indicators of chaotic time evolution. In other words, there may be numerous different *types* of chaos. The indicators of chaos are mostly based on various *entropy* measures (see, for instance, [133]; see also [51]). It seems likely that the generic behavior of nonlinear systems involves chaotic time evolution.

In a chaotic time evolution, either the whole of the phase space or some part of it is explored (by the point representing the state of the system) in a random manner. In contrast, time-invariant, periodic, and quasi-periodic behavior patterns are referred to as *regular* ones.

In the thermodynamic description of systems, one distinguishes between *microscopic* and *macroscopic* states and their time evolution. In the macroscopic description one often encounters an equilibrium state which is time-invariant, while the same system, when described in microscopic terms, may involve chaotic dynamics in the relevant part of the phase space (see, for instance, [44]). Likewise, the over-all statistical behavior of a complex system is often described in terms of relatively simple *distribution functions*, though even a distribution function may undergo notable changes on longer time scales.

The description of a nonlinear dynamical system may involve a number of characteristic *parameters*. For instance, in the case of a fluid flowing through a pipe, the nature of the flow may depend on the *coefficient of viscosity* and density of the fluid, the diameter of the pipe, and the pressure difference between the ends of the pipe driving the flow.

For a given set of values of the parameters, different parts of the phase space may correspond to various different behavior patterns. Thus, depending on the *initial condition*, the system behavior may turn out to be either time-invariant or periodic (or even quasi-periodic) in defferent regions of the phase space. Additionally, what is remarkable in the case of a nonlinear system is that, for various different sets of values of the characteristic parameters, the system may exhibit *qualitative* changes in the behavior pattern. Such a change is generally referred to as a *bifurcation*, though here again there remains a big gap in precision since it is not easy to classify various possible types of bifurcation (there are, for instance, *local* and *global* bifurcations or, again, bifurcations in conservative and dissipative systems, bifurcations with various possible *codimensions*, and so on).

Even a system with a phase space of a low dimension (as low as one in the case of mappings and three in the case of flows) can have a complex behavior pattern as it evolves in time. In other words, the complexity of such a system seemingly resides in its time course of evolution, while this complexity in time evolution may reflect a complexity in the underlying structure of the system itself.

1. In an early and influential paper by Robert May ([91]), one encounters complexity in apparently very simple systems (idealized biological populations) evolving in discrete time (successive generations, assumed to be non-overlapping) through a succession of bifurcations. The parameter whose value controls

the bifurcations in this system was related to the rate of production of offspring from one generation to the next. Evidently, this parameter is determined by a large number of factors relating to the life-cycle and reproduction of the species under consideration, the details of which is ignored in the simple set of nonlinear equations describing the population.

2. Seemingly simple systems exhibiting chaotic time evolution may result from a process of *reduction* from more complex ones, where numerous relevant variables pertaining to the underlying complexity are not explicitly taken into consideration, with the result that the reduced system evolves in a complex manner.

The set of characteristic parameters controlling the bifurcations of a nonlinear system can vary under the influence of other systems that may come in interaction with it. In this context, recall the feature of (approximate) *decomposability* of a system whose behavior pattern is determined by subsystems at a lower level belonging to a hierarchy of levels, while the same behavior is also conditioned by interactions of systems at the same level (say a number of biological populations interacting with one another) and by levels higher up (section 4.2.1). Because of the conditioning effects of the higher levels, the parameters characterizing the dynamics of a system do not remain ideally constant, and keep on changing relatively slowly in time, causing an unfolding of bifurcation scenarios whereby the behavior pattern of a system keeps on undergoing qualitative changes.

In a number of situations involving real-life systems, these qualitative changes in the time course of evolution appear in the form of *emergent properties* ([66]) and *self-organized complexity* ([138], [69]). It is to be mentioned, however, that the appearance of these novelties is not entirely the consequence of intrinsic properties of a system, since these critically require an appropriate *context* in the form of an influence of external systems — ones in the same level of the hierarchy of complexity pertaining to the system in question or in levels lying above. For instance, referring to the microscopic dynamics of a macroscopic system, the emergent phenomenon of a phase transition requires an appropriate condition, say, one involving the temperature in the case of a magnetic lattice where the dynamics decomposes into two *ergodic* components.

The term 'self-organized complexity' means the proliferation of structures at various scales in the course of time evolution of a complex system. At times, the emergence of such structures occurs under some kind of *driving* by external systems, but then one can think of an augmented system including the external systems in question, in which the said structures emerge spontaneously, i.e., in virtue of its own dynamics (hence the qualifying phrase 'self-organized'). A commonly occurring scenario in which self-organized complexity is found to appear is that of *self-organized criticality* (sec. 4.2.6) where there occurs a loss of stability under a relatively *slow* driving and the rapid transition to a stable configuration. However, such difference in time scales is a commonly occurring feature in the time evolution of complex systems in general, and nonlinear systems in particular.

The appearance of novelty in a system described by a set of nonlinear differential equations occurs as a consequence of *de-stabilization* of an existing stable behavior-pattern in the phase space and the attendant emergence of a new stable configuration or pattern. Often, one finds a succession of such transitions in an unfolding chain of ordering phenomena. For instance, in the case of fluid flow under dissipation (i.e., heat conduction and viscous effects), say, in the case of heating of a layer of fluid from below (against the pull of gravity), one observes a transition from steady conduction to non-steady motion involving convective rolls of a succession of various geometric forms and then, eventually, to *turbulent motion*. The characteristic parameter controlling the transition from one type of flow to the next in this case is termed the *Rayleigh number*, and the appearance of a succession of variously shaped convection cells associated with increasing values of this parameter constitutes an instance of self-organized complexity.

4.2.4 Complexities of time evolution

The successive regimes of *instability* and *stability* commonly associated with nonlinear differential equations are found to be present in complex systems of more general descriptions (e.g., in complex adaptive systems) where a precise mathematical description in terms of differential equations may not hold. In electronic control systems these regimes of instability and stability are associated with *positive* and *negative feedback* between various different parts of a circuit. Generally speaking, instability and stability (of *local* and *global* varieties) are consequences of the large number of subsystems making up a complex system (at some specified level in a hierarchy of complexity) and the spectrum of interactions between these subsystems.

It may be mentioned here that the complex and chaotic time evolution of a system may, under certain circumstances, be related to *computational* or algorithmic complexity [66] encountered in computation theory. For instance, for a system with positive *Kolmogorov-Sinai entropy* [44], if one tries to describe computationally a long time series characterizing a trajectory of the system or to specify as accurately as possible the initial condition in the phase space giving rise to the trajectory then the length of that description diverges along with the time-interval of evolution.

Thus, a complex system, in virtue of being composed of a large number of subsystems with a wide spectrum of interactions between those, and of being a part of a hierarchy made up of various levels exhibits, in general, a complex time evolution, the latter being characterized by several time scales. While we have spoken of the asymptotic regime of time evolution in the case of nonlinear differential equations, one may observe asymptotic behavior in relatively shorter and longer time scales as well, making up a *hierarchy of time scales* along with the hierarchy of levels of complexity mentioned above.

In other words, complexity is manifest *across numerous scales* — both in time and in the phase space, where the latter may be of an arbitrarily large dimension.

Over a limited time horizon, a complex system may exhibit a certain pattern and structure — either in a high dimensional phase space or even in the familiar three dimensional space, such as the patterns of oceanic and air currents in a geographical region. Over a longer time scale, these patterns may give way to currents having different characteristic features. Along with the temporally changing patterns, there are commonly found distinct patterns in various different regions of space (once again, either in the phase space or in spaces of lower dimensions, including the familiar three dimensional space). Such patterns and structures are indicative of the emergence of *order* in space and time.

The intricate and inscrutable nature of the time evolution of a complex system commonly results from *co-evolution* [133] — i.e., evolution not only due to fixed interactions between the constituent subsystems (along with the ubiquitous context effects), but due to changes occurring in the nature of these very interactions, and due to changes in the subsystems themselves. Commonly, this owes its origin to the fact that one cannot separate clearly the intrinsic dynamics from the context — small changes in the environment may trigger an instability. In other words, *everything evolves* and can affect the dynamics in major ways — by way of altering the subsystems, the interactions, and even the context — and the entire evolution becomes an enormous tangle of nested correlations relating to causes and effects. Faced with this tangle, one can at best untie a few knots here and a few knots there — *locally* in space and time, and learn the 'laws' governing the evolution only locally as well. This will engage our attention later, in sec. 4.3.

> Typically, a mathematical formulation of the evolution of a complex system remains a tall order — one describes the evolution at best in the form of an algorithm, and that too as only a partial description. Even when one describes the dynamics in the form of a set of nonlinear differential equations (once again, by way of reduction to a model), the solution to the equations can only be obtained in qualitative terms — for instance, as information pertaining to the structure of certain invariant sets and their stability characteristics, and to their dependence on relevant parameters, while bifurcation scenarios are also obtained numerically. An algorithmic description [133], on the other hand, is usually a flexible one, yielding a lot of relevant information about the system — depending on what one wants to know.

In this context, it is important to distinguish between microstates of a complex system and its *macrostates*, the latter being in the nature of statistical averages (along with low order fluctuations) of microscopic variables relating to its detailed phase space description. For instance, major breakthroughs in the understanding of complex systems have come about in the form of *scaling* and *power law behavior* relating to macroscopic data, and their dependence on parameters (or data) specifying the *structural features* of the system. A convenient way of representing the latter is in the form of *networks*. All this will be briefly touched upon in the following.

4.2.5 Complex systems as networks

It is often convenient to represent a complex system in the form of a *network*. In this representation, the subsystems making up the system under consideration are said to form a set of *nodes*, to be visualized as dots or circles strewn around in space (we will imagine the nodes to be located in the familiar three dimensional space, though the dimension of the space may not be of much relevance; for instance, the nodes of any finite network may be imagined to lie on a single line). In the case of a CAS, the nodes are often referred to as *agents*. The interactions or correlations among the subsystems are represented in the form of *links* connecting the nodes . If the correlation between a pair of nodes has a sense of asymmetry associated with it then the link is visualized as being a directed one (example: *husband-wife relationship* in a community); if, on the other hand, the correlation is symmetric then it can be represented by a single link without any direction — or by a pair of oppositely directed links (example: *relation of friendship*).

The network representing a complex system is, generally speaking, a *multi-layered* one [133] since its nodes may be connected by various different *types of links*. For instance, in a human society, nodes may be correlated in terms of religion, political commitment, family relations, occupation, and so on. These multiple layers constitute an important characteristic determining the nested and tangled structures that appear within a complex system. The multi-layered structure of complex networks will be found to be relevant in interpreting the feature of incommensurability in successively revised theories in science (see sec. 4.4).

The time evolution of a complex system is described by specifying whether and how new nodes are created and old nodes get removed from the system, together with the way new links are established and existing links get removed *and*, additionally, how the layered structure of the network gets modified. All this goes to describe the co-evolution of the system under consideration. Such a description of the evolution of a network may be a deterministic one or may have an element of randomness in it.

> Links in a network often carry *weights* that quantify the strengths of correlation between pairs of nodes. The weights are important in determining how the network evolves, and may themselves co-evolve along with the other network features.

It is often an impossible task to describe the detailed structure of a complex network, though certain features represented by quantitative measures can be identified as being relevant ones in various contexts. For instance, one can talk about the *degree distribution* in a network. The *degree* of a node specifies the number of links attached to it, and is one of its basic properties. It is rare to have a real life network with all nodes of equal degree. More generally, a network with N number of nodes is characterized by the probability distribution $P(k_i)$ $(i = 1, 2, \cdots N,\ k_i = 1, 2, \cdots)$, which gives the probability of a node i (we assume the nodes to be labeled with numbers) to have a degree k_i. The *average degree* of the network with specified degrees (k_i $(i = 1, \cdots, N)$) is given by $\langle k \rangle = \frac{1}{N} \sum_i k_i$. A distribution carrying somewhat less information is the probability $P(k)$ that any arbitrarily chosen node has a degree k $(= 1, 2, \cdots)$.

The degree distribution gives the most basic information about a complex network, but is only one among a large number of characteristics indicative of its structure in quantitative terms. The degree distribution gives an indication of the relative importance of the nodes in the network connectivity and dynamics, telling us, among other things, how the important nodes are distributed within the network. A related measure is the *clustering coefficient*, which gives the probability that any two neighbors of a node (i.e., nodes connected to some particular node) are also neighbors of each other. A *cluster* in a network refers to a set of mutually connected nodes.

Also of major relevance in describing network structures are concepts relating to *walks and paths* [133]. A walk is a succession of nodes such that each pair of successive nodes is connected by a link. The number of links in the walk, in between its terminal nodes, is referred to as its *length*. Closed walks or *loops* are ones with identical terminal nodes, and provide information about *feedback* in the network. A *path* is a walk that visits no node more than once. The shortest path between a given pair of nodes can be made use of in defining a distance measure on a network.

Without going further into the issue of structural measures characterizing a network, we now refer to another aspect of network structure, namely whether or not it is a *random* one, a widely referred instance of which was introduced by Erdős and Rényi in early days of network theory, and is known as the Erdős-Rényi (ER) network. In a random network, some or all entries in the *adjacency matrix* (a matrix expressing the connectivity between nodes of a network [133]) are random variables. At times, the entire adjacency matrix can be one drawn from a random ensemble. The class of random networks includes ones in which, as mentioned earlier, the randomness is generated dynamically, where nodes and links can be randomly added or removed, i.e., the elements of the adjacency matrix evolve as stochastic processes.

An ER network (also referred to as an Erdős-Rényi-Gilbert network) with a specified number of nodes is characterized by a fixed (i.e., time-independent) probability (say, p) of any pair of nodes being connected by a link, independently of other pairs. Various possible realizations of this network may differ in their number of links and in their probabilities of occurrence. In a second type of ER networks, in which the number of links (L) is specified along with the number of nodes (N; the links are picked randomly between pairs of nodes), all realizations are equiprobable in the corresponding ensemble. Though an ER network is of a simple structure, it admits of a *phase transition* analogous to one observed in a percolation problem.

Complex networks, however, are generally not of the ER type, though the ER model often serves as reference in describing the properties of such real-life networks. At times, networks are defined and described by appropriate modifications of the ER rule. Networks differ widely in their characteristics, and a large number of such characteristics are often needed to analyze and understand any given network. Typically, the properties of a complex network are addressed by modeling the underlying network formation process where one specifies the mechanisms that drive the process, i.e., the dynamics of the nodes and links. Since analytical methods do not usually suffice to adequately characterize the resulting network, computer programs are most often resorted to.

4.2.5.1 Small-world networks

The description and analysis of complex systems in terms of networks is of remarkable value in numerous areas including the one of *social networks*, where it has been employed since early days. Social network analysis has led to the discovery and understanding of a number of features observed in complex systems, such as the *small-world phenomenon*, where the average length between arbitrarily specified nodes turns out to be small in some sense. This phenomenon is observed in numerous complex systems, among which are the ones first described in the celebrated work of Watts and Strogatz. Their approach was to start from a regular network and gradually introduce random links between nodes so as to obtain network structures intermediate between regular and random ones.

Random networks such as the ones of the ER type are characterized by a low degree of clustering and a relatively short separation between pairs of nodes chosen randomly. Most regular networks, on the other hand, are characterized by high clustering and large separations.

> The separation between a pair of nodes in a network is defined as the smallest number of links to be traversed in succession in moving between the two — if the two are not connected by an uninterrupted chain of links, the separation is defined to be infinite.

> As mentioned earlier, the degree of clustering of a node is defined as the probability that any two nodes linked to it are also mutually linked. On averaging over all the nodes in the network, one obtains the overall degree of clustering in it.

If the links in a regular network are replaced (in a random succession) with links establishing random connection between nodes, one obtains, even with a relatively small number of replacements, a network with high clustering and a *low* separation. It is typically found in the case of social networks that nodes have, on the average, a separation spanning only six links. More generally, networks have the *small-world* property even as these may have only a small degree of randomness in their structure. The small-world phenomenon arises because randomly established links dramatically reduce the separation between nodes (for instance, a link established between previously unconnected nodes reduces the separation from infinity to one). An analogous manifestation of a similar effect is the one referred to as the 'strength of weak ties', where weak correlations between strongly tied clusters result in conspicuous phenomena — often running counter to intuition.

> Network analysis is potentially of great value in understanding how *neuronal aggregates* function in the brain, where effects such as the small world phenomenon are likely to play an important role [126]; a number of related considerations will be found in section 4.3.11.

4.2.6 Complex systems: scaling and power law statistics

We now briefly mention the features of *scaling* and *power law* distributions in complex systems. A variable y is said to scale as a function of a second variable x if the relation between the two is of the form of a *power law*, $y \propto x^{\alpha}$, where α is referred to as the exponent or the degree of the scaling. Scaling relations are often not exact since the relation between the two variables may be influenced by small effects arising from other relevant factors. Scaling-based arguments are useful in understanding diverse phenomena of interest where some fundamental scaling assumption can be invoked to relate various features of a system in a simple manner, regardless of the underlying complexity. For instance, Galileo established the scaling relation $R \propto L^{\frac{3}{2}}$ (an instance of *allometric scaling*), between the radius (R) of a weight-bearing bone of an animal and the linear dimension (L) of the latter, based on the assumption that the strength of a bone varies as the square of its radius. Scaling relations are remarkable in that they relate diverse features without direct reference to the details of the system concerned. Of course, not every functional relation pertaining to a system can conceivably be a scaling, but the manifestation of the joint operation of a large number of correlated factors often appears as one.

Scaling is typically associated with *self-similarity*, in virtue of which a system appears similar when observed in various different scales. A homogeneous object is trivially self-similar, while interesting self-similarity properties are exhibited by *fractals*. It is common for a complex system to be generated in a self-similar manner for reasons of economy and adaptability. A remarkable power law relation in biology (another instance of allometric scaling) — corroborated by a large number of observations — is that the metabolic rate of an animal of linear dimension L, when compared with other species of different linear dimensions, scales as $L^{\frac{3}{4}}$ so as to keep them cool. This is explained by noting that, for efficiently transporting metabolites to all the cells in the body, the blood vessels proliferate in a tree-like manner, with branches forming a self-similar pattern. Scaling laws are known to arise in the context of critical phenomena, in which there emerge long-range correlations among components, whereby systems become effectively *scale-free*.

More generally, scaling laws arise in the statistical description of stochastic complex systems where distribution functions (probability distributions of relevant variables) typically exhibit scaling behavior — this contrasts with non-complex statistical systems (ones with non-interacting components or with interactions resulting in a simple network structure) where distribution functions are generally of the exponential type. A large class of distribution functions characterizing complex systems of diverse descriptions includes those of the *fat-tailed* type where the distribution has a power law tail. In other words, a fat-tailed distribution is characterized by a distribution function that goes like $f(x) \sim x^{-\alpha}$ $(\alpha > 0)$ for large x, where x is a random variable associated with the occurrence of a certain set of events characteristic of the system under consideration, large values of which correspond to rare events in the set. It is likely that the fat-tail phenomenon is associated with interactions among the components of a complex system that conspire to generate events that would be exponentially rare in the absence of interactions (*remote causes* in operation!). This, however, is not a precise statement, and no general or universal explanation is known for the generation of power-law or fat-tailed distributions, though a number of distinct mechanisms have been observed to lead to such distributions in large classes of complex systems.

Among these we mention here the occurrence of fat-tailed distributions in systems with *self-organized criticality* and those with *sample space reducing* (SSR) processes ([133], chapter 3).

Self-organized criticality is a widely shared feature of systems driven away from equilibrium by external means where a system approaches a critical state and becomes unstable, thereby exhibiting a new behavior. The term 'self-organized' refers to the feature that the critical state is approached, in the presence of the driving which need not be precisely controlled, due to the intrinsic interactions among the components of the system. Such critical behavior under driving is observed in the dynamics of a sand-pile that gradually builds up when grains of sand are gently dropped on a table-top. On attaining a certain critical slope, the pile collapses as more sand is dropped on it and an avalanche builds up.

The dynamics of the sand-pile is dominated by two opposing factors — the slow *driving* by the addition of sand grains, and the rapid *relaxation* by the movement of the grains along the slope of the pile — and is characteristic of a wide class of processes. The slow driving allows the system to find a local equilibrium till a state is reached when the local equilibrium becomes unstable and a rapid relaxation ensues.

Fluctuations in self-organized critical systems commonly exhibit approximate power law statistics in the relevant probability distributions, and the fat-tailed behavior is manifested in the form of a relatively high probability of occurrence of avalanche-like events. The slow-driving-rapid-relaxation scenario is commonly observed in geology, weather change, psychological processes, progress of diseases, onset of epidemics, crash in financial markets, and in many other diverse circumstances.

Sample space reducing (SSR) systems provide instances of the emergence of power laws in history-dependent (or *path*-dependent) processes, i.e., ones where the 'memory' of its previous states are relevant in determining the statistics of the system.

History-dependent processes are commonly observed in driven systems as in the case of linear response theory of non-equilibrium statistical mechanics where time-dependent *response functions* determine the evolution of expectation values of the observables of a system.

In a large class of SSR processes, the size of the sample space (the space of possible states, or behavior patterns — analogous to a phase space) gets altered as a process unfolds, and the statistics relating to the system dynamics becomes history-dependent. A common SSR scenario involves a driving force that takes a system to an 'excited' configuration, from which it relaxes towards an equilibrium state, where there may be a succession of such excitation and relaxation phases, with the system driven permanently out of equilibrium. In a simple model of such a process ([133]), one obtains *Zipf's law* (see sec. 4.2.6.1 below), with the probability of outcome i $(1, 2, \cdots, N)$ (in an initial sample space with outcomes marked $\{1, 2, \cdots N\}$, which then gets reduced at successive stages) given by $p(i) \propto i^{-1}$. Even as one widens the scope of the model, one finds that Zipf's law emerges as an *attractor* for the probability distribution.

More generally one obtains a fat-tailed distribution where, significantly, one can infer numerous details of the driving and relaxation processes from the form of the distribution.

SSR processes are of great relevance in a wide diversity of phenomena in different areas including those in science, sociology and linguistics. For instance, the probability distribution of words in meaningful sentences formed at random can be viewed as an SSR process. In a broad sense, *evolutionary* processes can be understood as being of the SSR type, while an alternative description in terms of self-organized criticality is also possible.

4.2.6.1 Power law statistics: the Zipf distribution

Perhaps the most widely known instance of power law statistics is *Zipf's law*, which constitutes a special instance of the Zipf distribution that gives the probability distribution of a set of discrete random variables (say, x_i, $(i = 1, 2, \cdots)$) in the form $p(x_i; \alpha) \propto x_i^{-(\alpha+1)}$, where α is a parameter characterizing the distribution.

The Zipf distribution often appears when the random variables x_i correspond to the *rank* in which the outcomes of an experiment appear when arranged in descending order. For instance, let the experiment consist of a count of the populations of a number of cities, with the ranks of the counts arranged in descending order being $1, 2, \cdots$. Repeating the experiment for a large number of sets of cities, one can then work out the relative frequencies of counts corresponding to ranks $i \ (= 1, 2, \cdots)$ and then the probability distribution p_i, which turns out to be of the form $p_i \propto i^{-1}$. This is referred to as Zipf's law.

Zipf's law or, more generally, power law distributions, are commonly interpreted as *emergent properties* (see sec. 4.2.8 below) of complex systems. The idea of emergent properties has had the dubious distinction of fueling the controversy between the viewpoints of *reductionism* and *holism* in the philosophy of science. I will put forward my own take on this issue later in this chapter. But emergent or not, the validity of Zipf's law can be traced to the rules of constitution of a system (i.e., the ones enumerating or codifying how the system is built up from its elementary constituents, or how these constituents are *correlated* with one another), in a number of instances. Thus, Li ([85]) demonstrated that Zipf's law emerges as the rank distribution of word frequencies for randomly generated texts simply from the rules of formation of the words and from the assumption that the texts are long sequences of the words.

4.2.7 Complexity and entropy

Complex systems admit of descriptions inherently statistical in nature. This may be due to the stochastic nature of the formation and evolution of a system or due to the fact that all descriptions of a complex system are, by the very nature of things, partial. Accordingly, the quantitative specification of the properties of a complex system is accomplished in probabilistic terms, in which *entropy* plays a central role. Depending on the nature of the system concerned, one or more of three notions of entropy can assume relevance (see [133] for details), namely, *Boltzmann-Gibbs* entropy as defined in statistical mechanics, entropy as *information*, and entropy derived from a *variational* *principle*.

> All the above three notions of entropy converge in the case of statistical mechanics of equilibrium states of thermodynamic systems where the complexity of a macroscopic system in equilibrium is described in terms of an equilibrium ensemble that specifies the probabilities of microscopic states of the system which, to be sure, is a partial description. However, for non-equilibrium processes, the notion of entropy is not well defined, though one referred to as the *diagonal entropy* in quantum statistical mechanics tuns out to be useful [31].

Among the three notions mentioned above, the one of entropy derived from a variational principle — commonly referred to as the *maximum entropy principle* — can be looked upon as being of relatively more general relevance in the context of complex systems. This principle proves to be extremely useful in the analysis of large data systems (an essential ingredient of modern day civilization with all its inhomogeneities, complexities and inherent irregularities) where *constraints* of various descriptions can be incorporated in deducing various statistical distributions by invoking the variational principle. In this, one uses the approach based on *Lagrange multipliers* — one associated with the name (among others) of E.T. Jaynes in the case of statistical mechanics.

4.2.8 Complex systems: emergent properties

With all this background outlined in the preceding sections, we now focus on the issue of *emergent properties* and *emergent phenomena*, widely thought to be the quintessential feature characterizing complex systems. Briefly, the notion of emergent properties tells us that a complex system is structured into various *levels* as in a hierarchy, and each level exhibits behavior that cannot be deduced from the properties of the constituents residing in the immediately lower level ('the whole is essentially different from the sum of parts') — there is novelty appearing at successive levels of the hierarchy. We submit that, like many other things in real life, all this talk of complex systems being characterized by levels and emergent properties is essentially a useful interpretation of our experience relating to complex systems — one that cannot be conclusively proved right or wrong on the basis of hard evidence, but certainly a rewarding one in the description and explanation of systems and events in various different *contexts*.

One of the most notable instances of emergent properties is *life*. When compared with isolated molecules, a living being is a most complex and remarkable object. As per our present understanding (I take lots of liberties here, just to make a point), isolated molecules came together to build up, stage by stage, more and more complex polymers when early life-like organelles made their appearance in an oceanic 'hot dilute soup', whose chemical reactions with the atmosphere enveloping the earth at that stage of evolution led to an oxygen-rich environment — thereby ushering in a phase of proliferating life forms. The point of this summary (if at all it can be called one) is to state that there is no *point of discontinuity* in the process of pre-biotic evolution as far as chemical reactions go — all the stages of the evolution were enacted strictly in accordance with the principles governing chemical reactions: *Nature did not know* that a momentous development was happening within its fold — one that would then lead to an even more momentous drama, if there could be one — the biological evolution. It is *our perception* that makes life so stupendously different from non-life.

> One may recall here the principle of 'quantity leading to quality' propounded by by Hegel and emphasized by Marx and Engels ([21]). This deeply philosophical principle has perhaps been interpreted and used rather shallowly in subsequent literature.

Does this mean that one can deny the enormous degree of *self-organized complexity* that we call life? By no means. But it is worth appreciating that the great distinction between life and non-life is a matter of *perspective*. When we stand back from the enormously complex chains of chemical reactions that occurred during the pre-biotic evolution and the various stages of biological evolution, and the even more complex reactions going on in a human body, and look only at the contrast between an inert cluster of molecules and a vibrant young person, we say that the latter is endowed with emergent qualities. This is the view of holism, which maintains that there is 'something more' in the whole as compared to its inert parts. But then rises the voice of reductionism: looked at from the point of view of the fundamental constituents and their mutual interactions, does a living human body differ from a single protein molecule?

It is not worthwhile to labor the point here but we will have to take it up again in sec. 4.3 — it can only be stated here that this book will have nothing new to present in the matter of the controversy between reductionism and holism since each of the two is only a point of view that depends on what perspective one adopts.

> It may appear that one is adopting here the attitude of portraying the viewpoints of reductionism and holism as vacuous, and the choice of one over the other as hollow and irrelevant. But that would be unfair on this book. All that is being said here is that it is no use trying to *resolve* the controversy relating to the two, or to establish one or the other as the one correct view of nature. Do we then prefer that the two views be *reconciled*? Once again, that would be a misrepresentation. Contrary viewpoints are neither resolved nor ever reconciled, but the very approach based on the *dichotomy* between the two views is fraught with problems. One has to be comfortable with the idea that *both the two* are points of view — Nature does not know if it is amenable to a description in terms of either the reductionist or the holist view — Nature is just itself and is completely indifferent to what our concepts about it are. Rather than trying to resolve between or reconcile the two seemingly contrary views, a more fruitful approach would be to accept the *indissoluble unity* of the two in our effort to understand nature.
>
> At this point, an analogy chosen from the context of philosophy of mathematics may be useful. In the so-called classical philosophy of mathematics and mathematical logic, the law of the excluded middle is accepted as a basic axiom. On the other hand, intuitionistic logic does not accept that 'law' as fundamental. It does not accept a statement as either true or false till one or the other is established by rigorous deduction. In other words, it tells us not to tag everything with *either* a 'yes' *or* a 'no'. At the same time, it does *not* assert that there lurks a *third* alternative — something beyond 'yes' or 'no'. But then, I must say no more on the philosophy of mathematics lest I should be putting my foot in my mouth.
>
> In summary, then, the good old Hegelian-Marxian dictum of *unity of opposites* ([93]) — a great aphorism if ever there was one.
>
> However, aphorisms are meant to be appreciated and marveled at, but they are too enigmatic to be adopted as working principles in the concrete and untiring work people undertake in interpreting nature bit by grinding bit. In other words, 'unity of opposites' is one way of looking at and interpreting this world of ours — a useful way as I understand it — more useful, perhaps, than adopting a dichotomous approach in life and in science.

For now, we will adopt the idea of emergent properties as we have adopted the idea of complex systems constituting nested hierarchies earlier in sec. 4.2.1. Indeed, it would be a sophistry to deny the existence of levels of complexity and of properties specific to those levels that leave no trace when looked at in the context of a lower or even of a higher level.

> The question that still remains is how to *interpret* the existence of levels and the emergent properties. We believe this question of interpretation remains largely open (questions of interpretation, of course, are never fully closed), and we will take this up again in sec. 4.3.8.2 later in this chapter. The issue of emergence comes up almost everywhere in any discourse relating to complex systems.

An instance of emergence is provided by the property of *wetness* of water [66], which is a property of water molecules in bulk, that leaves no trace in one single water molecule, because wetness arises in virtue of *interactions* between water molecules.

This is one way of looking at the phenomenon of emergence: an emergent property of a system appears in virtue of interactions between subsystems that populate a lower level of the hierarchy of complexity but is absent in the subsystems looked at individually. Even as this appears to be quite acceptable as a defining characteristic of emergence, we should point out that the idea of identifying an entity (a subsystem in the present context) abstracted from its interactions with *all* other entities is not of much relevance, as we discuss in sec. 4.3 later.

Emergence appears in other contexts as well. For instance, a system described in terms of a set of nonlinear differential equations exhibits bifurcations as some characteristic parameter is made to cross some threshold value (perhaps, due to a slow change induced by environmental systems), in consequence of which there occurs a transformation in the topology of the trajectories in the phase space, with attendant transformations in system properties. In the case of a complex adaptive system (CAS), an analogous transformation may arise due some change in the environment of a system, e.g., when some individuals in a population become better adapted than the rest and one or more traits specific to these individuals get preferentially transmitted to succeeding generations.

The two scenarios sketched in the preceding paragraphs tell us that emergence appears in situations that are, at least on the face of it, distinct. In one the emergence is due to the assembly of a large number of building blocks that interact among themselves — much like the putting together of letters of an alphabet that produces a meaningful word. One may imagine the process of generating the assembly to be carried forward through steps — words put together to form sentences, sentences assembled to make up a paragraph illustrating a theme, paragraphs put together to form a short story, and so on — perhaps generating a book, and then an entire library. The *other* scenario involves a changing environment inducing a momentous transformation in a system.

A *third* scenario involves a change in the rules of formation of a system from its building blocks when, at some stage some notable transformation in the system behavior emerges. This is spectacularly illustrated in the case of cellular automata ([66]), when a sufficiently complex rule of transformation is found to lead to the property of self-reproduction.

Finally, there is a very definite sense of emergence when one talks of the transition from the classical theory to the quantum mechanical theory, from Newtonian mechanics to the one based on the special theory of relativity, from the special to the general theory, or from the quantum theory to quantum field theory. In all these cases, a transformation of a theoretical scheme signals a distinctly new way of describing and explaining natural phenomenon as some parameter in the theory is properly taken into account without being dismissed out of hand (the Planck constant, the typical particle velocity in relation to the velocity of light, the strength of the gravitational field in relation to other forces in the theory, the rest masses of particles in relation to their energies).

In other words, there are numerous sources leading to *emergent properties* in complex systems, and emergence constitutes a powerful heuristic for the understanding of such systems across disciplines.

Still, the idea of emergence brings up philosophical and theoretical questions of a more fundamental nature. As promised, these will be referred to in sec. 4.3.8.2 later in this chapter.

4.3 Complexity and reality: a close look at scientific realism

4.3.1 Reality and our interpretation of it

The very first thing we start from is that 'reality' is something that may appear to be self-evident, but that there is a catch to it — appearances are treacherous.

The foundational position that we adopt in this book may be briefly stated in simple terms as follows:

> Reality exists independently of our mind — we ourselves are part of that reality — but all that we know and comprehend of reality is a matter of our *interpretation*. In other words, we have to clearly demarcate between 'reality itself' and *our conception of reality*.
>
> Signals of various descriptions are incessantly generated from all the innumerable parts of reality, some of which impinge on our senses. These include signals generated *internally* — ones produced by mental processes of diverse descriptions — reality is presented to us in the form of an external and an internal world, made up of *phenomena*. The signals are processed by our mind and form mental impressions and then our *interpretation* of our external and internal worlds — the mental processes themselves are part of the infinitely complex dynamics of reality at large. The interpretation is based on *concepts, beliefs, and inferences*, and is *all that we have* by way of our conception and knowledge of the world.
>
> All the while, 'reality itself' continues to exist beyond our conceptual world — ever instrumental in constructing and reconstructing that world and ever deceiving us into believing that our interpretations constitute a *true* representation of itself.

> Our mental conception of reality is not apart from and secluded from reality at large — reality and our interpretation of it are *implicitly* correlated and form a tangled whole. Nonetheless, the fundamental fact remains that reality exists by and of itself and is self-determined, and that our mental world exists as a phenomenon within it while, *at the same time*, that phenomenon constitutes our sole takeaway from reality — we *sense* reality, we form our concepts and theories relating to parts of reality, but can never know how faithfully those concepts represent 'reality-in-itself'.

This, of course, is not a new line of thought — there exists a long line of contributions from philosophers and scientists from antiquity to the present day, that cohere to form a certain position in the philosophy of science and, at the same time, help us in comprehending, explaining, and illuminating it — it is this position that we have tried to summarize above as we understand it.

> In speaking of philosophers and scientists, we do not mean to exclude the insights and ideas — often of remarkable value — offered us by authors, poets, artists, sculptors, musicians, and very many others from various walks of life. The point is that the authors, poets, and all the others *don't take it upon themselves* to explain the world around us and to fathom out our relation to that world. Scientists are supposed to explore and infer the mechanisms underlying the workings of nature, while philosophers engage with foundational issues relating to existence, reality, and knowledge. Any individual entering into such an engagement may be said to be taking on the garb of a philosopher, at least for a limited time and purpose.

At the same time, it must be mentioned in no uncertain terms, that alternative views of the world and of its relation to our mind exist and *there is no ultimate guarantee* of the *truth* of any of these alternative views (including the one we propose to adopt) — at the end of the day, all these remain nothing more than *points of view*.

> The point of view we adopt in this essay is similar in many respects to the one explained in greater details in [4]. We also refer to [82], which starts from the same basic position as Baggott's, and then sets off in a different direction, exploring the cognitive roots of how science inquires into Nature, focusing on how inductive inference

is enacted in the human mind and how hypothesis formation and (scientific) creativity may be realized in the mind of an individual. We also suggest [5], [6] as delightful and instructive readings that help in understanding the setting in which much of the present chapter is put together. Later in the essay we will point out a few areas in which the point of view adopted in this book regarding our changing conception of the world perhaps differs from Baggott's — but, once again, it is no big deal that one point of view may differ from another.

Even a child will agree that reality cannot be the same as its representation in our conception. But it is important to understand how this conception is formed and what knowledge it imparts of the world out there.

This is the concern of *scientific realism*, which inquires as to which items of our conception 'exist' out there and how our conception of reality *corresponds* to reality itself. In other words, scientific realism examines critically the nature of our interpretation of the world. And, it is precisely here that *complexity* enters in a big way.

Nature is one hugely complex system — it includes *all* the complexity there is. This is so self-evident a statement that one often takes it for granted — philosophy tends not to give explicit recognition to lessons learned from the *science* of complexity. In a number of sections to follow, we will first jot down a few basic things that the point of view of scientific realism tells us and will then indicate the directions in which one can hope to improve upon one's way of looking at reality by reckoning with the complexity of Nature — the *ultimate* in complexity, that is.

4.3.2 Observations on the viewpoint of scientific realism

This section is not meant to be an attempt at an exposition of the viewpoint of scientific realism but will include allusion to a few areas where one finds continuing controversy within its fold.

1. Philosophy, of course, is a discipline that thrives in controversy, and scientific realism is no exception in this. Still, within the camp of scientific realism, there are a number of questions that philosophers seem to be worrying about with somewhat more than average concern.

2. Generally speaking, scientific realism concerns itself with how science views nature and what the *method* of science is supposed to be. In the words of Baggott:

   "In fact, almost a century of intellectual endeavour and argumentation appears to have led the philosophers further and further away from a consensus on science and the scientific method." [4]

Among these worrisome issues is the one of the existence, i.e., the ontological reality, of 'unobserved entities' and the related one of the ontological reality of attributes of objects including ones that form part of the *theory* about these objects. Related to this is the question as to whether our theories are expressions of principles intrinsic to nature or whether, in contrast, they are *constructs* designed to describe and explain nature as captured in our conceptual world. And finally, there is the issue that poses the question as to whether successive versions of our theories represent progressively accurate descriptions of the mechanisms inherent in nature. All this needs a bit of explanation.

But even before we engage with the above worrisome issues, I will briefly comment on the apparently conflicting viewpoints of scientific realism and anti-realism.

4.3.2.1 Scientific realism and anti-realism

Scientific realism and anti-realism appear to be facing each other on the above issues across an unbridgeable gap. The gap appears to be even more unbridgeable in respect of the supposedly foundational issue relating to the very existence of reality outside and beyond our senses. As for me, I have already indicated that I am in tune with the viewpoint that there is a reality that is self-determined and is independent of our conceptions and theories. But, at the same time, I have also clearly stated that such an acceptance is no more than a point of view — the point of view that is associated with scientific realism. The seemingly opposite viewpoint that the existence of something called reality is a conceptual construct, is branded as anti-realism.

1. Differing (and even opposing) viewpoints emerge in our attempt at answering questions that relate to an underlying *complexity*. Complexity is something that is inherently beyond the practical possibility of a *complete* description or explanation. And life does not allow us infinite time or leisure to settle such

complex issues. In order to arrive at answers within our limited means and resources we adopt certain simplifying assumptions — all in the nature of beliefs induced from our prior experience. This is how all theories are built up and *meta*-theories are adopted — the latter being precisely what we call points of view. From a broadly general perspective, theories and meta-theories are all akin to beliefs generated in our mind. Among the web of beliefs, some are easily revised when weighed against evidence, but some are in the nature of durable beliefs that come to be formulated in such a manner as not to be in direct confrontation with evidence.

For instance, I may entertain the belief that 'very few people are honest' as against the belief of my friend's to the effect that 'most people of this world are honest'. It is entirely likely that we will pass through life with both our beliefs intact, without ever being troubled with contrary evidence. What is more, despite the beliefs being seemingly incompatible, we may spend our life being the best of friends.

In reality, no person is ever fully honest or fully dishonest. Honest and dishonesty are two descriptions that we use as tags attached to persons so that we can assess their behavior with relative ease without being burdened with too much of confusing intricacies.

2. The statement that a system is self-determined means that there is no external system or agency that determines its behavior. This, however, does not mean that the behavior is *determinable* or predictable through observations and inferences. Scientific realism makes an assumption that reality is self-determined, but does not bear responsibility of stating that the reality as a whole is determinable.

The opposition between apparently contrary points of view is never in the nature of an ultimate and indissoluble divide. There is always an extended middle ground between the two opposites, which bears testimony to the fact that the said opposites do not exhaustively divide the terrain one is looking at. It is often our shortness of vision that we do not see this and, having once formed a point of view, to which a contrary view is found to be posed, we become possessive of it (the other point of view similarly attracts its own takers) and we then come to be cursed with a *dichotomous* approach.

When scientific realism adopts the position that reality exists independently of our conception of it, that position is in fact one arrived at by way of a *meta*-induction. And when it further states that *all* our knowledge and belief about the reality is a matter of our interpretation since reality forever is only fractionally accessible to our senses while being the perennial source of our interpretations, it is actually adopting a certain position in the context of a *deep and complex problem* that we face in describing our relation to reality. To someone not versed in the nuances of philosophy, the opposing point of view that reality is something that lies in our concepts, does not sound too different.

This is not to say that the two points of view are not distinct from one another but only to state that the distinction is not a matter of hard evidence. Confronted with such seemingly incompatible points of view, one may profitably adopt the broader one based on the Hegelian-Marxian dictum of the indissoluble 'unity of opposites' (something like the belief that 'people appear to be honest or dishonest depending on circumstances') but that, again, would be no more and no less than still another point of view.

Two other areas of discourse where scientific realism and anti-realism appear to be in conflict are, *first*, the ontological status of 'unobservable entities' and their properties, especially ones that emerge in the context of the relevant *theories*, and *next*, the status of successivey revised *versions* of theory in various areas of science — whether or not they approach progressively to *truth* about Nature (see [100] for background).

4.3.2.2 Unobservable entities and theoretical properties

Broadly speaking, an 'unobservable entity' means one whose existence is inferred on the basis of indirect evidence, generally in theoretical terms. A commonly cited example in this context is the electron — an entity that cannot be observed with the unaided eye or even a microscope but one whose existence is still ascertained by numerous indirect means. Scientific realism accepts that the electron exists and that its ontological status is no different from other more mundanely experienced objects. It is to be noted, however, that the lack of directness in the observation of an electron means that there is an associated intrusion of theoretical concepts in what we refer to as an 'observation of an electron'. For instance, take the case of a *positron* — the *anti-particle* of an electron. The observation of a positron is yet more problematic as compared to an electron since it requires a certain minimum of energy to be generated and does not last long in material environment, and its existence, ascertained in specially designed experimental conditions, is yet more a matter of theoretical interpretation — indeed, the positron was first predicted on theoretical grounds in the context of relativistic quantum mechanics.

This raises the issue of how we interpret *theory*. As we see, the apparently simple question relating to the existence (or the *reality*) of an entity is bound up with the deeper question of the status of theories in relation to reality. Before dealing with this issue of what a theory signifies, let us focus on the apparently simpler issue of the *properties* of an entity such as those of an electron.

When we speak of an electron, it seems that we speak of it without reference to other entities in nature. The identity of an electron is established in terms of its charge and rest mass, in addition to its *spin*, the last named being a quantum mechanical property of the electron. While it seems that this identity is established without regard to other entities in nature, in reality it is not so and, moreover, it does not provide us with a complete understanding of what an electron 'really' is.

The assertion of the mere existence of an entity constitutes a metaphysical statement unless we also specify how that entity is related to the rest of the world — how it behaves in the company of other entities of nature. The mass, charge, and spin of the electron tell us a lot about how and where the electron is located within the infinitely complex web of inter-relations between the entities of nature but these still do not give us a reasonably complete information of how the electron behaves in this complex world of ours. Such information comes only at the cost of theory. For instance, the charge of an electron assumes relevance only in the context of the theoretical statement that the force between an electron and another charged particle conforms to the inverse square law named after Coulomb. While the Coulomb law describes the force between a pair of static charges, one can complement this with the force on a moving electron in a magnetic field so as to make up the expression for the composite *Lotentz force*. If the spin of the electron is also brought into the picture and one finally adds the gravitational force on the electron exerted due to other massive particles around it (commonly, however, the gravitational force turns out to be of negligible magnitude), one gets a reasonably complete description of how an electron is expected to behave in the company of other entities in the world. However — and this is of great relevance — such extended description is *still* inadequate for many purposes since it does not include the so-called weak interactions of the electron encountered in nuclear and sub-nuclear events, and what is more, does not refer to the *fermionic* nature of the electron.

All this goes to show that the mere existence of an electron is vacuous in so far as our understanding of it in this world of ours is concerned, and the fact of existence assumes significance as it gets concretely manifested through its behavior in the universe made up of other entities. And we find that this is already a matter of quite extensive *theory*. Though the electron is uniquely identified in terms of its rest mass, charge, and magnetic moment, that identification is hollow unless looked at in the context of an appropriate theory which gives meaning to the rest mass, charge, and the magnetic moment, these being still inadequate to completely specify what an electron is. What is apparent is that the theory is context-dependent, i.e., needs to be continuously upgraded as the electron is observed in circumstances more and more remote from the world of our direct experience.

When we speak of an electron, we give expression to a concept lodged in our mind. The electron we speak of in the context of the Bohr theory of the atom is certainly not the same thing as the electron of the standard model. All this means that when we speak of the existence of an entity we *refer to* some aspect of the reality out there (the *noumenal* reality, refer back to sec. 1.12; see also sec. 4.3.6 below) in some context or other, but that referent is not known to us as it is (and cannot be spoken of as such) — it is an entity (correlated with all other evolving entities out there) about which our conception keeps changing.

4.3.2.3 Theory as code

The really significant issue in respect of our knowledge of a particle such as the electron comes up when one considers it to be moving in the field created by other particles that themselves move under mutual interactions, since it is only then that the relevance of the *equations of motion* of a system of particles appears in its true light (we do not consider here the field theoretic description of the 'fundamental' particles). This is made clearer by referring not to particles with electrical and magnetic properties but to ones having gravitational interaction alone since the principles involved are then understood in simpler terms.

We do not refer here to the space-time dependence of the gravitational field in interaction with massive bodies or to the issue of its integration in a unified theory of a broader scope. We will come to the question of the emergence of successively revised structures of theories later in this chapter, in sec. 4.3.8.3.

Considering a hypothetical situation where a number of massive particles interact among themselves in accordance with Newton's law of gravitation, one writes down the set of *equations of motion* of the particles, which is then supposed to capture *all the theory* pertaining to the system in question.

There arise two questions that demand our attention at this stage. The first of these pertains to the set of particles along with the equations of motion as constituting a *model*. And the second relates to the issue of the theory describing this model being in the nature of a *code*. The former question will be addressed in a later section (sec. 4.3.5; the idea of a model was briefly referred to in the introductory section of this chapter, sec. 4.1), while the latter is of fundamental relevance for now.

Does the set of equations of motion based on Newton's law of gravitation *describe everything that can possibly happen to the set of particles in the model*? Evidently not, since theory is no detailed description of reality (the experienced reality, that is, within the restricted context of the model). If a theory is expected to provide a complete and detailed description of experienced reality then it misses its purpose. In fact, a theory is the essence distilled from reality and its usefulness lies in its ability to produce a description when appropriately *unpacked*, and to lead to *predictions*. In the present context, the unpacking is done by *solving* the equations of motion based on an appropriately chosen set of initial conditions. The solution constitutes a part of theory too but one that pertains to the mathematics of differential equations and is not an integral part of the theory pertaining to the model in question.

The initial conditions (and the boundary conditions in the case of a set of partial differential equations) involve items of information *external* to the system considered in a model — information whose relevance arises from the fact that a model is something *abstracted* away from the rest of reality.

As mentioned, a useful way of expressing what a theory signifies is to say that it acts as a *code* —- and, the unpacking referred to above is analogous to the process of *decoding*, with a key provided by the mathematical theory of solving a given set of differential equations. The code, evidently, is not the same thing as the *result* obtained in the process of decoding, which is the actual stuff that the code was designed for in the first place. What that result would be for a given set of initial conditions can be known only *after* the decoding is done. And, that result may contain huge surprises, vanishing without trace however hard one looks at the code alone. Imagine a computer program written so as to obtain the answer to an intricate problem, that answer being known only when the program is actually run on a computer, with the necessary data fed to it, and that answer actually adds to our knowledge while the program in itself does not. The famous 'four-color problem' was solved with the aid of computer programs, and its solution (and not the programs) constituted mathematical knowledge on a question over which mathematicians had struggled for a long time.

> Other analogies where the action of a set of rules operating on a basic package, in conjunction with additional information fed by hand, leads to the unfolding of diverse consequences that add to our knowledge of the relevant models include (a) the generation of the multitude of theorems of geometry from the basic axioms by the action of a set of rules of deduction, (b) the operation of rules of sentence formation on a set of core principles of a grammar, generating an entire language when employed in the context of a set of words as enunciated in a dictionary, (c) the solution to the Schrödinger equation (the code) in the context of an appropriate Hamiltonian, leading to the spectral characteristics of a molecule, when invoked in the presence of relevant data.

In summary, a theory pertaining to a model acts somewhat like a code in respect of the detailed behavior of that model when appropriate rules of decoding are employed in conjunction with appropriate information setting the *context* of the decoding process. *Knowledge* of how the model is expected to unfold under various situations is obtained only when the decoding is actually performed, and cannot be meaningfully said to have been 'already there' within the code.

What is more, when looked at in the context of explaining and predicting the behavior of the system (some part of the universe abstracted away from the rest) represented by a model, the 'code' (i.e., the theory) itself is generated from experience by a process of abduction, and gets revised from time to time in a process that is said to be a major feature of science. We set aside this process of continual revision of theory for the time being, and consider how the code *leads to* but cannot be said to *contain within itself* all the knowledge about nature that we can arrive at, including all the novelty that we can predict.

> This, incidentally, is the big difference between a theory and a code that one has to keep in mind. A code is written by someone who knows how it operates, but a theory is not like that, unless we adhere to the view that it is our guess at a code 'written by God'. God or not, a theory is just a clever guess at how various parts of Nature are expected to behave; it is our abstraction from observed reality and is supposed to be our guide to unknown terrain in our journey through the maze of nature at large. What the theory tells us on being unpacked is not known beforehand because we are no God who is supposed to have written the entire code of Nature. On being unpacked the theory may be found to be junk or else, may provide us with knowledge of how various parts of nature behave. In this book, we outline the view that our ceaseless attempts at revising our theories does *not* constitute a process of arriving at some ultimate code hidden within the folds of Nature. As we have repeatedly mentioned, this is just a point of view, nothing more.

In order to illustrate this we return to the consideration of the equations of motion (based on Newton's law of gravitation) of a system of particles. When solved with data corresponding to initial conditions for a system of two particles, these equations describe a regular behavior of elliptic motion of the particles around their common center of mass — indeed, Newton's law was arrived at in the first place principally by observations pertaining to the two-body problem. However, when one tries to solve the equations for *three or more* particles one observes a multitude of behavior patterns of the model including regular motions along with irregular or *chaotic* ones. In other words, the theory (comprising of a set of equations of motion — the 'code' in the present context) pertaining to a number of particles interacting by Newton's law of gravitation does not in itself constitute a description and explanation of the behavior of the particles under various possible circumstances, and the question that now assumes relevance concerns the way such description and explanation is arrived at. As we have pointed out earlier in sec. 4.2.1, the process of arriving at the description and explanation is more often than not at least *as complex as* constructing the theory itself. As a point of interest, one can, if one wishes, refer to Newton's law of gravitation as the *theory of everything* in so far as systems of gravitating particles are concerned, but that does not, in itself, have as much of relevance as one could wish for in providing us with knowledge about behavior of such systems.

> This is because we would not know what the theory entails, before the consequences of the equations of motion are worked out for all possible contexts (i.e., all possible numbers of particles and all possible initial conditions) and actually compared with phenomena in nature — not only the ones that we now experience, but ones that will be played out at all time to come everywhere within the infinite expanse of nature. Who knows, there may be surprise hidden somewhere — indeed, the lessons from the theory of complexity tells us that there *will* be surprises, perhaps when we least expect these (the fat-tailed distribution at work!) — and we will have to guess at a revised theory. We do not know beforehand what our theory is going to uncover.

Before we proceed further, we include here *a brief summary* of the topics touched upon up to this point in the present section.

> Nature generates all our conceptions pertaining to what it is and how it works, but those conceptions do not capture Nature as it is, in its entirety — the question of how nature conceived by us corresponds to nature-in-itself is a deep and complex one. Indeed, it does not carry much sense in talking of 'nature-in-itself', not as much because the latter is transcendental — it is *a priori* and independent of our senses, though — as because it is the ultimate in complexity.
>
> The mere *existence* of entities in nature is devoid of meaning and relevance unless we also talk of their *properties* and the way they behave in the world. And, properties are not inherent in those entities independently of their mutual interactions. Even assuming that an electron is uniquely identified by its charge, mass, and spin, its behavior is known only when the Lorentz force law

governing its motion in the presence of other particles is specified, in addition to the universal law of gravitation — we do not consider, for the moment, further revisions in the theory of particles and their interactions.

This means that not only is the question of the mere existence of entities devoid of content, that of their properties too is meaningful *only* with reference to appropriate *theory* — the issue of the existence of unobservable entities (indeed, of *all* entities) is essentially and inseparably tied up with the theory describing their behavior. Theories governing the behavior of entities constitute an important component of *our interpretation of nature*. Statements about the existence of entities are meaningful only to the extent that theories explain our observations on them, though customarily the question of existence is seen as one separated from that of theories. Put differently, 'existence' and 'theory' are inseparably tied with each other.

Theories, however, are much like codes that are to be unpacked appropriately to actually lead us to knowledge of how the entities making up the world behave and evolve in it — theories are meant to *explain* why the entities (or systems of entities) behave the way they do and to *predict* how they (the systems, that is) are expected to evolve. What is more, theories themselves are arrived at by a process of induction from our observations of parts of nature in interaction with one another, and are liable to revision with accumulating experience of ours. Finally, theories in themselves would be meaningless unless these were supplemented with relevant data — the initial and boundary conditions — so essential for arriving at concrete results from the code, which is an abstract and distilled essence of our partial experience of reality.

It is necessary to dispel here a possible misconception. Theories are constantly in the process of being updated — concepts come and go. Does this mean that the existence of entities is also in a state of flux — that an entity may cease to exist and a new entity appears in its place as our conception changes? That, certainly, is not the case since, reality exists and evolves by and in itself. It is only our concepts that change. Our concept of an entity and of its existence can undergo revision since that concept is arrived at by notionally isolating the entity from the rest of the complex world. In fact, however, no entity exists in isolation from the rest of reality.

4.3.2.4 The question of ontology

We continue here our observations on scientific realism. Our explanations will entail occasional repetitions which, one hopes, will help clarify matters.

While the assumption — we have called it a point of view — of a mind-independent nature (the 'reality') — is said to distinguish scientific realism from anti-realism, the ontological assertion of existence becomes meaningful only when it is complemented with the ontological significance of the *properties* of entities that exist and then, as we have seen, of *theories* from which we can derive the behavior of entities and systems (of those entities).

However, it seems too far-fetched to talk of theories being 'real' in any sense. Various points of view within the fold of scientific realism do not unequivocally assert the ontological reality of theories, mostly because theories are endowed with a *fluidity* — they come and go, and do not *reside* in the entities of the world. This is a view that makes a clear separation between the existence of entities, and the theories that explain and describe their behavior — however, this needs to be examined closely if scientific realism is to come clean on consistency.

When we talk of an electron — an electron describing a circular arc in a magnetic field and ionizing a gas it moves through — we are actually making statements within our experience of reality and our interpretation of it. It is important to recognize that these statements only *correspond* to things and their behavior 'out there', the interpretation being made possible by means of signals sent out from these components of reality to our senses and our instruments. However, the exact nature of this correspondence remains fundamentally undefined. In other words, when we speak of the existence of an electron, we actually make a statement of *our perception* of a tiny bit of nature — and that perception is a complex one, including in itself a chunk of interpretation that, essentially, is in the nature of theory.

Among these statements regarding the behavior of entities, the *existence* part is implicitly (and almost universally) distinguished from the *theory* part (Lorentz force, ionization potential, and all that). It does not seem problematic to state that the electron spoken of in our interpretation *truly* corresponds to the electron residing 'in reality'. When it comes to theory, however, the correspondence is taken to be true but only 'approximately' (whatever that may mean) — there appear revisions from time to time in the theory, supposed to be taking us progressively closer to the 'actual' behavior of electrons 'out there'.

> Indeed, the very notion of an entity is deeply theory-dependent, though one commonly glosses over this aspect in philosophical discussions: as we have commented earlier, the 'electron' featuring in the non-relativistic quantum theory of atomic spectra is certainly not the same entity as the 'electron' in the standard model.

From the ontological point of view, theories are taken to 'correspond' to objectively existing 'laws of nature' — ones that are supposed to determine the behavior of entities. And, what is more, all these laws of nature (Newton's law, law of gravitation, Maxwell's theory, and so on) are commonly assumed to be *embedded* in an all-embracing 'unified' theory, also inherent in nature, which all our theories progressively lead us to.

The point of view — to be further explained below — adopted in the present book differs from the one expressed in the last paragraph. It posits that theories are constructs put together in our interpretation of reality (see sections 4.3.4 and 4.3.7 for further elaboration) and have no counterpart inherent in nature — they are contextual and undergo *non-monotonic* revisions. What remains objective is the behavior of entities that they predict — the behavior we experience in our phenomenal world, since theories are constructed precisely to explain this behavior of systems at various levels of complexity. It is the behavior of entities that gives them identity — an identity that is once again contextual, being the phenomenal aspect of some tiny part of the real world out there. One may think of entities along with their behavior — the two can be separated only notionally — as a 'projection' onto the context set by our senses and our instruments of observation, along with our prior beliefs and theories emerging in our interpretation of the world.

> Indeed, speaking in all generality, *everything* is a projection from the nouemanl world to the phenomenal (for the terms 'noumenal' and 'phenomenal', refer back to section 1.12; see also sec. 4.3.6) — the entities we perceive, their interactions and properties, and our experience of events in this world. In the same vein, everything is context-dependent — what we perceive and what we explain by means of theories change radically as the context is changed. This is the fundamental position that lies underneath all we present in this book, as we explain in the following paragraphs.

It all appears to be confusing stuff, but confusion is avoided only at the cost of inconsistency.

4.3.3 Complexity and incomplete descriptions

Complexity throws new light on all the discourse about *correspondence* between *the* reality out there and our interpretation of it.

It is precisely because of the all-pervading complexity of nature that *whatever* we say or think about it *has to be partial and incomplete*. And, it is precisely because of this too that *no* statement of ours ever *truly* corresponds to the reality out there. There is no question even of an approximate correspondence since theories change dramatically as the context of a model (to which a theory applies) is changed (see sec. 4.3.5 below).

> As we mentioned earlier, the distinction between our conceptions and the reality out there is not difficult to recognize and accept. What is more difficult, however, is to see why they need not correspond exactly and truly to each other, subject to inessential corrections from time to time. It is here that complexity assumes command over everything and overrides all other philosophical considerations.

The fact that it is essentially metaphysical to talk about reality outside of and away from our interpretations, does not owe its origin to any transcendence or any similarly deep mystery that lies beyond our comprehension — it is due to the *pervasive complexity* of nature. This is something that we infer from our conceptions of parts of nature that we are confronted with in the course of our everyday experience and our scientific explorations. Consider, for instance, a gas made up of an enormous number of molecules. It is only a tiny part of nature, and is not even a highly complex system if one assumes that the molecules are all identical classical particles with fixed and identical two-body interactions between them. Even so, nobody can make meaningful statements about *all possible* types of behavior of the gas — anything we say about it is by the very nature of things a partial and incomplete statement — one where some aspect of its behavior is abstracted out from the rest. The idea of the gas considered in isolation from the rest of the universe is itself an idealization and constitutes a *model*. Even when some kind of interaction with external systems is considered, say, with a specified field or with a heat reservoir, one is still left with a model — an extended one to be sure, but a model nevertheless.

All our experience, all our scientific explorations, always occur within some limited *context* set by the range of objects whose interactions are thought to be relevant in respect of the experience, and by the limitations of our senses and our scientific instruments, along with our current state of personal and inter-subjective beliefs and concepts. All these together interact in the setting up of a *model* — an abstraction from the complexity of nature — complexity to which no limits have ever been found to apply. Nature, in other words, is *infinitely complex* — with an effectively infinite number of *dimensions* to it, i.e., an effectively infinite number of independent aspects relevant for a supposedly complete description pertaining to it.

In other words, *everything* that we consider relates to some model or other — whether or not that model is precisely defined. Models in physics and chemistry are mostly defined with great precision, where the role of our vaguely defined beliefs and concepts is done away with in terms of a number of precise mathematical assumptions and mathematical rules of derivation. Even so, the theory applying to such a model does not always lead to precise and definitive conclusions because of mathematical difficulties involved, and one needs schemes of *approximation* in respect of the models. These approximations are symptomatic of the complexity of the model in relation to known rules of mathematical derivation. One can go further and say that no theory, approximate or otherwise, ever applies to reality at large which is infinitely complex and cannot be experienced or explored *as a whole* either by means of our senses or by means of scientific instruments.

Put differently, any and every theory applies to some part or other of nature at large as reflected in our experience, sought to be described in some context that may or may not be precisely defined. If one focuses on some part of our experience without specifying the context even by implication or even vaguely, then the model itself becomes undefined.

On the other hand, it might appear that one and the same system can be the subject of different models under various different contexts. For instance, a gas made up of an enormously large number (say, N) of molecules, all confined to move within a specified volume (say, V) may be looked at either in isolation from all other systems around it or in interaction with a large heat reservoir that can exchange energy with the gas, where particle exchange between the two systems (the heat reservoir and the gas under consideration) is not allowed. In the latter case, the energy of the gas can fluctuate along with fluctuations in the energy of the reservoir, covering a range of the microscopic states of the latter. One can, however, *average* over all these fluctuations over the microscopic states of the reservoir and focus upon the states of the gas alone and set up the systems in such a way that this constitutes a model of the gas (without overt reference to the reservoir) for specified values of V, N and the *temperature* T, as compared with the other model for the *isolated* gas where , along with V and N, its *energy* gets fixed at some specified value (say, E).

> We have to keep reminding ourselves from time to time that, when we speak of a model as an abstraction describing some part of nature, we actually mean some part of our phenomenal experience of nature.

Can one then, in all fairness, say that the two models refer to the *same* gas? This apparently simple question to which everyone will answer yes, takes us back to the issue as to whether and to what extent a 'system' or an entity can be abstracted away from its behavior resulting from its interactions with systems around it. It is only in rare situations that one can talk of a system all by itself, as in the case of the isolated gas. The complexity of the world implies that even very weak interactions with the rest of the world can lead to essential modifications in its behavior under the appropriate context. In other words, various models of what appears to be the same system are to be considered as essentially distinct entities since they correspond to distinct contexts in which its interactions enter into the definition of these models, and to very different behavior of the 'system' as such. Here the term 'interaction' is meant to refer to both the interactions between the subsystems and those with the surrounding systems.

Significantly, *the theory pertaining to a model depends to a remarkable degree on the context defining it*. This question will engage our attention to a marked degree later in the present chapter (sec. 4.3.8.1).

4.3.4 Ontology of reality: entities and correlations

The ontology of reality pertains to the issue of what resides 'out there'. Truly speaking it falls within the domain of metaphysics to speak of entities and their properties that constitute ontological reality. Nonetheless, it can and does fall within the scope of our world-view — our philosophical point of view or, in a manner of speaking, our mindset — that forms the backdrop of the more substantive statements that we make in science and philosophy. Though a point of view cannot be either proved or disproved, it does influence the *mode* in which we inquire into the workings of nature.

> An illustrative analogy: I may have the mindset that few people in this world are honest, while my friend working with me in our project lab may have the one that most people are honest. Evidently, this differing viewpoint will influence the evaluation of and working relation with a new entrant to our lab that each of us will develop.

Thus, when we speak of an electron as it exists in reality, we actually speak of something else — an electron as it exists in our interpretation of nature in some context, specified either precisely or vaguely. As we have indicated in earlier paragraphs, this distinction between entities in the *noumenal* and *phenomenal* worlds is devoid of meaning when applied to an entity (such as an electron) *abstracted away from its properties* — its behavior vis-a-vis other entities in this world. What is more, the behavior of entities cannot, in a similar vein, be abstracted away from *theories* abut these. Theories are mental constructs meant to provide us with explanations and predictions, and hence belong to the phenomenal world of our interpretation. The question that scientific realism now faces is the following: to what, if anything, in the noumenal world do the theories correspond?

As we have seen, scientific realism is inclined to posit an unambiguous correspondence between phenomenal and noumenal entities. So, according to the same trend of thought, there has to be an unambiguous correspondence between a theory and the associated noumenal posit — an intrinsic *law of nature*. Just as phenomenal entities acquire meaning only in relation to appropriate theories, similarly, it is only 'natural' that the corresponding noumenal entities acquire 'meaning' within the fold of laws of nature. Let us, for the time being, ignore the fact that it is *meaningless* to talk of 'meaning' in the noumenal world unless we allow ourselves to tread dangerously close to the assumption of a Creator.

> *I*, for one, don't have an issue with a viewpoint that looks up to His powers, but scientific realism seems to have; this makes me want to examine closely the posits of scientific realism as I do in this essay, since I should, at the end of the day, like to call myself a realist — nonetheless, I certainly don't have an issue with other points of view such as anti-realism and social constructivism. I can only clarify what my viewpoint is — there is no intent whatever to prove its *correctness*, if only because that is an impossible task anyway.

In the course of all our experience in the world, what we find in the behavior of entities is their all-pervading mutual *correlations*. More precisely, the world of our experience is made up of an infinite multitude of mutually correlated *events*, where an event refers to an entity (a 'particle' in a description arrived at by abstraction) marked with *space* *and time* co-ordinates. Theories are constructed out of all the multitudes of correlations between events by a process of abduction — as we mentioned, theories are the distilled essence of our experience.

What corresponds to the phenomenal correlations are, once again, correlations in the noumenal world — correlations existing 'out there'. In other words, the ontological reality of entities acquires meaning only in association with the correlations among events existing in reality, outside our conceptual world, independent of all our conceptions and beliefs. Once again, this is an assertion without proof, but this, at the end of the day, constitutes the point of view of scientific realism.

While this completes the story as far as we are concerned in this book, the question remains as to what our *theories* correspond to. This question arises when we extrapolate from the original posit that events in the noumenal world correspond, in some sense, to phenomenal events. This seems to raise the expectation that *something* has to exist so as to correspond to the *theories* too.

But theories, as we have said, are nothing but the distilled essence of our experience, obtained as *constructs* in our conceptual world resulting from the interpretation of our experience, and scientific realism has absolutely no responsibility to seek out correspondence regarding everything that exists in our conceptual world. For instance, it is absolutely not a fact that all our beliefs and concepts have a correspondence with something out there. And, theories are nothing but glorified beliefs about nature that apply to *models*. Of course, theories are beliefs of a very special kind — ones that have to bear the responsibility of producing results that must have a close correspondence with some part of our experience. Even so, they have no responsibility whatsoever to have a correspondence with something existing out there independently of our conceptions.

At this point, we'll *sum up* this part of our discussion.

> Our experiences in life are generated by multitudes of signals perceived by our senses, aided by scientific instruments, that give rise to impressions

> and concepts in our mind. As such, they must undeniably correspond to real events happening 'out there', in the noumenal world that lies beyond all our conceptions and interpretations constituting our phenomenal world. This correspondence is fundamentally made up of two parts — a correspondence between events in the two worlds, and one between *correlations* between events. Having said this, however, we must hasten to add that events and their correlations form one indissoluble whole — splitting the two must lead to philosophical pitfalls.
>
> The task of *theories* is to generate conclusions *within the context of models* that fit with experience. A model is a cleverly constituted object that abstracts from actually existing events in our phenomenal world so that the conclusions of a theory applying to it agree closely with actual experience in some domain or other.
>
> A theory is like a belief formed in our mind and does not have to correspond to anything in the noumenal world, though the conclusions generated from it have to have a close agreement with our experiences and hence, with events and their correlations in that world — this, indeed, is precisely the reason that a theory is constructed in the first place.

It is the notion that theories have to correspond to, or to fit with, what are referred to as 'laws of nature' that lead to questions and paradoxes. In particular, there arises the question as to whether our theories, by means of successive revisions, approach more and more closely to mechanisms inherent in nature that explain and predict the behavior of reality as a whole. The notion of *incommensurability* of successive versions of a theory is at odds with this commonly entertained idea of approach to the ultimate *truth* inherent in nature. We will come to this in sec. 4.3.8 later in the present chapter. For now, we focus on a number of issues that will finally lead us to an appraisal of this idea of correspondence of theories with intrinsic mechanisms of nature and then to the one of *one single* grand theory describing the foundational principle governing reality at large.

Further elaboration of the the concept of incommensurability of theories and of how it fits in with the viewpoint of scientific realism will be found in the sections 4.3.8.3 and 4.4, the latter being where all the present considerations will be summed up.

4.3.5 Models and their significance

Experience with complex systems we encounter in the real world tells us that these are generally characterized by the feature of *co-evolution* (sec.4.2.1), where *everything* pertaining to a system evolves with time, including the constituent entities and the nature of their mutual interactions. Often, the interactions can only notionally be separated from the entities themselves — the two being represented by links and nodes in the network representation of a complex system — and it is their co-evolution that is fundamentally responsible for the behavior of the system itself.

> Recall that, with reference to the noumenal or the phenomenal reality, it is more meaningful to refer to *events* rather than to entities — events are entities marked with appropriately chosen *space and time* co-ordinates. There is a subjective or observer-dependent aspect to space and time co-ordinates, though the entire multitude of co-ordinate systems corresponding to different possible observers are related to one another in an objective way — one that determines the *structure* of space-time. The structure of space-time is part of the theoretical framework of physics, and is objective in the sense that it is independent of the mode of description adopted by this or that specific observer. The *concept* of space and time (or, briefly, space-time) applies to our phenomenal world and refer to the noumenal world only partially and incompletely. More precisely, space-time has the same status as entities and events, in respect of which there is a partial and context-dependent correspondence between our experience and the noumenal reality.

A model is an abstraction where one or more aspects of the co-evolution may be ignored for the sake of simplicity of analysis. The fact that the model can be useful *in spite* of such simplification needs separate attention on our part. Recall that a complex system, in its time evolution, generally involves a number of distinct *scales* in space and time, where the term 'space' need not mean the three dimensional physical space but a phase space that can have an arbitrarily large number of dimensions. The existence of such scales (sec. 4.2.1) is generally a manifestation of a spectrum of interactions between the components of such a system. Such multiple scales characterizing the behavior of the system results in co-existing nested regimes of stability and instability where a stable regime (in space and time) can, in an approximate sense, be described in terms of simpler, *reduced* systems, or *projections*.

As a simple instance, we refer back to the gas (with specified values of V, N) isolated from its surrounding systems with some specified energy (E; see sec. 4.3.3). Observed on a *large time scale*, the gas attains a state of stable equilibrium when its properties can be modeled in relatively simple terms, though even such a simple model may involve formidable mathematical difficulties in yielding conclusions that can be compared with experimental observations. On the other hand, one can compare this model of an isolated sample of a gas with a model where the gas attains equilibrium while in contact with a large heat reservoir at some specified temperature T (which now replaces the energy E of the gas). In either case, the behavior of the gas can, in a sense, be described as a projection from a complex, non-equilibrium evolution.

In this context, we recall our earlier comments regarding different possible models describing the 'same' system. One may feel justified in saying that the two models of a gas referred to above describe the *same* system — a gas. Strictly speaking, however, the two models refer to distinct entities described in differing contexts relating to the interactions of the gas with other systems around it where the interaction may result in exchanges of various types, including the exchange of chemical species. And, since it is meaningless in principle to talk of a system without reference to its interactions with other systems in the world, the two models are to be looked at as distinct ones, though in a practical (and loose) way of speaking, they may be said to involve the same system. Indeed, a model is made up of a certain set of subsystems (say, the molecules of a gas) with certain interactions characterizing these *without regard to* what happens elsewhere in the universe. Such models yield results in close agreement with what is observed in real systems, if the interactions ignored in the model are in fact of negligible consequence in the case of the relevant experimental situation within some specified time horizon. In the case of the gas in contact with the thermal reservoir, this is the situation if the interactions with the reservoir are sufficiently weak.

> *In summary*, complex systems are characterized by stable and unstable regimes in space and time, distributed over a range of scales, as a result of which one can have systems that can, in approximate terms, be represented by means of models — ones that lead to results in close agreement with observed behavior of the systems. Various different models can be constructed for any given system, where the latter interacts in various different manners with other systems within a bigger and more complex system. Though one can loosely say that all these different models pertain to the same system, they can imply very different behavior of the system because it is characterized in these models by interactions of different kinds. In this sense, they represent different *entities*.

As we see, it is in principle not quite right to speak of 'a model of a system' since, more pertinently, a model describes *a set-up* involving one or more systems along with their interactions with other systems, where the latter may be considered as *external* to the model. However, the interactions with these external systems have to be incorporated in precisely defined terms (commonly, in the form of boundary conditions) in order that the model may be useful.

Thus, simply stated, a model is a chunk scooped out from a bigger complex system, where all relevant interactions, both among internal subsystems and with external systems are suitably specified, as part of an *idealization* — though an idealization that is often close to observations. In the case of biological and social systems, or more generally for complex adaptive systems (CAS) (refer back to sec. 4.2.1) the interactions cannot be precisely specified, but one can specify in general terms how the states of a constituent in a complex system change under the influence of other constituents it interacts with. Interactions among subsystems lead to correlations that constitute the basis of the behavior of systems. Generally speaking, the behavior of a system obtained from a model, may depend markedly on the *context* in which the model is defined.

The statement that theories apply to models (and *not* to systems) and that the behavior of a system that is derived from the theory describing a model may depend markedly on the context — a simple enough statement on the face of it — has far-reaching philosophical consequences. To see this, we have to take into account one other important set of facts about theories and models that we will now have a look at.

To begin with, models are useful only when they lead to behavior of systems in close agreement with the behavior observed under conditions that may be actually realized. For instance, the model of a gas in interaction with a large thermal reservoir where the interaction is of such a kind as to correspond to some specific value of the temperature T, very closely reproduces (under an appropriate set of additional conditions that we keep implied for the sake of simplicity) the behavior of an actual gas in a metal cylinder kept exposed to air in a room under stable atmospheric conditions.

Further, when one commonly states that a model describes the behavior of a system on the basis of a theory, it seems on the face of it that the theory is *known beforehand* and the behavior of the system under consideration *follows* from it. In reality, the theory underlying a model often results by a prolonged process of abduction and inference based on observations on real-life situations, under proper laboratory control, if necessary. In other words, one has a complex and interwoven relation where theory is induced from experience and then that theory is applied to models so as to lead to behavior that closely reproduces the behavior observed in real-life situations.

In contrast to real-life situations where the complexity of the phenomenal world is often of non-trivial consequence, models are useful because they can be defined with precision, regardless of the complexities existing in the real world beyond the scope of their definition. In the case of the physical sciences, the definition of a model can be as precise as one likes, and the rigorous rules of mathematical derivation can be invoked to work out the consequences of the theory applying to the model, subject to the errors of mathematical approximation that are often found to be essential in such derivations (these statements apply only qualitatively to complex adaptive systems; see comments below). What is more, in the mathematically well-defined models, one can also work out the limits of error within which the results can vary so that, on comparing the consequences of the model with experimental observations, one can determine whether and to what extent the model deviates from reality in virtue of the abstractions and simplifications involved in setting it up. All this goes to make the models and the theories useful and essential in the physical sciences.

On the other hand, in geology, meteorology, biology, population genetics, epidemiology, economics, finance, administration, social studies, and similar other fields, one meets with progressively diminishing mathematical rigor, though the use of high-powered computers have brought all these fields within the fold of what can be loosely referred to as the 'scientific method', where the consequences of models can be worked out (to within limits) and compared with experience. Significantly, as we move along the above list of fields of study, the focus shifts progressively from CPS to CAS, and the variety and complexity of behavior increases.

In all such areas of study, the complexity of nature makes it imperative to make use of models and theories — induced from real-life observations by a process of abduction and inference — where appropriate rules of derivation are invoked so as to work out the consequences of the theories describing the interactions characterizing the models (among subsystems and with external systems as set by the context of a model), and to finally compare the findings with experience in real life, gained under controlled conditions wherever possible.

In summary, theories are constructed in a process of abduction, inference, and abstraction, and are applied to models — theories are nothing but constructs that are of vital necessity in making sense of the infinite complexity made up of entities and their correlations that we experience in our phenomenal world and, eventually, of the *real* world from which the phenomenal world derives in a process of conceptualization and interpretation.

4.3.6 The noumenal and the phenomenal

The terms ‘noumenal’ and ‘phenomenal’ have been taken from Kant, from whom the modern era of discourse on scientific realism can be said to have originated (see [42], esp., chapters 1, 6). However, Kant’s terms of reference regarding the two worlds were different as compared with those used in the current discourse on scientific realism and, moreover, numerous distinct points of view relating to the exploration of reality undertaken in our scientific enterprise continue to exist from days preceding Kant.

In contemporary terms, the noumenal world, which is the real world beyond the phenomenal one captured in our concepts, exists in and by itself, regardless of any kind of ‘intellectual intuition’ that transcends our ‘sensible intuition’ ([128]) — which is where the ‘noumenal’ referred to in the present essay differs from that in Kant’s point of view.

> However, we intend not to harp on differences but to explore where different points of view interpenetrate so that there may result a deeper understanding of these, along with the possible emergence of broader points of view — more fruitful in making sense of our existence and our experience in this complex world of ours.

As our understanding of the current literature goes, we refer to the noumenal world as being made up of events (i.e., entities located in space-time) and their correlations that correspond to entities along with with their interactions in the phenomenal world captured in our interpretation of the noumenal world. As for the entities and their interactions in the phenomenal world, these are manifested (i.e., their *being* gets expressed in their *becoming*) only through their *properties*, where the latter correspond to correlations of diverse types in the noumenal world. In this sense, we speak of interactions and correlations in the phenomenal world as having their counterparts in the noumenal or the ‘real’ world.

The infinite multitude of correlations among events in the phenomenal world (and, correspondingly, in the noumenal world too) are sorted out in our conceptual universe by means of *theories*. Theories, however, are constructs in our mind, constituting the distilled essence of our infinitely complex experience, and are meant to explain and predict the course of occurrence of events. Theories apply to models and capture partial truths about the phenomenal reality. They do not possess a counterpart in the real world, and do not correspond to purported laws of nature leading us to intrinsic mechanisms underlying the processes in the real or noumenal world. What are termed 'laws of nature' can only refer to the theories resulting from our interpretation of events experienced in the phenomenal world.

Here we include a few words on space, time, and the structure of space-time. Space and time (or, space-time in brief) are identifying indices ('co-ordinates', somewhat like book-keeping entries) that constitute an ordering among events in the phenomenal world, where we assume for the sake of convenience of reference that a similar ordering applies to corresponding events in the noumenal world as well. The space and time co-ordinates assigned to an event are observer-dependent, though the ordering itself, expressed in terms of the space and time co-ordinates that can vary from one observer to another, is observer-independent. The theoretical descriptions applying to the phenomenal world posit a *structure* of space-time that is made explicit in the general theory of relativity where, once again, the structure is an objective (i.e., subject-independent) concept, depending on the mass-energy distribution in space and time (the so-called 'stress-energy tensor') and determines a set of quantities depending on the gravitational field strength.

> The structure of space-time is an objective thing in the sense that it can be treated on the same footing as events in our phenomenal world that correspond (in a sense that cannot be specified completely) to events in the noumenal world. However, the way this structure is accounted for in the current theory of gravitation is model-dependent and contextual.

Does the theory of space-time and gravitation apply to a model or to our entire phenomenal universe? As with every theory, it does apply only to a model — one where all interactions and events on length scales smaller than the so-called *Planck length* and time intervals smaller than the *Planck time* are ignored. This actually sets the context in which the general relativistic theory of space-time and gravitation is defined since it allows one to deal with quantum mechanical effects and gravitational effects independently of each other.

As stated earlier, general relativity is nothing but a theory that forms part of our interpretation of nature and, strictly speaking, is not intrinsic to the 'real world'. As a theory, it is no doubt a remarkable one that explains a vast range of natural phenomena but is still a distilled essence of our experience within this range — one where quantum mechanical effects have no influence on gravitation. Beyond this range, one needs a fundamental restructuring of the theory, where quantum field theory is to be integrated with the theory of gravitation.

To continue, we have spoken of entities and their correlations in the noumenal world. This is where we have implicitly taken a liberty in projecting our concepts arrived at by abstraction from the phenomenal world back to the noumenal. It is an abstraction to separate entities or events from their correlations, manifested in the form of various properties of the former in the phenomenal world. Strictly speaking, we do not have hard evidence to make specific statements about the stuff the noumenal world is made up of. But points of view do not wait for hard evidence, and that of scientific realism asserts that the latter is made up of entities and events *corresponding* to the ones encountered in the phenomenal world. Indeed, our senses have evolved in such a way that some such correspondence holds because of the obvious adaptive value of it. However, the same adaptive process ensures that the entities and their interactions we sense are context-dependent. Thus, what we sense and observe with our bare eyes provides us with only a cross-section of the entities so observed (for instance, a tree leaf as a flat object of green color), while more demanding and specialized contexts reveal other particulars (the venation of a leaf, and its stomata when examined by a botanist). It then becomes impossible to speak of an entity 'as it really is'. Entities existing in the infinitely complex reality are complex systems themselves, and there is no final or ultimate description of such a complex system — only so much of the system is revealed in any given context.

For instance, in a hypothetical situation corresponding to a more pervasive experience of nature on our part, the 'noumenal stuff' may even turn out to be an infinitely extended field satisfying a nonlinear evolution equation.

> Even if one could write down a Hamiltonian density describing such a field, that would still not qualify as the ultimate theory of reality since a further expansion of our range of experience may demand a renewed attempt at a radical revision of the theory since, simply stated, the latter is nothing more than a construct.

Still, the noumenal reality is not something intangible and transcendental — it is a concrete thing, though infinitely complex. If we prevent ourselves from making any assumption whatsoever of this reality, we can speak only of the 'noumenal stuff'. But our points of view and our inferences arrived at inductively do not retreat timidly when called upon to make assumptions just because of lack of hard evidence — we constantly keep on making assumptions till some are proven wrong by evidence. Bold assumptions, consistent with whatever evidence we have, are things of thrill and are capable of making us more and more adaptive in our journey through life.

Scientific realism does not care much to be non-committal and to describe the real world as one made of just 'noumenal stuff'. When we have the perception of touching a table-top our senses give rise to some specific conception in the phenomenal reality, but this does not mean that we have had no contact with something tangible in the 'real' reality out there — we did indeed have such contact (as scientific realism is not afraid to tell us) when signals were sent out to our senses (to be processed in our mind), giving us the impression of the table-top in our phenomenal reality. What is more, when we act back upon reality, it is the noumenal world that our action operates on, though the *result* of that action is once again a matter of interpretation and is revealed in the phenomenal reality.

There is, however, a rider to the statement that the noumenal reality is not just a philosophical construct — one that it does not pay to ignore. While accepting that the noumenal reality is real enough and not just a cleverly made up concoction, it does not pay to be too confident to say that the conception of the table-top, based on the signals sent out to our sense, is *all there is to it.* Every conception, every interpretation, is incomplete and partial, having had its origin in some partial and contextual interaction of our senses (extended by the use of instruments) with parts of the noumenal reality — the latter is infinitely extended and an infinite-dimensional complex system, incessantly sending out an infinite multitude of signals of all kinds in all directions, to be captured again by parts and constituents of the same system, setting up correlations between all the innumerable parts of the same noumenal reality, while some of these signals are received by our senses which are themselves parts of the same reality. In other words, our conceptions are nothing but consequences of correlations between parts of the noumenal reality.

We repeat that all this is nothing but a point of view, an assumption that may be termed a meta-induction, one we adopt as a guide to our scientific quest — no more and no less.

Entities that are partially and contextually sensed by our interactions with parts of reality (the noumenal reality, that is) are not sensed in isolation from their properties — their interactions with other entities sensed in a similarly partial manner. Like the perceptions of the entities, the properties are also a matter of sensation and perception, i.e., in the ultimate analysis, of *interpretation* in our mind. Finally, the properties of entities are explained by *theories* that are likewise constructed in the mind by a process of abduction in which induction and deduction go hand in hand.

4.3.7 Do theories correspond to 'laws of nature'?

The statement that theories are constructs does not receive open-hearted approval from scientific realism, since the latter seeks to establish a correspondence between theories that reside in our minds and the purported 'laws of nature' residing in the real world lying beyond the mind.

> We disregard here the flaw inherent in the notional separation between the mind and the 'real world' — as if the mind is something distinct and apart from nature — and accept it as a way of simplifying our discourse. Even as the mind is actually a part of nature, it is a special entity capable of forming impressions of the rest of nature and even, to some extent at least, of itself.

The only things that one is to accept as 'real' are entities and their correlations (or, more precisely, the indissoluble unity of the two) — *all* the rest are our interpretations aimed at making sense of the infinitely complex experience arising from these, where 'making sense' means a process of adaptation of our existence with the world of our experience.

> Bas van Fraassen, from his philosophical position, refers to theories as constructs ([139]). Likewise, Nancy Cartwright ([22]), from a different philosophical position, denies ontological status to theories. Theories, indeed, are constructs applying to models meant to represent parts of the phenomenal reality, having no counterpart in the noumenal reality.

To repeat, the correlations among the entities in the noumenal world create an impression, by means of multitudes of signals (these being, in the ultimate analysis, in the nature of correlations themselves), in our mind that we call experience. Theories are nothing but constructs in our attempt at sorting out the complex experiences and making use of these for the purpose of explanation and prediction.

The notional flaw that, at times, afflicts scientific realism consists of trying to project our theories on to the noumenal reality — to assume that *something* must reside out there that generate the theories in our mind just as entities and their interactions in the world of our experience are generated from the 'real' entities and their correlations by means of signals. That 'something' is referred to as a 'law of nature' specific to some domain of inquiry. However, in order that such a correspondence may exist between theories and the purported laws of nature, either of two things has to happen: *either* signals of some kind are to generate this correspondence, which is a possibility we discount as having had no evidence to rest upon unless such signals are of divine origin, *or* the laws of nature are ones generated by a process of abstraction *analogous* to the one in which theories are arrived at in our mind.

Looking at the second possibility, which is in the nature of an extension of the first, one has to assume that the 'laws of nature' are to be inherent in the noumenal reality, containing in them the distilled essence of all entities and their correlations, and that these 'laws' get impressed in our minds as theories by some circuitous and mysterious process operating through the world of our experience.

All these possibilities involve the operation of some mysterious factor (factor not accounted for by experience in any way whatsoever) in virtue of which our theories can be accepted as counterparts of laws of nature residing in the reality out there — something quite antithetical to the spirit of scientific realism.

I, for one, consider myself an adherent of the viewpoint of scientific realism where I use that term to mean *the reality of entities and correlations*, everything else being relegated to our phenomenal world and our interpretations of it. As part of this interpretation, we assume that *interactions* occur between the phenomenal entities that generate the impression of their properties and regularities of behavior, and in a further surge of philosophical fervor, one may go on to assume that these interactions, properties, and theories, all have their counterparts in reality. However, 'being generated by reality' is not the same thing as 'having counterparts there'.

But there are trends in scientific realism that find it hard to desist altogether from notions generated in fervor: unless there is some regularity inherent in nature in the form of laws, how can our theories be so highly successful in explaining our experience and in predicting so accurately the behavior of phenomenal entities?

Here, however, one must be loyal to the lessons learned from complexity. A vast number of models of complex systems have by now been analyzed in mathematical terms and in more general approaches involving algorithms and computations, including ones based on AI systems that are endowed with learning abilities. All such studies point towards one single feature that all these systems have in common — all are characterized by a enormous range of spatial and temporal scales showing a multitude of stable regimes of behavior existing in space and time. The 'regularities' of the phenomenal world reflect one or more of the stable structures in one or more of such scales.

As mentioned several times earlier, the term 'space' means one made up of possible states of a system that can be of an arbitrarily high dimension. Every stable regime has its own effective state space and refers effectively to some subsystem generated as a projection of the entire system on to some lower dimension that captures the stable behavior in question.

The reality out there is the *ultimate* in complexity, and it is completely beyond our conceptual ability to capture all these spatial and temporal structures that possibly characterize the noumenal reality in its present state of *self-organized complexity* (see sec. 4.3.8.1 below; see also sec. 4.2.1). It is highly likely that our mental apparatus — almost infinitely complex as it itself happens to be — and our scientific set-ups can access only a few of these regimes, and our theories, obtained by a method of abstraction from the world of our experience reflect the regularities of only these few regimes. It is essentially philosophy in fervor — out to resist contamination from what is referred to as anti-realism — that *projects* the regularities captured in our theories on to 'reality at large'.

Scientific realism tolerates much debate and dissent within its own fold, but it gets emotional and puts its foot down when it comes to the question of the *truth* of theories supposed to describe the mechanisms underlying the workings of nature. Theories, according to major trends in scientific realism, may be revised from time to time but that is only because of our limited means in grasping the vastness and the intricacies of Nature, in virtue of which there always remain a gap that succeeding waves of theory building and theory revision are to bridge in the future. As these major trends claim, relinquishing the claim to truth is taken to stand for an attitude of pessimism and surrender to anti-realism.

But we must not ourselves get carried away in our own fervor and project ourselves as a critic of these major trends. Viewpoints are not to be fought over, as we must constantly keep on reminding ourselves. The worry is not over the correctness or otherwise of the position we adopt (such correctness can never be proved on hard evidence), but over the tendency of maintaining a sharp and irrevocable divide between what are described as 'scientific realism' and 'anti-realism' based on the single issue of locating truth in our theories and seeking a correspondence between these theories with purported 'laws of nature'. The assumption of the existence of such a sharp demarcation is better avoided if we are not to get shackled under the weight of our own viewpoint.

The point of view of anti-realism makes no bones about asserting that theories are mental constructs, and it may be argued that the position we adopt is then one adopted by anti-realism. One need not be too worried at being identified with this camp of philosophy or that — the practice of attaching tags to philosophical positions has its uses but cannot be stretched too far.

> We have stated that points of view cannot be fought over. But that must not mean that *discourse* over points of view is pointless. A philosophical mindset is a durable thing and does not go away overnight. But we are not born with our respective points of view — these are generated in the course of our journey through life, depending on how we confront and look at our accumulating experience. Likewise, points of view can change as well — in the course of experience once again, and by discourse and communication. There is no point *fighting* over points of view, but one can certainly try to *understand* a contrary point of view without being *dismissive* of it. Only then can a *synthesis* of the contraries be brought about, freeing us from permanent bondage to some mindset or other.

This is where we have to pause and, once again, *summarize our position* on the point of view of scientific realism as outlined up to this point of the present chapter — we will add to it in subsequent sections.

> Scientific realism accepts the mind-independent existence of a reality made up of *entities and their correlations* — the two being inseparably linked into an integral whole. Myriads of signals of an enormous diversity are received by our senses and our instruments, from which is generated a huge canvass of *interpretations* of this reality in our mind. Reality in itself (the noumenal world) is known to us solely in terms of our perceptions and our interpretations, that generate the phenomenal world for us. The noumenal entities and their unfolding correlations generate the experience of phenomenal entities, their interactions, and their behavior. One can, in principle, think of a correspondence between the phenomenal entities and their behavior on the one hand, and noumenal entities and their correlations on the other.
>
> The reality that generates phenomenal sensations in us is an infinite dimen-

sional and complex system, parts of which are captured in our conceptual world through these sensations. In keeping with the behavior of complex systems in general, that reality is co-evolving along with all its components and their correlations, and develops a multitude of *structures* made up of stable and unstable components in a multitude of *scales*. It is precisely the constituents in some stable components in some particular set of scales that we sense as the entities captured in the phenomenal world in any given context. Thus, what appears as a tabletop in the context of our ordinary everyday observations, appears as a collection of molecules in a different context of observation on a finer scale.

Theories are formed in our mind by a continuous process of abstraction involving inductive and deductive inference (a special class of inductive processes is referred to as abduction). It is the theories that play the essential role of sorting out the complexities of behavior of entities that we experience, of making sense of that experience, and of providing us with explanations and predictions. Strictly speaking, theories are applicable to *models* constructed out of our experience by simplification that allow us to define these with precision and apply precisely defined rules of deduction from these theories — the precision being understood in relative terms. The consequences deduced from theories are compared back with our experience, thereby resulting in a process by which we make sense of our world and adapt ourselves to it.

Theories are constructs in our mind and do not necessarily correspond to counterparts in the noumenal world. The purported 'laws of nature' residing in the reality beyond our mind are suppositions based on a projection from the phenomenal to the noumenal world. This goes against the fact that theories constitute the distilled essence of our experience constructed in a process of abstraction. Theories, moreover, are applicable to models that are incomplete and partial representations of the phenomenal world in the sense of being obtained as simplified versions of parts of the the latter. In this sense, theories are relevant as *locally* valid (in space and time) coded descriptions of the phenomenal reality.

An analogy of a mathematical nature may make things a bit more clear. One can look upon all our experiences and interpretations, including all our theoretical constructs relating to models, as *projections* from an infinite-dimensional noumenal reality on to finite dimensional 'sub-spaces'. A projection is a well-defined mapping that sets up a correspondence between two universes. However, the *inverse of a projection makes no sense — it is not well-defined.* Analyzing our interpretations and our theories, *we are not entitled to project back* because that, indeed, is metaphysics. Scientific realism consists of accepting the reality of the infinite dimensional noumenal world and of the 'projection' on to the phenomenal world and from there to our world of interpretations.

The questions that now remain are, is their a *global* theory valid for the entire phenomenal world of ours, embracing all the local versions? And, is there a *corresponding* grand unified law of nature? We in this book have no answer to these questions since, strictly speaking, these do not carry meaning — theories are mental constructs of a local nature and have no counterpart in the noumenal world. It is here that our position bears a strong resemblance to anti-realism. Nonetheless, we still consider our viewpoint as being realist in virtue of adopting the assumption of a mind-independent reality, as we have stated above in clear terms.

We will have occasion to briefly discuss this again later in this chapter, in sections 4.3.8, and 4.4.

This brings us to the concluding part of our discussion of theories of reality where we look at a number of issues arising from the above position regarding reality and our conception of it, notable among these being the question of successive revisions of theories.

4.3.8 Theories of reality: in search of the 'ultimate theory'

We begin by recalling that theories are mental constructs that apply to models, that models and the theories applying to those are contextual, that various different models refer to different cross-sections of the phenomenal reality, and thus, to different parts and different cross-sections of the noumenal reality too. We recall also that a theory pertaining to a model is a distilled essence of observations made on it, arrived at by a process of abstraction and abduction.

Theories, in addition to being partial and contextual, are often found to require *radical revisions* from time to time. This is a matter of major discomfort to dominant trends in scientific realism. In the present section we are going to put forward our view on this vexed question in scientific realism, when a number of related issues will also come up for consideration, based on the standpoint we have outlined in the previous pages of this chapter.

4.3.8.1 Theories are contextual and domain-specific

We first note a few examples that tell us how and in what sense theories apply to models, and are contextual and domain-specific.

Domains refer to specific areas of inquiry in our scientific endeavor. Thus, physics, chemistry, biology, medicine, geology, meteorology, social sciences, economics, business administration, governance, all are instances of domains. Among these, the social sciences, governance, public policy, business administration and such other domains were not traditionally considered to be related to science. But these are now routinely studied with the help of high-powered computers in terms of network representations of complex systems, and theories are routinely formed and applied to such systems. *Psychology* and the behavioral sciences form an interface between the domains of the physical and the social sciences.

Domains are made up of sub-domains. Thus, physics has innumerable sub-domains such as electromagnetic theory, gravitation theory, statistical mechanics, and so on, where there are further sub-divisions too. As sub-domains make up domains, the latter in turn can be grouped into broader fields where each field has certain common foundations and common methods, with broadly common features in the structures of their theories. Indeed, these broader fields, along with the domains, sub-domains, and the sub-divisions of the latter, can all be looked upon as forming a complex system among themselves, that can be represented in terms of a network having a hierarchical structure, as revealed in the multiplicity of journals devoted to general subject areas and to specialized topics, and in the papers published in those.

Theories are to be found in the papers published in the various journals, where the lineage of the theories can be traced from the domains and subdomains these papers refer to. Taking, for instance, a domain in physics, such as electromagnetic theory, one can see that there are innumerable sub-domains (and further sub-divisions too) with corresponding *theories* such as short wave asymptotics (i.e., ray optics), interference, diffraction and scattering theories, theory of antennas and waveguides, non-linear optics, quantum optics, and so on. There is a sense in saying that all these theories (with some exceptions that are based on the quantum theory of radiation) are just special topics within electromagnetic theory, but that does not go against the observation that each of these require special and specific approaches, and journal papers devoted to any two of these special topics bear little resemblance between them. It is very much a matter of *context* as to how one is to look at a theory devoted to any of these special topics — as a distilled essence of investigations into phenomena relating to the sub-domain (or a sub-division of a sub-domain), or as a part of a bigger theory.

There is a *correlation* between all these theories, and actual *structures* inherent in reality — the noumenal reality, as we have called it, where the terminology is borrowed from Kant. As mentioned earlier, reality is a hugely complex system — the *ultimate* in complexity, made up of an infinite multitude of dimensions, i.e., the number of independent entities required for a complete description of it which, truly speaking, is an impossible limit to achieve. It is, moreover, a dynamic system characterized by the feature of *co-evolution*, where all its components and their correlations evolve in a mutually determined manner. In consequence, reality is endowed with an infinite multitude of self-generated stable and unstable *structures* at an infinite multitude of *scales* in space and time. This, at times, is referred to as *self-organized complexity* (see, for instance, [138], and references therein) — in a loose manner of speaking, one can say that various different parts of a complex system pass through successive stages of self-organized criticality — there can be even be a large number of simultaneous instances of such transition in various different regions of the phase space, with self-organization emerging in between successive episodes of criticality. This results in a huge canvass of self-organized complexity for a real-life complex system.

Instances of such structures and scales are to be found in solutions to sets of non-linear differential equations (see [51] for background), mostly obtained in numerical computations. All these structures at various different scales provide a multitude of different *contexts* in which observations can be made (referring to those that are accessible to our senses and our scientific instruments), experience gained, and theories constructed as distilled essence of experience.

What is the relation between all these theories in the various sub-domains and domains, all having some correlation with actual structures existing in nature at various scales?

Theories pertaining to various sub-domains belonging to a given domain *dovetail* with one another, analogous to the way the languages and customs of different communities located in contiguous geographical regions are related, with overlapping features. All these languages and customs may have a common denominator, depending on the physical features of the region the communities are located in, and on the history of the migrating groups of people they are descendants of, but these common denominators do not determine the specific features pertaining to the communities. Likewise, theories pertaining to sub-domains may bear common birth-marks as descendants from a parent theory, but all are arrived at by *independent* processes of abstraction and abduction from experience gained in the respective specific contexts. These may all even be embedded in an overarching theoretical framework, but that does not detract from their autonomy — the overarching theory does not determine the specific features of these 'smaller' (but not lesser!) theories.

The electromagnetic theory encoded in Maxwell's equations does definitely provide the ground in which the ray theory and the diffraction theory of optics have germinated, but both these have had independent histories of development, with independent processes of abstraction having led to these. Both can be shown to have a common lineage in the form of limiting relations to the electromagnetic theory, but they are in no sense *determined* by the latter. Moreover, the two theories have a common boundary where they are related in a complex manner.

> Individuals having distinct mind-sets start talking at cross purposes at some stage when they are engaged in a discourse, because they *construe* differently — words and phrases carry different *meanings* for different people. This, does not, however, mean that communications and attempts at understanding one another must stop, having reached a dead end. Much of our understanding in this world depends on *tacitly* held views, where apparently contrary items of thought coexist without annihilating each other, integrated in our conceptual world — concepts have interpenetrating boundaries, while explicit statements show them as belonging to two incompatible groups of ideas, separated by a sharp boundary.

> The term 'determined by' or something 'determining' something else, results in a lot of confusion and misunderstanding. Do the Maxwell equations determine the theory of ray optics? Does the theory pertaining to the structure of the DNA molecule determine the behavioral diversity among people? Does the Schrödinger equation determine the structure of a complex molecule? Does a code determine the decoded script and the consequences that the latter can lead to? In each of these cases, one has to understand the complexities of meaning hidden in any attempted response to the question — there is a sense in which the response can be either yes or no, with none of the two being wrong; and again, there is a sense in which the two together make up a complex response — one it is difficult to make explicit.

Complexities in meaning only result from the complexities of the system it is supposed to refer to. In the above instance, the complexities of meaning of the term 'determined by' depends on the enormous complexity of our conceptual space and then, eventually, to the complexities of the parts of reality the term refers to. Thus, the electromagnetic theory encodes the behavior of electromagnetic fields in general while the ray theory encodes the behavior of short wavelength fields — the unpacking of the theory in the two cases describe the complex behavior of systems of which one is a part of the other, but the very complexity of the systems prevent the corresponding theory from being redundant. In other words, *complexity resides at all levels* and leads to the necessity of theories to be developed independently though, possibly, with one (say, ray theory) being enveloped by another (Maxwell's theory), where the latter is in the nature of an overarching theory. As an analogy, one can think of a novel by a great author with a complex plot spanning a huge spectrum of space and time, and with a sub-plot in it describing the intricate relation between a man and a woman that is itself a deep and troubled one — a relation that is in the nature of a hopeless tangle, generating a multitude of contrary emotions that can never be resolved. Here the sub-plot develops within the overarching plot and is compatible with it, without the latter determining the former.

Every theory has to have a *context*, even though of a vague and non-specific nature. And the context goes a long way to set the entire tone and texture of the theory. In other words, theories that dovetail with one another may nevertheless have different contexts and very different *structures*. The ray theory describing the behavior of short wavelength electromagnetic radiation has a completely different structure when compared with the quantum theory of radiation where one has to take into account the interaction of radiation with matter. Or, again, take the example of a group of men in a religious congregation and one in a political demonstration. Even when the two groups of men have a common cultural background, the (vaguely formed) theories that generate our anticipation of the behavior of the two groups are completely different, though the two behavior patterns are only apparently incompatible with each other and have underlying common links. This last observation goes to show that theories are only incomplete guides to understanding the behavior of complex systems.

4.3.8.2 Emergent properties and emergent theories

We recall (sec. 4.2.8) how complex systems are often characterized by *emergent properties*. Emergent properties are commonly associated with emergent *structures* appearing in the ceaseless dynamics of a complex system arising out of the varieties of interactions among the constituents at various levels of it — and both, in turn, are associated with *emergent theories*.

As stated several times in earlier sections, stable and unstable structures in a complex system appear at all scales in space and time. At any given point of time, the stable structures constitute the *self-organized complexity* of the system. It is by means of self-organized complexity that the hierarchy of levels referred to in sec. 4.2.1 make their appearance in the course of dynamical evolution of complex systems in various different contexts. For instance, the unconscious mind of an individual and its interaction with his conscious mind are explored by a psychologist in her consultation room, while the interaction of many such minds assumes relevance for the manager of a business organization. Decisions and policies of many such organizations, on the other hand, affect the economy of an entire country. At each such level, there is a measure of *decomposability*, i.e., the subsystems or the constituent units interacting with one another have an identity of their own. It is their interaction, along with the interactions with external systems (external to the system of interest, that is), that determine the behavior of a complex system perceived at some particular level.

As an instance, the properties of a liquid in bulk are determined by interactions of its constituent molecules and also its interactions with surrounding systems such as the atmosphere. Likewise, the properties of the molecules arise as consequences of the interactions among the electrons, protons, and neutrons making it up. However, the interactions between the electrons, protons, and neutrons are not directly involved in determining the bulk properties of the liquid, as determined by the interactions among its molecules. Moreover, the behavior of the liquid under diverse circumstances can be described mostly in terms of its bulk properties (such as its density, viscosity, compressibility) without direct reference to the molecules — these bulk properties emerge as *statistical* averages of molecular interactions. One then says that the bulk properties are *emergent* ones as these are independent of the *details* of the molecular interactions. On the other hand, the *theory* describing the properties of the liquid — one which links these bulk properties to averages over the molecular interactions — is an emergent one with reference to the theory describing the behavior of the individual molecules. All this, along with the observation that the liquid state itself emerges in a phase transition from a gaseous aggregate of molecules, establishes the statement that emergent structures, emergent properties, and emergent theories are all linked by a common thread — the one of *complexity*.

The theory of the liquid state (let us call it the 'A-theory') is an emergent one with reference to the theory aimed at establishing the structure of molecules (the 'B-theory', for easy reference), based on quantum mechanical principles. Broadly speaking, the two theories share a common ground but apart from this common lineage, they have little in common with each other — they are, in a manner of speaking, *independent* theories. Certain basic ingredients of the A-theory (ones relating to the inter-molecular potential) can be understood in terms of the B-theory, but that is about all though, of course, both the theories share the common language of physics and mathematics. There is some sense in saying that the A-theory emerges from the B-theory though, here again, there is a certain reverse relation in that certain features of inter-molecular potentials can, to a certain extent, be inferred from the behavior of liquids.

The question arises as to whether and in what sense the A-theory can be said to be *reducible* to the B-theory. This is a question relating to the deep and complex relation between theories, analogous to the one of the complex inter-connections among our *concepts* — and is ultimately related to the complexity of reality itself. There is no easy or 'satisfactory' way of settling this question, as indeed there can never be.

All our theories arrived at in scientific investigations form a hugely complex system, analogous to the enormously complex webs of our beliefs and our concepts formed in the course of our accumulating experience in this world. Some of these theories are nested within others, some broadly imply others, some are in the nature of interpenetrating theories, while some are distant kin of others. In this complex web of theories that is an incessantly evolving one, there are structures on all scales, but no theory can be said to be the ultimate foundation of all others. Theories, in other words, are somewhat like words whose meanings are explained in a dictionary, where these meanings make sense only by *mutual* reference. For instance, the words 'ball', 'sphere', and 'round' can be found to occur in the entries for all the three. In other words, there is no *basic* or foundational set of words in terms of which all the others are explained. There may, however, exist certain groups such that, in each group, only a few words are mostly used to explain the meanings of the rest. This is because words correspond to concepts and categories as these are formed in the course of our experience, and experiences come piecemeal, having no systematic tree-like structure in them.

A theory is an abstraction from experience, with the latter pruned and idealized suitably so as to refer to a model — this happens with scant regard to the currently existing structure of the network representing all the accumulated theories in various different fields, domains, and subdomains.

Two questions that stand out are the following: *first*, if theories are mental constructs, then how come they are so successful, and *next*, why should all the theories in the various domains and sub-domains of experience not ultimately reduce to a single grand theory of the universe as a whole — some ultimate theory of fundamental particles and fields coupled with the theory of gravitation so that the small scale and large scale theories of the cosmos are accommodated within it? Related to these two questions is the one that asks whether the successive revisions to a theory lead us closer and closer to truth about some part of nature, with all these partial truths embedded in one single all-embracing *foundation* of all theories — the ultimate 'law of nature'?

None of these questions can ever be settled conclusively and to the satisfaction of all because these are related to the metaphysics and to the ontology one is prepared to accept. The way we see it in this book, one has to take as guide the lessons grasped from our experience of complex systems.

A complex system involves levels and layers nested within it, and can only be experienced partially — in bits and pieces. The question of looking at the behavior of a complex system *in its entirety* is an abstract one because of the infinitely extended web of links to other systems equally complex, and to systems at levels located higher and lower in the hierarchy — indeed, even the idea of a single hierarchy is an abstract one, since all the 'hierarchies' are tangled together in this world of ours.

While we have focused here on the relation between emergent properties and emergent theories, the idea of emergent properties has led to questions being asked as to whether it is a philosophically and logically sound one. Scientists and philosophers subscribing to the viewpoint of reductionism complain that ‘emergence’ has a mystical aura about it that does not bode well for either science or philosophy. In this essay we adopt the position that one needs to have a better understanding of emergence from the scientific point of view — how emergence is related to co-evolution and to the appearance of stable and unstable structures on all scales in a complex system — before a more meaningful philosophical discourse can be engaged in. For background, we suggest [101], and [33], along with references cited therein.

> In this context, the following lines from Crick, quoted in [33], may be of some relevance:
>
> ”There are two meanings of the term emergent. The first has mystical overtones. It implies that the emergent behavior cannot in any way, even in principle, be understood as the combined behavior of its separate parts. I find it difficult to relate to this type of thinking. The scientific meaning of emergent, or at least the one I use, assumes that, while the whole may not be the simple sum of the separate parts, its behavior can, at least in principle, be understood from the nature and behavior of its parts plus the knowledge of how all these parts interact.” [29]
>
> However, this passage from Crick notwithstanding, we submit that, as of now, the deep link between emergence and complexity is not sufficiently well understood (on scientific terms, that is) to make ‘in-principle’ statements as enlightening or meaningful as they should otherwise be.

4.3.8.3 Successive revisions of theories

Theories are arrived at inductively by abstracting from experience, and are revised as and when they fail the test of observations accumulating subsequently.

On the basis of clinical observations and pathological tests, my family physician diagnosed that my son was having a certain problem with his blood circulation (a ‘theory’). *Treatment prescribed by him produced early results, indicating the correctness of his diagnosis. However, there was a relapse of earlier symptoms and a more serious symptom started showing. The doctor then patiently went through his history once again and came up with a completely new diagnosis* (the revised theory).*A new course of treatment rapidly cured my son.*

An inductive inference is, in principle, *defeasible* — the conclusions get modified under new evidence, as in the case of the medical diagnosis of my son's ailment. And, the modification need not be 'small' in any sense. As one arrives at a new inference in the place of an old one, the latter may differ quite markedly from the former. An alternative way of saying this is that an inductive inference is *underdetermined* by evidence — alternative choices are possible on the basis of one and the *same* evidence, and an added evidence may tilt the balance away from an earlier one to a novel and more justified alternative. But the justification is never complete.

Early on in my life I chose the career of a teacher — unfolding circumstances propelled me to a research career; and then, finally, I left that too so as to adopt the life devoted to social work — all this while choosing from among alternatives that life presented to me, and finding one choice better than an earlier one under added experience gained in my life's journey.

But perhaps hard scientific evidence and well-tested theories are not like these chaotically changing choices so common in our social experience? Surely, the fact that the gyromagnetic ratio of the electron is known correct to twelve decimal places makes it an essentially accurate conclusion of the current quantum field theory, possibly open to only extremely small corrections under the impact of further evidence and further revision of theory?

This book does not claim authority to comment upon the current state of affairs with the standard model that has successfully explained and predicted a wide range of phenomena at the sub-atomic level, and on the efforts under way to patch up its loopholes, but one does feel inclined to say that the 'loopholes' appear to be gaping ones (see [149] for a detailed and serious assessment, one that is not too technical).

Here, to put things into perspective, we will point to the spectacular accuracy with which Newtonian mechanics, along with Newton's law of gravitation, explains and predicts the orbits of gravitating bodies that makes space travel possible. And we will also point out that Newtonian mechanics, including the relativistic corrections and possibly also the corrections due to the space-time curvature caused by the sun's gravitational field is a marvel of a theory, but only *within a context*. And *that* context differs equally spectacularly from the one in which the collision data from the large hadron collider at CERN assume relevance. In other words, the stupendous success of either of the two theories (the Newtonian theory with appropriate corrections, and the Standard model) notwithstanding, both are contextual and are arrived at independently of each other (in a manner of speaking, that is) — the infinitely complex reality existing out there has space enough to accommodate both and, who knows, many many more. At another level of reality, Newtonian mechanics is replaced with the general theory of relativity in dealing with the behavior of cosmic bodies where the latter, initially thought to lead to results in the nature of 'small' corrections, was subsequently found to generate astoundingly novel consequences.

We will not speculate on whether a possible future theory will connect up gravitation with the standard model not only because I have no competence for such a thing, but no less pertinently because the job that I have set for myself in this book is the much more modest one to see how the lessons learned from our experience with complex systems help us on this issue of inter-theory relations and theory revision.

The accuracy of the predictions of a scientific theory is contextual, and so, in a sense, is the explanatory power. The gyromagnetic ratio of the electron arising from its spin was found to be 'anomalous' (as compared with the value of the ratio arising from 'orbital' motions of the electron) in connection with spectral characteristics of atoms, but it could not be determined with very great accuracy within the context set by the Dirac equation. A remarkable improvement was made possible within the context of quantum field theory. The important thing to note is that the improvement in accuracy was, in some sense, 'small' (depending on what is referred to as the fine structure constant) so that the succeeding theories (quantum electrodynamics, and then, the standard model) could conceivably be interpreted as a 'small' correction over the preceding one (relativistic quantum mechanics), but the frameworks of the earlier and the later theories differed spectacularly. In other words, the 'small' difference in the value of the gyromagnetic ratio was symptomatic of a very big structural revision in the theory. Quite often, a succeeding theory, in addition to being responsible for small corrections in predicted values, unearths altogether new phenomena as in the case of quantum field theory predicting and explaining the existence of new sub-nuclear particles. In that sense, then, not only the terms of reference of the new theory, but its predictions too differ to a very large extent from those of the earlier one.

This once again raises the question as to whether successive revisions of theory can be considered to be in the nature of 'small corrections', indicating a convergence to some final theory describing reality.

> For instance, considering the case of the gyromagnetic ratio of the electron once again, a possible revision of the standard model in the future will most likely engender only a minute correction to this quantity, but what is more likely is that the emerging theory will have a freshly added layer of meaning to concepts that will be inaccessible from the existing version, though it will be possible to interpret the latter from the point of view of the revised theory. It may even be that the term gyromagnetic ratio itself will lose its current relevance and will turn out to be the limiting case of some new conceptual entity (or may get related to components of some higher dimensional quantity) whose significance will be apparent only in the new theoretical scheme: the latter will be *incommensurate* with the current theory.

> There is, in all probability, a complex substratum underlying the currently accessed microscopic world that the standard model does not visualize or address. In this sense, the standard model, along with all its results, can be said to be a 'simple' one — an idealization. If and when the substratum gets opened up, theory is likely to acquire a complex texture, the standard model will have lost its simplicity, and the gyromagnetic ratio, while acquiring only a minute correction, will in all probability come out as a projection of something more complex.

This is a question that cannot be settled one way or the other to the satisfaction of everybody since the answer depends on the meanings that one attaches to phrases like 'small correction', 'convergence', and so on, and these meanings, in turn, depend on the metaphysics that one has in mind such as theories being the reflection of an underlying 'regularity' and 'harmony' of nature. As for me, I do not find myself impressed by the idea of regularities and harmony buried deep within the bosom of nature. The idea of regularities and harmony is specific to our thinking mind which is always interpreting, sorting, always formulating 'simple' rules for our survival and onward journey in life; and, what is more, the idea is specific to the context in which reality presents itself to us. What can indeed be considered as 'ultimate' in nature is, precisely, its complexity — a complexity that generates islands of regularity within itself that we get a hold on, but ones that cannot be said to be indicative of 'intrinsic harmony' of Nature.

The only safe extrapolation — if there could be one — from our experienced phenomenal reality to the 'real' reality out there is the one of complexity, of an immense spectrum of correlations among its constituents when looked at within any given context, where the term 'context' is now used to mean a given level of self-organization of the infinitely tangled system that we refer to as the noumenal reality.

As mentioned earlier, reality, as a complex system, is a *co-evolving* one, where the entities it is made up of keep changing, their interactions keep changing, and their levels of self-organization keep changing. What we can observe of this reality is, by its very nature, some chunk of it within a limited horizon of space and time, however vast and varied that may appear to us. As we focus on some part of reality, we seem to zero in on some regularity inherent in it and may have the feeling that our theories, in the course of successive revisions, are approaching the point of correctly capturing that regularity. But on actually approaching that point, the convergence seems to dissolve in thin air and 'divergence' raises its ugly head. This happens because, with accumulating experience, *a new context opens up.*

Within the confines of purely combinatorial considerations, one can refer to *Ramsey theory* ([46], chapter 4), results in which imply that within every sufficiently large structure, there has to exist a regular or ordered substructure. This, admittedly, is a vague and incomplete paraphrasing in a subject that has attracted attention of great mathematical minds and has had interesting applications, but will have to suffice for our present purpose.

In the present section we have focused on the dynamical evolution of networks, whose nodes (or vertices) denote systems that interact with one another, as represented by the links (or edges) in it. As we have mentioned, a network representing a real-life system is generally a multi-layered one and undergoes co-evolution. In this process of co-evolution that is likely to have disordered and ordered aspects built into it, the network passes through a succession of structures where, looking at the structure at any particular stage of the process, one can find 'islands' of regularity, in tune with results in the Ramsey theory.

Turning our attention to the infinitely extended and infinitely complex noumenal reality, one imagines that our scientific theories capture the order and harmony built into these islands of regularity within a vast sea of

> complexity. Evidently, there is nothing to guarantee that the order inherent in these islands of regularity can be extrapolated to nature as a whole.

Our experience of reality always occurs within the constraint of certain borderlines that are, in a sense, objective ones — objective, that is, with reference to the current context in which science can access the universe in space and time. For instance, even the hugely successful standard model of fundamental particles and their interactions works within the context set by what is referred to as the *Planck scale*. The science of fundamental particles and their interactions, along with the theory of the 'early universe' ignores all inhomogeneities and structures of the noumenal world down to the atomic scale and strives to access even smaller distance and time scales by using highly energetic particles as probes, investing fabulous amounts of resources into the job. The idea is to explore the possibilities of a theory that fits with the standard model at the scales of length and energy currently accessible and, at the same time, makes it complete by weaving gravitation seamlessly into its fabric. Whether and to what extent that effort is going to meet with success is anybody's guess. Meanwhile, the Planck scale sets the context of the standard model — what lies on the other side of it can only be conjectured. This little book of ours is meant to make a statement that the complexity of the real world has to be reckoned with in setting our mind on what to expect and how to direct our efforts on this issue of extending the standard model.

In stating that the context to our theories are not arbitrarily chosen by us but are set by objectively determined limiting boundaries, what one means is that these boundaries depend on the part of reality the theories try to probe and access, and on the current state of organization of that reality. It is in this sense that the limits within which the standard model is expected to work can be said to be related to the Planck scale because that sets the context within which gravitation can be included in the theory as a *classical* field, independently of the quantum mechanical interactions among particles — interactions of the electro-weak and the strong variety.

Instances abound where a theory gets modified to a 'large' extent as some objectively set boundary or other is crossed and discrepancies with observed facts emerge that may initially appear to be be 'small' but eventually turn out to be 'large' ones. Thus, even where a small but persistent discrepancy makes necessary the modification of a theory, one eventually finds that the terrain on the other side of the objectively existing boundary setting the context of the earlier theory is replete with phenomena quite out of the range of capabilities of that earlier theory. In this sense, the revision of an existing theory can be said to be substantial or extensive in respect of both the framework of the theory and its terms of reference *and* the concrete predictions of the theory.

> *Theory revision*, which can be looked upon as a particular case of belief revision in general, is a fundamental concern of scientific realism since the ontological status of theories is a debated issue within its fold. In particular, the issue of incommensurability in respect of successive revisions of theory which, in turn, is related to the question of an ultimate theory of nature, has been referred to several times in this book, and will be taken up again in sec. 4.4

This is what appears to have been the case of the general theory of relativity as it emerged as a revision of the Newtonian theory of gravitation. Even as the terms of reference of the revised theory differ markedly over those of the earlier one and entirely new conceptual ingredients are introduced, the predictions of the two theories differ to only a small extent over relatively small scales of space and time. However, when looked at over larger scales and in the presence of gravitating bodies of relatively large mass (one can attach quite specific meaning to the terms 'small' and 'large' here though we will not enter into it), the general theory makes predictions that differ spectacularly from those of the Newtonian theory, in keeping with the remarkable difference in the conceptual framework of the two.

The relation between an earlier theory and the revised one is *asymmetrical* — referring to the border separating the two theories (the one corresponding to small velocities and a small strength of the gravitational field in the case of Newton's theory of gravitation), the predictions from the revised theory approach those from the earlier theory as one approaches the border in some limiting sense (where the former reproduce the latter along with small correction terms), but the converse does not hold. As one other instance of such a relation between an earlier and a succeeding theory, one can refer to the motion of a particle approaching a sufficiently high 'potential barrier' as described in the classical and quantum theories. In the classical theory, the particle fails to propagate to the other side of the barrier, while in the quantum theoretic description the particle 'tunnels' to the other side, though the probability of tunneling decreases exponentially as the height and width of the barrier become large in comparison with its energy in some well-defined limiting sense (the same limit corresponds to the Planck constant going to zero). In this case, the two theories differ spectacularly in their conceptual framework, and an asymmetry is quite manifest: the predictions of the quantum mechanical theory makes it possible to understand and interpret those of the classical theory close to the limiting situation mentioned above, but there is no way the classical results can be used to interpret the quantum mechanical ones in an analogous manner.

This asymmetrical relation finds expression in certain special features of the predictions of the revised theory close to the border setting the two theories apart, since the border is seldom a sharp one. These special features can be described in mathematical terms in the case of theories in the physical sciences. As one moves across the border, or approaches it on one side, quantitative prediction can be expressed in terms of an *asymptotic series* instead of a *convergent series* commonly encountered in mathematical approximation schemes. This corresponds to the fact that the relation between theories can often be described in terms of *singular limits*.

4.3.9 Digression: Asymptotic Series and Singular limits

4.3.9.1 Asymptotic series

A *convergent series* is one where one can unambiguously attach a meaning to the idea of 'summing up' an infinite number of terms. In principle, one can perform a term-by-term addition to obtain successive *partial sums* of the series, which approach as close as one wishes to a fixed number — the *sum* of the infinite series in question. Each partial sum differs from the sum of the series by an 'error term' that gets smaller and smaller as successive terms of the series are added up.

Innumerable examples exist of such convergent series representing mathematical and physical quantities of interest. One such object is the number 'pi' (π), the ratio of the circumference and the diameter of a circle. In decimal terms it is approximated by 3.14159265, but this value differs from the actual value of π by a small error term — the error never vanishes even when one fills up a large number of decimal places. There exist several convergent expansions where successive partial sums approach π at a rapid rate.

Convergent series are useful not only to represent numbers but *functions* as well. Thus, a function $f(z)$ depending on the variable z (commonly one taking up complex values of the form $a + ib$, where a, b are real numbers) can be represented by a convergent series for every specified value of z within some specified domain.

Contrasting with the case of convergent series, there exist examples of infinite series — of great relevance in mathematics and the physical sciences — that are endowed with *contrary* significance. Such a series, referred to as an *asymptotic series*, can be used to approximate a function with great accuracy but is typically a *divergent* one. Thus, a series of the form $(a_0 + a_1 z + a_2 z^2 + \cdots + a_N z^N + \cdots)$ can be used to approximate a function $f(z)$ at a point z in some neighborhood of any given point, say, $z = 0$, by evaluating the partial sum up to an optimum order $N = N(z)$ (where it is possible to estimate $N(z)$ quite accurately), but on evaluating the successive partial sums beyond $N(z)$ one finds the series to diverge. Early exponents of the power and potentiality of asymptotic series were George Stokes and Henri Poincare among others, who reinstated these divergent series in the road map of mainstream mathematics and physics following a phase when these were banished from respectable research programs.

4.3.9.2 Singular limits

The noted mathematician-physicist Michael Berry illustrated the idea underlying a singular limit by means of the following interesting observation, made in a light spirit: half the bodily remains ($\delta = \frac{1}{2}$) of a worm discovered in an apple after a big bite is more revealing (and revolting too) than a full worm ($\delta = 1$) since it indicates that the other half is now residing in your digestive tract; by the same token, say one-tenth of the remains ($\delta = \frac{1}{10}$) is even more revolting, and so on, till you discover to your delight that one of the apples in the lot does not reveal a worm ($\delta = 0$) even after several bites, because that indicates that the apple is *worm-free* (discounting the other appalling possibility). Here $\delta = 0$ is a *singular limit* since something entirely novel emerges in this limit as compared to small values of δ, close to it.

Other well-known examples of the phenomenon of singular limits in physics are: the limit of the viscosity of a liquid going to zero (no turbulence in the singular limit), the limit of wavelength of light going to zero (in relation to the size of an obstacle; no interference and no diffraction fringe), the Planck constant going to zero (in relation to the size of a typical action integral; classical mechanics: no tunneling through a potential barrier, no explanation for the hydrogen spectrum, ... no nothing).

Berry and a number of other mathematicians and physicists (see, for instance, [13], [24]) have worked on what a theory looks like *close* to a singular limit because the limit itself is not smooth, and it is of great interest to know what transpires close to the limit ($\delta \gtrapprox 0$) as against the situations corresponding to $\delta = 0$ and δ substantially away from zero. This sheds much light on what is referred to as *theory reduction* — a singular limit corresponds to some limiting value of a relevant parameter (denoted by δ here), close to which a theory assumes a complex form. The complexity, originally hinted at by Stokes, melts away as δ takes up the value zero and also as delta moves substantially away from zero where, however, the theory is of a notably different structure.

More generally, singular limits illuminate the *transition* between different *levels* of reality — they tell us how the levels differ 'qualitatively' and yet can be understood in terms of the continuous variation of a single parameter δ (or of a number of parameters). They tell us that the qualitative difference is the result of a certain 'violent' behavior ('Stokes phenomenon') close to the limit — a 'violence' that can nevertheless be understood in terms of the smooth variation of a single parameter. What is more, this 'violence' can typically be related to the appearance of an *asymptotic series* (sec. 4.3.9.1) describing some typical physical prediction of the theory.

4.3.10 The truth of theories

At this point, we look at the idea of *truth* inherent in a theory. The concept of truth is a vexed one. Even within the rigorous domain of mathematical logic, it is difficult business arriving at a precise formulation of what is meant by truth. This is achieved within the framework of the *correspondence theory* of truth built up, among others, by Alfred Tarski.

Truth is a property characterizing a statement (a 'sentence' in some formal language) but is semantic in nature. In other words the truth of a statement says something about the state of affairs in some 'universe' of discourse, i.e., some *set* in the context of mathematical logic. Tarski derived the definition of truth from that of 'satisfaction'. However, instead of following the rigorous logical route to the concept of truth, we follow the broad outlines of the formal approach (see, for instance, [83]) and adopt the position that statements derived in a theory (ones that can be taken to constitute a 'language') can be true if they 'correspond' to some state of affairs in a universe, where the 'universe' may mean our phenomenal world or some part of it that one may choose. The term 'model' as used above in respect of theories of various parts of reality is, in a broad sense, analogous to what is referred to as a model for a set of sentences of a formal language: a model (relative to a scientific theory) is some part of the phenomenal world defined contextually in which the statements derived from the theory turn out to be true. Here the term 'truth' means that a state of affairs in the model, *corresponding* to some particular statement derived from the theory happens to hold in the model. For instance, a system made up of a specified number of particles (with no charge and with zero 'lepton number — whatever that term means), imagined to be isolated from the rest of the universe, qualifies as a model for the theory based on classical mechanics along with Newton's law of gravitation, but is not a model for the theory of electro-weak interactions.

An important observation on the semantic theory is that the truth of a statement acquires meaning only when it refers to some state of affairs *within* the universe under consideration, i.e., within our phenomenal world in the present context. However, a statement *about* that universe as a whole is not admissible as one whose truth can be ascertained. This is a stricture that prevents the *liar paradox* and other anomalies from vitiating the idea of truth as outlined in the semantic theory (also referred to as the correspondence theory).

However, it is not only the matter of a formal logical paradox that stands in the way of ascertaining that the entire phenomenal world of ours is governed by some 'ultimate' law, though it is certainly food for thought as to whether it is meaningful to talk of the truth of a statement pertaining to 'nature at large' since our world as a whole is not embedded in a larger world where we are located as observers. In other words, in stating that there is an ultimate law of nature, we repeatedly come across an impasse where we need the so-called God's-eye or God's-design point of view to bail us out.

A more earthly option to adopt is to accept that there is no such thing as an ultimate law of nature — a point of view arrived at on the basis of lessons drawn from our experience. This is the experience that tells us that theories are constructed piecemeal and apply to models within specific domains of experience, a model being defined with reference to some part of our phenomenal reality delimited by means of a context that may be made explicit or else left implied. That same experience tells us that when our range of exploration and observation gets extended across certain objectively existing boundaries, a theory gets revised so as to explain anomalies and to accommodate new phenomena, and a broader theory emerges covering a newly emerging domain within our phenomenal reality. Of the two theories, the one arising in the process of revision is a broader one in that one can interpret certain features of the previously existing theory within its framework, but the converse relation does not hold. Further, the two theories have incommensurate features, and do not conform to the picture of a monotonic progression towards an all-embracing theory, valid across domains to the entire phenomenal universe.

1. Further considerations on the issue of theory revision are to be found in sections 4.3.8.3, 4.3.9.2 and 4.4. In addition, refer to sec. 4.5, where theory revision is considered with reference to a restructuring of our *conceptual space*.
2. Consider a theory 'A' revised to a new broader theory 'B' in terms of which theory 'A' can be understood

in a limiting sense close to the border separating the two. It is only close to the border that the relation between 'A' and 'B' acquires meaning. The theory 'B', in turn, may be revised to a theory 'C' across a new border separating the two. In this case, the relation between 'A' and 'C' need not be a significant one, since 'C' cannot be interpreted in relation to 'A' in the same sense in which 'B' can be. Put differently, the relation between 'A' and 'B' cannot be extrapolated to pairs of theories separated from each other by several successive revisions. Stated simply, such extrapolation would be an exercise in naivety.

There is a major trend in scientific realism that asserts that theories make true statements about reality, and in order to make this assertion compatible with the idea of a progression towards an overarching theory, there has emerged a trend to replace the idea of 'truth' by one of 'truthlikeness' or 'verisimilitude' as it is referred to. Apart from the question of how consistently one can formulate the idea of verisimilitude ([115], [5]), one has also to recognize that, instead of a monotonic convergence, successive waves of theory building are likely to reflect the behavior of an asymptotic series that is found to converge only up to a point, after which it starts diverging. Past the point of divergence, an emergent theory can again converge to observed features of phenomena in an expanded domain only to diverge once again as some new boundary is crossed. It is this picture of non-monotonic and incommensurate behavior of successively revised theories that may make redundant the efforts at consistently replacing the idea of truth of a theory with that of verisimilitude.

Finally, truth is related with the question of *theory choice*. In contrast to the formal theory of truth, truth in real life is conditional on its *acceptance* by human society at large. This may sound paradoxical since truth of a statement is commonly supposed to depend on objectively existing 'state of affairs' in the world. But recall that the world we are speaking of is the phenomenal world where a statement pointing to an objectively existing state of affairs is, all said and done, a matter of interpretation. And, in this complex world of ours, one cannot simplify things by just saying that the interpretation concerned is the interpretation 'of mankind' and acceptance of truth of a statement is acceptance by 'mankind'. This brings us to the domain of social reality where science is engaged in a complex interaction with human society. In mathematics, everything is formalized so that nothing is left to vagaries of human psychology (or, is it? — but we will let that go) and acceptance of truth is reduced to *proof*. Major logical systems are *complete* in the sense that one can produce proofs for all true statements. Real life is much more messy and innumerable conflicts among men prevent a universal yardstick for the judgment of truth — as a result of which elaborate legal systems have come into being. And even then truth remains elusive in the end.

In scientific exploration, the situation is somewhere in between where, broadly speaking, peer-reviewed journals constitute a system of establishing the truth of theories. In reality, the system is tolerably 'objective' so far as the acceptance of a major portion of the totality of all scientific contributions is concerned, but exhibits gaping loopholes when it comes to instances of *theory revision*. This is precisely because of the *incommensurability* inherent in successive theories that arise in what have been referred to as *conceptual* revolutions. As a result of the incommensurability, successive theories cannot be compared in their totality in terms of reference common to both since the succeeding theory contains new conceptual ingredients that the preceding theory lacks. It is here that the acceptance of theories depends to a major extent on *points of view* of various different groups of scientists and even among larger social groups. This is why the idea of incommensurability is, at times, branded as *social constructivism* — a point of view that is supposed to be inimical to scientific realism.

However, as we see it, scientific realism, in order to be consistent, has to make room for the idea of incommensurability. Looking at the case of what Thomas Kuhn calls a scientific revolution ([80]), conflicting points of view that can never disappear from human society can impede the acceptance of a theory, though only up to a point since there always exist a set of referents in the two theories close to the border (refer back to sec. 4.3.9.2) separating the two that can be subjected to reality check.

> The fact that mathematical truth pertains to some model that constitutes the universe to which mathematical statements apply, and that the truth of theories in science are also meaningful when referred to models of various parts of reality, helps us understand the 'unreasonable' effectiveness of mathematics in the natural sciences [145]. For instance, the quantum mechanical states in an atomic model may make up a set to which the language of linear vector spaces (including the language of group theory) is applicable in the sense of a mathematical model. In this sense, the evolution of mathematical concepts and theories and the development

of theories in the sciences converge upon models to which both the two apply.

4.3.11 The brain and the mind: a vast sea of complexity

4.3.11.1 Brain and mind: introduction

The human mind is an infinite universe in itself. It is an impossible task even to describe what the mind is and what it is capable of. The complexity of the mind is related to the enormous complexity of the neuronal connections in the human brain, which operates within a no less complex biological system, including a host of chemical messengers, that modulates the actions of the neuronal circuits. The relation between the brain and the mind is itself an intriguing one since the mind has no materiality of its own. The operations of the mind (more precisely, of the self-system lodged in the mind) define and constitute the psychology of an individual, which is an *emergent* phenomenon generated by the electrical and chemical activities in the brain, modulated by a number of associated physiological systems. Psychology, in turn, involving a vast array of ingredients, states, and processes, leads to an endless variety of human behavior spanning an almost infinitely wide spectrum.

The term psychological *ingredients*, or 'resources', refers to such things as concepts, beliefs, preferences, emotions, memories, hopes, aspirations, and an infinity of other items involved in making up the psychological *states* of an individual. Psychological *processes* pertain to ceaseless changes in the mental states by means of which the mind undergoes a stupendously complex evolution in the lifetime of a person. Incidentally, the terms 'mental' and 'psychological' will, at times, be used interchangeably though, in reality, the latter constitutes only a part of the former. And, further, 'states', dispositions, and processes of the mind, all correspond to spatio-temporal excitation patterns in neuronal assemblies, characterized by time-scales varying over an enormous spectrum.

In this book we adopt the position that mental ingredients, states, and processes are based on activities of *large scale neuronal aggregates* — mostly in the brain — interacting with one another by means of neuronal circuits and pathways, where these interactions are modulated by systems of chemical transmitters embedded within a larger biological system and a still larger *environment*.

We, moreover, adopt the view that the mind is an *essential* intermediate level in understanding human behavior and is not simply an identifiable derivative of neural activity that can, in principle, be discarded in the description and explanation of behavior. This is what is meant by saying that the mind and its endlessly complex spectrum of processes constitute *emergent properties* of neuronal interactions. In particular, in spite of being devoid of materiality, the mind is nevertheless as real as the neuronal assembly itself.

We will be mostly concerned here with the complexity of the neuronal organization in the brain while referring more briefly and informally to the complexities of the mind. In this, the starting point will be the assumption that the mental ingredients like concepts, beliefs, and preferences are based on *dynamic excitation patterns* in neuronal assemblies and these dynamic patterns themselves evolve in time as, for instance, our concepts and beliefs keep on changing in our lifetime. Thus, a concept is itself made up by association between a number of other concepts (forming a constituent of a vast web of mutually related concepts) and corresponds to a dynamic excitation pattern formed as a composition of the patterns making up these other concepts where, however, the 'correspondence' is a strange and complex one that cannot be described precisely with any degree of completeness. The excitation patterns are set up, depending on their nature, in different neuronal assemblies, and patterns in distinct assemblies are integrated into a single whole depending on the functional relations between these. Ultimately, all the excitation patterns in all the distinct but interrelated neural assemblies appear as restrictions (to the respective networks) of a single but pervasive dynamic pattern in the entire brain (where we keep implied the role of chemical transmitters in modulating the activities of the various assemblies of neurons).

A psychological state is assumed to result from patterns in some subset of neuronal assemblies, and a mental or psychological process corresponds to an evolution in time of these constituent dynamic patterns. However, as mention above, the question of the correspondence between neuronal circuits and psychological ingredients or states remains, in the very nature of things, as one with no final or definitive answer.

1. We repeat that, strictly speaking, psychological processes constitute a sub-class of mental ones, being related to the self of an individual.
2. A class of mental processes of great relevance is made up of those referred to as *cognition* — ones relating to memory storage and recall, learning, decision making, inference making, and *creativity*, among others.

4.3.11.2 The complexity of neuronal assemblies

Neuronal assemblies are organized *at all scales* in the brain. Such a wide spectrum of scales is characteristic of complex systems in general where the emergent structures (both in space *and* in time) are symptomatic of the feature of self-organized complexity (refer back to sections 4.1 and 4.3 in the present chapter). Such structures can be inferred from the spectral features of electrical and magnetic signals picked up at various points of the cranium under diverse experimental set-ups and also from the direct time-series data and their correlations (refer to [126] for details on this and related issues).

The several billions of neurons in the brain are assembled in distinct structures ranging from small to large ones, with interactions of various strengths among the neurons belonging to any particular structure and also among those located in distinct structures. In addition to the structural complexity of the neuronal assemblies, there are *functional* relations of a complex nature among various assemblies. Thus, looking at some particular structure, it may be found to participate in a large number of functional processes (generation of affect, creation of a new belief, recalling an experience from memory, and so on) by interacting with other structures, at times even remote ones in the brain. Generally speaking, the functional correlations among neuronal assemblies do not explicitly involve causal dependencies — the latter define a large sub-class of interactions relating to planned, purposeful and causality-based actions of the mind.

The strength and nature of the interaction or correlation between two neuronal assemblies or between members of a group of such assemblies can in principle be expressed in terms of a mathematical indicator referred to as the *relative entropy* ([126]) that can be related to time-series of data obtained from electrical and magnetic signals collected from various different regions of the brain. The dynamics of signal propagation along the axon of a neuron is by now well studied by means of mathematical equations resembling the propagation of voltage signals along electrical cables with specified physical characteristics (the *Hodgkin-Huxley* equations, for instance). Starting from basic models of this type and from models of synaptic interactions, one can simulate the dynamics in larger and larger neuronal assemblies by means of computer simulation and then compare the results with time-series data mentioned above. The relative entropy and other related entropy measures provide information on the complexities in and across various structures of the brain.

Even the propagation of signals along a single neuron may be possessed of complex features involving chaos (see, for instance, [62], [96]))

In keeping with what has been mentioned in earlier sections of this chapter, there is no single and well-defined measure of complexity in real-life systems, though there exist a number of qualitatively defined features common to numerous systems that are supposed to constitute instances of complex ones — the neuronal organization in the brain being one of the most commonly referred among these. According to Sporns ([126], chapter 13), a major feature underlying the complexity of the action of the neuronal organization is the coexistence of 'segregation and integration', 'enabled by the small-world modular architecture of brain networks'. What is of relevance in this context is the element of *conflict* between the two, since the segregation of the networks tends to make those act independently of one another, while their integration makes them functionally dependent. There are myriads of possible modes of integration between a large number of components approximately meeting the criterion of decomposability, more so when the components interact by means of a wide spectrum of strength and types of interaction.

An instance of complexity generated by interactions (distributed over a range in accordance with a given probability density) among a large number of components is provided in condensed matter theory by a *spin glass* system, in which the interaction between any two components, mediated through distinct loops of intermediaries, is characterized by the feature of *frustration* [130], i.e., a conflict that may or may not be a pronounced one, depending on the loops considered.

A foundational position to adopt is that these myriads of modes of action of the system of brain networks constitutes the basis — in a manner that is likely to prove undecipherable — for the equally innumerable variety of mental states and processes.

But the above is only one of several distinct (but related) aspects contributing to and enhancing the complexity of neuronal activity and of mental states and processes. One other aspect of great relevance relates to the role of *remote interactions*, typical of complex systems. In a system made up of a large number of components, a certain function or behavior is often explicable to a good degree of accuracy by taking into account a certain subset of components and disregarding the large number of remaining ones that can be assumed to be remote in the context of the function under consideration. But such an assumption is *unlikely to remain valid over all time scales*. Accumulating influences from remote interactions are inevitable so as to cause the emergence of new behavior in a major way. These accumulating influences cause instabilities of various types, symptomatic of *sensitive dependence on initial conditions*.

The same sensitivity marks the dynamics of a complex system as a function of *changing context*. As the environment of a complex system gets modified even to a small extent (due to its own evolution as a complex system in its own right), instabilities analogous to those arising from small changes in initial conditions get into the act and lead to the emergence of distinct behavior patterns.

However, the said instabilities are only of a *local* nature, resulting from positive expansion rates along only a few directions in the phase space of the system under consideration. The spectrum of expansion rates ('Lyapunov exponents') ([51]) of a complex system involves both positive and negative exponents, capable of causing both *amplification* and *suppression* of perturbations.

In the case of neuronal assemblies in the brain, the amplification and suppression of perturbations arise as a result of *feedback* loops, generating *re-entrant* signals. Another major aspect of the role of such signals relates to the ability of the mind to *map its own activities*.

In other words, the instability is generically confined to only certain subspaces of the phase space, the dynamics in the other subspaces remaining substantially unchanged (coexistence of sensitivity and stability).

> Refer back to sec. 1.16, where we introduced the useful analogy of looking at mental processes as dynamical evolution in the 'phase space' of the mind, the phase space being of an effectively infinite number of dimensions. Specific thought processes can be said to be confined to low-dimensional 'sub-spaces' of this phase space.

Further, even the diverging perturbations get suppressed as the deviations from the original dynamics become relatively large since the local expansion rates vary from one region to another in the phase space. In brief, the dynamics of a complex system generally remain globally stable, switching from one mode of behavior to another without systemic breakdown.

When one speaks of the context in which a complex system is observed and analyzed, it is not only the environment (in which the system is embedded) that is relevant, since the context also includes those remote components of the system that are necessarily left out as being of negligible significance. As mentioned above, neuronal assemblies in the brain, characterized by the small-world phenomenon (refer back to sec. 4.2.5.1 for a brief outline), are interdependent in such a way that, during the unfolding of some particular pattern of activity, only a few assemblies interact to an appreciable extent while other assemblies, not initially involved, become relevant eventually so as to cause a switch-over to a distinct pattern by means of local instabilities mentioned above.

Finally, the activities of neuronal assemblies in the brain are *history-dependent* (refer back to sec. 4.2.6). An obvious source of history dependence is, of course, the activity of neuronal assemblies responsible for *memory* (refer back to sec. 2.6). All present experience is colored by past memory, and the multifarious ways that this coloring can be effected lends infinite variety and complexity to our mental life. However, one can look at this from a broader point of view, based on the likely role of *dynamic excitation* *patterns* in neuronal assembles. The term ‘memory’ is commonly assumed to refer to persistence of past experience. However, whatever persists in the mind is memory in a broader sense. For instance, a belief of mine is memory in this broader sense, because it persists and can be made use of — though one does not ‘recall’ a belief in the same way as a memory is recalled. Similarly, a concept, or a preference or, say, the encoding of a set of basic emotions, all must likewise be implanted in neuronal assemblies in the form of dynamic excitation patterns. The activity of the brain is, in a broad sense, a ceaseless ebb and flow of such dynamic excitation patterns and, in a similarly broad sense, signals generated as inputs from the external world (or even from within the brain itself) interact with existing patterns in numerous neuronal assemblies generated at earlier times so as to give rise to new patterns in the course of time. Thus, in this broader sense, dynamic patterns generated in the past may be said to give rise to *structures* within the brain that cannot, strictly speaking, be distinguished from the ‘physical’ ones produced by means of neuronal pathways and circuits — indeed, looking at the process of development of the brain during the lifetime of an individual, one observes that the process of generation of brain structures is heavily experience-dependent (the phenomenon of neuronal *plasticity*; refer to [11] for background). This makes the phenomenon of brain activity an exquisitely complex and intractable one since it is determined by the myriads of incidents during the entire developmental history of an individual (occurring in both the inner and outer worlds of hers). The history-dependence of neuronal processes is especially significant in the case of self-linked ones since the self is a unique product of the developmental history of an individual.

4.3.12 The complexity of the mind

The psychological (or, more generally, the *mental*) ingredients, states and processes constitute the mental world of an individual which, once again, is a vast and fathomless one. Take, for instance, the huge array of *concepts* that keeps on growing during the lifetime of a person (indeed, the entire mental world keeps on growing in extent and complexity, discounting the processes of shrinkage at advanced age), where the set of concepts as a whole is, at times, referred to as a *conceptual space*, with mutual associations among the elements of this space (for instance, the concepts ‘sphere’, ‘round’, and 'ball' are mutually associated). Or again, consider the equally vast web of *beliefs* lodged in the mind of an individual. Beliefs are in the nature of correlations (of a certain specific kind) among concepts, and the great thing about beliefs is that they *need not be mutually consistent* since these are tied together by means of *emotions*. Since beliefs act in a large measure as guide to behavior, these lead to *conflicts* among the goals, preferences, and behavior patterns of the person concerned, and consequently, to an exquisite complexity in her mental world.

Other ingredients and processes in the mental world add to its complexity in a like manner. In this context, one observes that the mind has a conscious and a non-conscious layer, where the two are related to the activities of the underlying neuronal assemblies in a manner that may never be known to any degree of completeness. It appears that the operations of the non-conscious and the conscious minds are distinct in nature — though with analogous features — where, once again, an element of conflict enhances the variety and diversity of modes of activity of the mind.

> The non-conscious mind appears to operate mostly on principles analogous to PDP computation, where inputs are *associated* with one another in various and diverse combinations; the conscious mind, on the other hand, operates *sequentially*, analogous to the operation of logical principles. While the non-conscious mind primarily registers *regularities* in the world accessed by it, the conscious mind seeks out *rules* and establishes relations of *implication* among inputs received from the world. The combination of these two makes the operations of the mind infinitely flexible and adaptable.

Generally speaking, the sources of complexity of mental states and processes are analogous to those of the underlying neuronal organization and can be summarized as follows: (a) an enormously large number of components (concepts, beliefs, etc.) involved and a wide spectrum of interactions among these, (b) an element of conflict, coexisting with compatibility, in the interactions among the components, (c) the existence of features, corresponding to amplification and suppression of perturbations (or deviations), in the interactions, and (d) the pronounced history-dependence of processes, i.e., the dependence on prior states.

Among these, the amplifying and suppressing (or moderating) influences in mental processes owe their origin to *emotions*. The emotions lend great variety to psychological processes and make them exquisitely sensitive to small perturbations and also to small changes in the context in which the processes occur. In addition, there often arises a large or small degree of conflict among the emotions, faced with which the mind seeks out a 'resolution' in accordance with principles of its own.

Considering all this and *summarizing*, the human mind is a vast ocean of fathomless complexity, generating an infinite diversity and variability of behavior that is dependent to a large extent on the developmental history of an individual, and is extremely sensitive to context. Above all, the mind has the capacity of *self-reflection* that engenders untold capacity for transcendence from behavior captive to constraining circumstances. This complexity of the mind is emergent from an underlying complexity in the neuronal organization of the brain — the latter in turn, operates within a complex environment provided by the physiological systems of the body and by the social environment of an individual.

4.4 Summing up: complexity in reality

> In this chapter dwelling on complexity and reality, we have traveled far and wide, in order to see how our conception of reality is shaped by its complexity. We have at times been somewhat desultory and at times repetitive, partly because complexity is a messy thing — there is no neat, cut and dried account of how it operates. It is not inherently beautiful, elegant, or harmonious, though there are islands of 'simplicity', 'beauty', and 'harmony' in it. In one's attempt at interpreting reality, one often has the satisfaction of getting to converge on to what appears to us ro be an elegant and beautiful theory when an extrapolation appears to be in order, extending the terrain ruled by the same simplicity, elegance, and harmony, and then, at some stage of extrapolation, one is met with blatant divergence. This then calls for a renewed hunt for elegance and harmony in a new terrain. This, as we see it, is the best that we can expect of science in its attempt at understanding and explaining reality — an infinitely extending mass of complexity that it is.

Even as we have offered partial summary of our wandering discourse on several occasions earlier in this chapter (see, in particular, sections 4.3.1, 4.3.2.3, 4.3.4, 4.3.5, and 4.3.7), we find it expedient to mention a few salient points of it before we finally call a halt to this part of our considerations, adding a few explanatory remarks on the way. In winding up, we'll go back to the question of 'incommensurability' and to the one of a 'final theory' because these are where we feel that scientific realism continues to remain kind of hesitant and undecided.

1. Scientific realism is based on the idea that there exists a mind-independent reality (the 'noumenal' reality as we have called it) that is an infinitely complex system and is comprised of entities and their correlations — the two making up an inseparable whole that can be separated only notionally. Various parts of this reality send out signals, some of which are captured in our senses and our instruments. These signals lead to perceptions of the noumenal reality, generating our phenomenal world. All our experience relates to this phenomenal world, which we interpret and sort out in the form of concepts, beliefs, and theories, the latter being specialized systems of beliefs — justified by evidence and accepted as true. Theories are meant to explain and predict the behavior of systems in the phenomenal world, where the systems and their interactions have a correspondence with entities and their correlations in the noumenal world.

   The noumenal world is concrete and real (though it is captured only partially in our perceptions), and bears the effects of our actions as we react back on nature in the course of our journey in life, though these effects can be perceived only in terms of our interpretations within the confines of our phenomenal world.

2. Theories are domain-specific and are built up by a process of inference from observed behavior of systems, that behavior being generated by their interactions. They constitute, in a sense, the distilled essence of our experience and, strictly speaking, apply to *models*, where a model repre-

sents some specified chunk of reality with some specific *context* added to it — the context represents the effect of the rest of the complex reality. The very complexity of reality makes a model sensitively dependent on the context. The truth or falsity of a theory is a question that can be settled only with reference to a model, defined in some context — however, the definition of the model along with its context may not be overtly precise, either or both being left implicit.

3. A model along with its context may, in a sense, be thought of as a *projection* of the infinite-dimensional phenomenal reality into some smaller domain — these projections being the faces of reality that we come to observe, depending on the limits of our senses and our instruments. Theories can then be described as constructs sorting out the behavior that we experience within these projections, and appear as codes from which their predictions can be obtained by a process of unpacking — one that often requires elaborate schemes of approximation. These predictions, on being compared back with experience, provide the 'reality check' on the theories.

4. Predictions obtained from a theory can be astoundingly accurate, testifying to the remarkable effectiveness of the inferential process in which these are arrived at, and also to the fact that the models to which these apply are, despite appearances, actually simple ones where much of the complexity of the real world are left out. For instance, the Newtonian theory of gravitation augmented with corrections from the general theory of relativity applies to a system of massive particles, *with all other interactions imagined to be switched off and all other complex structures ignored*, and predicts to an excellent degree of approximation a vast range of phenomena. Some of the ignored structures are then introduced into the theory in successive stages of approximation, such as the rotational motions of heavenly bodies, considered as rigid ones (fluidity of the cores of these bodies can then be introduced in the next stage of approximation).

All this is possible because of the *decomposability* (sec. 4.2.1) property of complex systems that owes its origin to islands of stability generated in the process of self-organized complexity. An analogous situation arises in the case of the standard model which, after all, is a simple one (again, with credit going to the remarkable inferential ability inherent in the way it was arrived at), and is applied to scattering processes where the gravitational interactions and all other structure in the world are left out of consideration — these, indeed, do not matter in the domain defined by these scattering processes.

As the context defining a model changes, as we look at some chunk of the complex reality from a different perspective, the behavior pattern within the model changes too, and can change dramatically since it is a different 'projection' that now acquires relevance. This results in a notable change in the structure of the theory describing the model, though there may exist common referents in the two theories (the ones applicable before and after the context change) regarding which these give near-identical predictions in some common domain. For instance, quantum electrodynamics introduces only a small (though crucial from a theoretical point of view) correction to the gyromagnetic ratio of the electron over the value obtained from relativistic quantum mechanics (the Dirac equation) since the latter ignores certain complexities in the electromagnetic interaction that the former takes cognizance of. However, the 'small' correction notwithstanding, the theoretical structure of quantum electrodynamics differs spectacularly from that of non-relativistic quantum mechanics.

Small corrections or small anomalies that appear to be insignificant at first sight (another case in point is the Lamb shift of spectral lines) actually act as pointers to complexities of the world lurking behind the apparent simplicity of a theory just as specks of dust visible at the boundaries of a rug on the floor point to a big mess of dust hiding under it.

5. Thus, one starts from a theory applying to a given model within some particular domain of experience, constructs a theory, checks for its validity within the context of the model, and then proceeds to incorporate some more complexity into the theory by attending to anomalies that the theory cannot account for. This corresponds to a new model along with a changed context and, even as the model addresses apparently small anomalies, an attempt at a correct explanation of these uncovers a big change where hitherto ignored complexities come up for consideration. This is how theories are built and re-built. Successive theories leave unscathed some predictions regarding the behavior of systems since these relate to common referents within the domains of these theories, but the small corrections to these obtained at successive stages of theory building are symptomatic of big changes in the structure of the theories arising from new complexities crowding in.

   In other words, *predictions regarding the common referents of the successive theories may change in a commensurate manner, but the theories themselves are not commensurate* and, at the end of the day, entirely new and novel phenomena are uncovered by the new theory.

   This can be illustrated by way of referring to the behavior of a solid. This behavior is captured in the spectral characteristics of the solid in various different ranges of the overall frequency spectrum (ranging from zero to infinity), as revealed by its response to various types of *probes* scattered from it and by a multitude of other types of response. One finds in gradual succession that the solid — a chunk of complexity that it is — admits description in terms of a bewildering variety of *collective excitations* that can be uncovered only bit by tiny bit. At each stage, one sets up a theory applicable to a situation where only one or a few types of these excitations are relevant, as revealed by some simplified *effective Hamiltonian* (expressed in terms of so-called *quasi-particles*), and then moves on to some other context where some other effective Hamiltonian describes a different set of excitations.

To be sure, the solid as a whole can be conveniently described by a Hamiltonian based on the electrostatic interactions between its charged constituents, shutting off all considerations of quantum electrodynamics and of the weak and strong interactions (where, in the process, electrons, protons, and neutrons are divested of possible internal structures, and gravitational interactions are ignored), but even this Hamiltonian, looked at as a code, is too difficult to unpack when the quantum mechanical symmetry principles on sets of identical particles (leading, among other things, to the *Pauli exclusion principle*) are taken into consideration. A piece of solid — a tiny chunk of the complex reality that it is — is nevertheless an unwieldy package of complexity in its own right and is a messy thing so far as a theoretical description of its behavior is concerned. One has to be content with a mosaic of partial theories, all differing from one another in their concrete ingredients, rather than one single overarching theory having a 'simple' structure.

As with the solid, so with the world of the so-called elementary particles. But here one lacks hard evidence in support of the viewpoint one holds (not that viewpoints are easily changed on the basis of hard evidence), because the wherewithal necessary for that is difficult to come by, even as fabulous — perhaps too fabulous — amounts are already being invested for the purpose. We will not comment on the investments being made because that is not under consideration within the confines of this book since it requires a discourse, among other things, on the *power structures* linking science to the rest of the human society. As higher and higher energy scales are accessed in investigating the scattering events among particles, attempts at zeroing in onto a simple and 'all-embracing' theory appear to prove futile ([149]), even though what appears to be futile to one may appear highly promising to others. The world of elementary particles is a complex one, carrying in itself the imprint of the complexity of reality at large, and lessons learned from the experience on complex systems are likely to apply here as well.

6. The incommensurability inherent in successive waves of theory building often appears in the form of *singular limits* in the transition from one theory to the immediately preceding one. Modifications in the predictions of a theory close to the border separating the domains of applicability of successive theories, commonly marked as 'small' corrections, are actually telltale signs of incommensurability, in that these appear as terms in an *asymptotic series* symptomatic of a singular limit.

7. Let us now imagine a hypothetical scenario where various different domains of experience are separated from one another by borders described in terms of not one single parameter (denoted above by δ) but of a host of relevant parameters (say, $\delta_1, \delta_2, \cdots$). As some particular parameter (δ_1) approaches a singular limit (say, $\delta_1 = \overline{\delta_1}$) from one side (say, $\delta_1 \to \overline{\delta_1}^+$), there appears an asymptotic series describing the value of some relevant physical variable, signifying the transition to a distinct theoretical framework appearing at $\delta_1 = \overline{\delta_1}$. The theory valid for $\delta_1 > \overline{\delta_1}$ on the other hand, involves other relevant parameters, any one of which (say, δ_2) becomes significant as some *other* border of reality is approached in an expanded domain of experience. The preceding theory corresponding to $\delta_1 = \overline{\delta_1}$ (for concreteness, one can think of δ_1 as the Planck constant $\hbar$, for which $\overline{\delta_1} = 0$) is also characterized by a similar border corresponding to some other physical parameter. Successive waves of theory building may correspond to such incommensurate transitions, as a result of which mankind goes on to build up a mosaic of theories in its perennial attempt at sorting out and making sense of the enormous complexity of nature that it is confronted with.

8. It remains to describe and understand the *asymmetric* relation between theories built in order to explain the behavior of systems within models

appearing in successive stages of expansion of our domains of experience. This is the problem commonly referred to as 'theory reduction' where a theory 'A' appears to reduce to a relatively simpler theory 'B' as some parameter (say, δ) approaches a singular limit (say, $\bar{\delta}$). The 'simpler' theory obtained with $\delta = \bar{\delta}$ can, to a certain extent, be understood in terms of the reducing theory 'A', but only in a close vicinity of the limiting value $\bar{\delta}$, though the converse is usually not true — the terms of reference of the theory 'A' cannot be understood within the folds of the theory 'B'.

However, the theory 'B' may, in turn, involve additional parameters that may correspond to similar transitions to other theories none of which are, however, related to the theory 'A' in an analogous way. This is how science builds up a mosaic of theories that in itself constitutes a complex system — a reflection of the complexity of Nature at large.

9. It is here that we point to the internal *structure* of a theory that helps us understand how two theories may be incommensurate with reference to each other and still have a set of common referents and a common set of ideas underlying both. Theories are made up of *concepts* correlated with one another by a multitude of relations of association, causation, and implication. The totality of these correlated concepts forms, in turn, an immensely complex network where, moreover, the network is a *multi-layered* one (sec. 4.2.1) involving several layers of relations among the concepts. For instance, there is one layer where the relation between concepts is expressed in plain language without scientific connotations, another layer expressed in mathematical terms, another one expressed in terms of theory with a limited domain of validity, still another layer expressed in terms of a broader theory, and so on.

   The conceptual network is, at the same time, a *co-evolving* (sec. 4.2.1) one, along with the structure of the concepts and of all these layers, where some concepts and some layers are added afresh, some retained, some get modified, and some get deleted as obsolete. In other words, in the event of a theory revision, some of the layers remain substantially intact while some others are modified in major ways and still others are added afresh. It is the set of concepts and layers of correlation that remain substantially unaltered that provide a common ground of mutual reference between the theories while the modifications and fresh additions are responsible for the relation of incommensurability between the two. In other words, successive revisions of theory correspond to *layers of meaning* being added to existing ones, in virtue of which an earlier theory cannot interpret one emerging in a subsequent revision while the latter is capable of interpreting the former in its own terms.

It is precisely this that incorporates the viewpoint of incommensurability within the framework of scientific realism without relegating it to the domain of social constructivism — a name rather disparaging from the viewpoint of the former.

10. Finally, what applies to 'reality at large', applies to the *human mind* as well — where the human mind is the relevant universe of discourse — there can be no ultimate theory in terms of which all mysteries and questions relating to the human mind can be answered, and the behavior pattern of an individual can be predicted under all circumstances. As we attempt to explain the behavior of an individual in given contexts, it is eventually bound to turn out that 'there is more to it than meets the eye' — the mind being a vast and fathomless complex system, its activities evolve in a complex manner, undergoing local instabilities, switching from one mode of behavior to another from time to time, throwing up a variety of patterns, exhibiting great diversity arising from dependence on remote causes, and evincing sensitivity to the context. Indeed, the last word here is that the *context itself is a complex system.* On the other hand, in this great sea of uncertainty and diversity, there remain *islands of stability, regularity, and predictability* at various *scales* in the relevant

'phase space' and in time.

The next section (sec. 4.5 below) will be devoted to an attempt at understanding *creativity*, which involves a *restructuring* of the conceptual space (refer back to sec. 1.10). The latter is one of vast expanse and complexity that, in the ultimate analysis, reflects the complexity of the world at large, including our mental world — there exists a huge repertoire of concepts regarding our mental entities and categories in addition to those referring to the external world.

4.5 Creativity: restructuring of the conceptual space

Creativity is one of those things that make our existence and our life blossom with colors while themselves remaining elusive and intangible. Since antiquity, humankind has never ceased to marvel at its own creative moments. Creativity is one of those things that make us believe that God resides in our soul.

Creativity has many — almost too many — aspects to it. To begin with, there have been anecdotal accounts of it and insightful but essentially conjectural attempts at understanding what the phenomenon of creativity involves. In more recent decades there have been systematic explorations and investigations into numerous aspects of this phenomenon, though the anecdotal — mostly introspective — and the conjectural accounts continue to remain relevant. Even as more disciplined lines of psychological and neuropsychological investigations are opened up, a satisfactory theory of creativity remains essentially elusive and conjectural. However, the conjectures are gradually acquiring more substantive content and creativity is beginning to acquire a less ethereal form.

In this part of the present chapter we will collect and put together a number of ideas relevant to the understanding of the phenomenon of creativity, confining ourselves mostly to the domain of cognitive psychology while we will, on a few occasions, hint at neuropsychological underpinnings as well. Without any claim whatsoever to originality, we will hope that the following pages will constitute a useful *point of view* for the understanding of the phenomenon of creativity. In the process, we will include a tentative suggestion regarding the way an apparent *non-determinism* enters into an act of creativity — one that can more generally be traced to the making of an inductive inference. Indeed, creativity can be described as a supreme act of inductive inference. Essentially the same non-determinism will be later seen to be involved in the exercise of *free will* (section 5.1) as well.

We will be concerned here mostly with the phenomenon of *scientific* creativity, and creativity in other domains will not be addressed separately. Creativity in various distinct domains has a number of basic aspects in common, though each of the domains possesses special features too.

In the following, we will steer clear of the vast literature on implementations of features of creativity in *AI systems*. While there remains little doubt that AI systems can replicate many of these features taken in isolation or in clusters, our focus in this book will be on creativity as it is realized in humans.

More specifically, we will adopt the position that a creative act is fundamentally analogous to the making of an *inductive inference* and the exercise of *free will* in that all these are intimately associated with the *self* of an individual. As mentioned earlier, the self (refer back to sections 1.9, 2.7) is a pervasive psychological engine based on the entire developmental history of a person, involving such mental ingredients as preferences, emotions, memories, personal beliefs, heuristics, capacity for generating analogies, repressed dispositions, and moral and social views, and is to be distinguished, at least notionally, from shared mental and psychological resources that she possesses in common with members of social groups or communities to which she may belong. The latter may broadly be described as resources shared with her *cultural* environment. At the same time, however, many of the self-resources are also imbibed from the *same* cultural environment. Indeed, a near-identical material and cultural environment leaves distinct imprints on different individuals, depending on their widely differing developmental histories where the latter includes the hugely complex succession of psychological and neuropsychological states of a person. Two persons sharing a lifetime of common associations and a common background may yet differ remarkably in their self-resources — this is explained by noting that the neuronal organization (integrated with the underlying biological system) and the emergent psychological makeup of a person are *complex* systems, possessed of unpredictable intricacies in their time-evolution.

We will not, in this book, try to state the defining features of creativity, to enumerate the various different kinds of domain-specific creativity, or to enter into the issue of measuring and comparing creativity. As stated above, we present nothing more than a point of view of looking at what the phenomenon of creativity involves, and will assume that the answers to many basic and specific questions are known, at least in an intuitive and commonsense way. As regards the defining features, we will just assume that creativity stands for the discovery of a *novel* concept in some domain of exercise of the human intellect, by means of which a new set of emergent ideas are glimpsed at, promising to open up an entire new field of engagement, possibly effecting a merger of erstwhile domains. In this, we will mostly consider *scientific* creativity, while occasionally referring to other areas of human endeavor as well.

4.5.1 Restructuring of the conceptual space

As underlined at several places in this book, *concepts* are the fundamental entities based on which we organize our perception of the world and act upon it. In this, *beliefs* play the fundamental role of establishing correlations among concepts and guide us in the all-important decision-making and inference-making activities (refer back to sections 3.1 and 3.2) — processes of enormous significance in our survival and adaptation. Many of these processes are enacted in our mind beneath the level of conscious awareness.

All the concepts — dormant and active in the mind — make up a stupendously convoluted and complex structure in our mental world that we have referred to as the *conceptual space* ([14], [53], [54]). The structure is convoluted in that the concepts are *mutually* related instead of being defined in explicit terms. As a fruitful analogy from mathematics, one can refer to a huge set of *implicit equations* where a large number of variables (say, $x_1, x_2, \cdots, x_N$) are determined in terms of a set of functional relations, each of the functions being determined in terms of *all* the variables taken together.

On the face of it, it appears that if the variable x is determined in terms of y (e.g, $x = y^2$) and y is defined back in terms of x (e.g., $y = x - 1$), then that is enough to send us in circles if we want to determine x and y in numeric terms. In reality, however, the two equations can give precisely determined values of x, y though in the example given above, one obtains two instead of one single solution in the pair of variables. In other words, a set of implicit equations is perfectly capable of producing a solution for the variables involved, *provided that* the equations are appropriately framed, since there may remain ambiguities and vagueness in the solutions obtained. For instance, a pair of implicit equations involving three variables x, y, z corresponds, in general, to values of x, y, z lying on a curve in the three dimensional space made up of the three variables. In the case of a large number (say, M) of implicit equations in a large number (say, N, with $N > M$) of variables, the former determine one or more regions in a N-dimensional space to which the possible values of the variables get confined. In the case of concepts formed in our mind, this residual vagueness and ambiguity is of vital relevance since it gives flexibility and versatility to our language and to our mental activities.

> Concepts are not formed on the basis of precise pen-and-paper definitions, but are constructed from our continuing experience of reality. The experience of reality (both the external reality and the inner reality of the mind) perceived at any given point of time or in a short duration cannot be exhaustively, unambiguously, and precisely registered in the mind, far less on pen and paper, if only because that experience has infinitely many aspects — we perceive only a limited number of these, a large part of that perception depending on our current state of mind — and, what is more, it involves affective associations that remain implicit.

Continuing to refer to sets of implicit equations, let us, for the sake of concreteness, consider a set of two variables ($N = 2$) x, y, satisfying a pair of implicit equations with, let us say, a single well-defined solution $x = a, y = b$. This defines two numbers (analogous to two concepts defined by mutual reference) related to each other by the pair of equations we started with, and a simple *network* made up of two nodes and two links between them. Analogously, the conceptual space residing in the mind of an individual can also be represented by a *network* made up of the concepts as nodes and their correlations (set up by means of heuristics and beliefs) as links between the nodes. The difference between the two cases is that, in the case of the implicit equations, nodes of the network are located on a single line made up of numbers from $-\infty$ to $+\infty$ (the real line, as it is called; more generally, the network is located in the *complex plane*) while the conceptual space can be one having an arbitrarily large number as its effective dimension.

Let us now imagine that a third variable z is added to the pair x, y referred to above, and a third implicit equation is added to the pair we started with (with that earlier pair also getting modified by inclusion of the new variable z), and see how the solution to the augmented set is related to that of the earlier smaller set. A more pertinent question is one involving not just two or three variables but a large set of equations relating variable $x_1, x_2, \cdots, x_N$ $(N >> 1)$, augmented to a larger set, where a new variable x_{N+1} is introduced and a new implicit equation is added to the set one starts with (where the new variable is introduced in the earlier set of N number of equations as well). The solutions to the initial set of N number of equations (in which the variable x_{N+1} may be imagined to be present as a parameter with a constant value) and the augmented set of $N+1$ number of equations form two networks — both, once again, located on the real line. The advantage of referring to networks arising out of sets of implicit equations is that, questions relating to the *geometrical structures* of networks can be discussed in concrete and precise terms, making use of the notion of the *separation* between any pair of points.

One can, for instance, compare the structure of the network obtained as the solution of the N number of initial equations referred to above and the one obtained from the augmented set of $N+1$ number of equations (we refer to the two networks by symbols 'A' and 'B' for the sake of easy reference). For instance, one can refer to pairwise distances between the nodes of 'A' and those of 'B' (this gives $N(N+1)$ numbers) and look at the minimum and the maximum of the distances so obtained (i.e., the range of separations between the nodes of 'A' and 'B'). In terms of these and other appropriate numerical measures relating to the structures of the two, one can form an idea of the extent of *restructuring* as 'A' gets transformed into 'B'. In the case of large N $(>> 1)$, the extent of restructuring is to be specified in statistical terms.

For networks obtained from sets of implicit equations, the restructuring can, in principle, be quite dramatic. In other words, the two solution networks can be markedly distinct from each other as compared in terms of the locations of the solution values on the real axis or in the complex plane, even when one single implicit equation, involving one single additional variable, is appended to the initial set of N equations ($N >> 1$) where the additional variable is assumed to be present in the earlier set of equations as a parameter with some pre-assigned constant value.

This metaphor of solutions to implicit equations can throw some light on the restructuring of the conceptual space that becomes necessary when additional concepts (only a few in number) are sought to be correlated with a large number of existing ones so as to provide an explanation of observed phenomena for which an existing theory fails to provide one. Such a state of affairs corresponds to the replacement of an existing scientific theory by a more effective one in which the conceptual space undergoes a marked restructuring. This is precisely where *scientific creativity* comes in.

Glimpses of the new theory first appear across the conceptual horizon of a lone scientist or of a small number of individuals who may or may not be part of a collaborative group. This happens when an individual (or a small number of individuals; we will, for the sake of convenience, refer to the conceptual restructuring initiated in an individual scientist) seeks to infer new correlations in her conceptual space, trying to find a consistent explanation of observed facts (or anomalies) that could not be explained on the basis of the previously existing correlations between concepts relevant in a problem domain.

Creativity, in brief, is the discovery of a new *dimension* to the complex and infinite dimensional reality we find ourselves in.

4.5.2 Creativity: Beliefs, inferences, and the self

The conceptual space of an individual is never static — its structure is perpetually in a state of flux due to ever new concepts being formed, and ever new associations and correlations between concepts being established. However, not all of the myriads of concepts lodged in the mind are involved in the evolution of the conceptual space at any given point of time, since only those belonging to a certain domain of inquiry, in which the mind is seeking to arrive at some inference, assume relevance in the inferential effort, in which fresh concepts are formed and fresh correlations are established within the periphery of that domain in the course of the process.

Among the concepts located within the inferential domain, some are well connected with others through existing associations and correlations, but some others may form an isolated island, being uncorrelated with the rest, or having tenuous correlations — established through beliefs characterized by low credibility, or ones that are not coherent with those currently activated in the inferential exploration. Among such uncorrelated concepts one is likely to find those that relate to experimental observations that are not tightly explained by the existing body of correlations among the well-correlated group of concepts.

> It may be recalled that correlations between concepts are set up by means of beliefs and heuristics in inferential processes where an inference can be described as one or more correlations set up in this manner. The correlated concepts often form new beliefs and, at times, lead to concepts of a higher degree of complexity, and myriads of such inferential process give the conceptual space an enormously convoluted and nested structure in an ongoing process of evolution. The process of correlations being set up between concepts, which is essentially an inferential one, is, at times, referred to as an 'exploration' in the conceptual space [14].
>
> Referring to the complex network representing concepts in the conceptual space, a set of concepts tightly correlated to one another corresponds to what is commonly termed a 'cluster' in network theory. A restructuring of the conceptual space corresponds to the setting up of new correlations between such a cluster and a sparsely connected set of concepts, when the extant correlations get transformed so as to generate a *consistent* set of correlations among the enlarged set of concepts embracing the ones freshly linked to the previously existing cluster.

Creativity — we are primarily interested in *scientific* creativity in this book — consists of a restructuring of the conceptual space with the following features characteristic of it.

(a) It establishes correlations between the mass of already correlated concepts and a few uncorrelated ones forming anomalies or unexplained experimental observations.

(b) In forming these new correlations, it causes a major restructuring among the concepts in the relevant domain of inquiry, whereby a large number of previously existing correlations get modified, for the sake of consistency or coherence, with the new correlations formed.

(c) As the relatively small number of concepts corresponding to the anomalies or unexplained experimental evidence are correlated with the previously correlated set (these earlier correlations considered collectively constitute a *theory* — one that now gets revised through the restructuring of the conceptual space), an early indication is, *at the same time*, obtained that an entirely new terrain of concepts lies *beyond* the ones representing the anomaly that initiated the restructuring in the first place and that get coherently correlated with the entire set of previously existing concepts. A new theory with a broader scope is thereby made visible where the relevance of a new set of conceptual correlations now becomes apparent.

All these supposed features of scientific creativity, that often bring in a major revision of existing theory and make visible the outlines of a new theoretical scheme, constitute a *plausible* description of how creativity is likely to be related to a restructuring of some relevant domain of the conceptual space. We recall that the fundamental process leading to such a restructuring is the 'exploration' of the conceptual space — a term that stands for an extensive inferential effort on being faced with an anomaly, i.e., unexplained evidence that cannot be explained on the basis of the previously existing set of correlations among concepts in the relevant domain of inquiry. We recall further that inference is precisely the cognitive-psychological process of explaining new evidence by the one of establishing correlations by making use of the existing body of beliefs and heuristics. The restructuring, in turn, brings in a process of *belief revision*, equivalent to the formation of new theory.

The only 'justification' of the new set of beliefs, i.e., of the emerging theory, is that the unexplained evidence in question is explained in a coherent manner. There is no guarantee that the restructured beliefs will be sufficient to explain *further* unexplained evidence as those are encountered in subsequent practice and experience. Let us, at the cost of being a bit abstract, call the set of beliefs generated in the restructuring process as 'B' and refer to the unexplained evidence that this set of beliefs makes explicable as 'U'. The only justification for 'B' is that it should explain 'U' and, at the same time, form a coherent set.

> However, 'coherence' is often a deceptive concept, being partly a matter of facile appearance, where emotions and affect play their role in causing an *impression* of coherence. Generally speaking, coherence depends on some *yardstick* against which it is judged, and that yardstick itself is likely to be an enveloping belief appearing as a *mindset* that makes a set of beliefs (such as 'B' referred to above) appear to cohere among themselves while the coherence with some *larger* set of beliefs is left unexplored. In other words, coherence is not a sure guide to justification. What is more, explanation of hitherto unexplained evidence is also an equally fallible guide, since the appraisal of evidence is itself theory-laden [82].

Now suppose that some *new* evidence (call it 'U′') is unearthed, requiring a further revision of the belief system whereby 'B' gets transformed to 'B′'. Within this scenario, consider the two sets 'B' and 'B′' in reference to the evidence 'U', i.e., the one that was considered to be anomalous prior to the observation of 'U′'. Evidently, 'B' and 'B′' are *both* sufficient to explain 'U', i.e, 'B' is *underdetermined* by 'U', but it still appeared to be acceptable *in the context where* 'U′' did not appear in the horizon of experience. In *that* context, the more inclusive revision to 'B′' was not on the cards (imagine Newton to have discovered quantum theory, bypassing the classical theory altogether — quantum theory, after all, is supposed to be inclusive of the classical theory — Newton did not lack in the talent but had to be, by necessity, consistent to his *context*). In other words, the belief revision corresponding to a restructuring of the conceptual space is dependent on the *context* of that revision.

Evidently, the above hypothetical scheme of conceptual restructuring and belief revision is essentially conjectural, but it generates a useful *point of view* in terms of which the phenomenon of creativity becomes, to some extent, amenable to concrete analysis, thereby shedding some of its elusive character.

4.5.3 Intuition, insight, and creativity

As mentioned earlier (see, especially, sections 2.7.8, 2.8), our beliefs have two aspects inherent in them — an aspect of being shared inter-subjectively, and one of being private and personal. Among these, beliefs that are predominantly inter-subjective are mostly accessible to the conscious mind, where the holder of a belief can be aware of it. On the other hand, predominantly self-linked beliefs are not always open to being accessed consciously, since one can be only dimly aware of those, and that too after much soul-searching.

We have also observed that beliefs may act as 'if-then' type rules for the establishment of correlations among concepts and that such activity of beliefs may take place in unconscious depths of one's mind.

Finally, the process of establishing correlations among concepts commonly appears as inference, leading to the evolution of the structure of the conceptual space. A certain type of major restructuring of the conceptual space ([14]), as outlined in 4.5.1 above can then be associated with creativity.

In this context, it may be of interest to briefly explain the terms 'intuition' and 'insight' before continuing with our attempt at understanding the phenomenon of creativity.

Generally speaking, the process of inference has both a conscious and an unconscious component, where the former involves the use of inferential rules — mostly in the form of beliefs shared inter-subjectively — that one can access by conscious reflection though, in reality, even these operate automatically in many situations, i.e., below the level of awareness (for instance, a deeply ingrained suspicion towards fellowmen). On the other hand, it is likely that the beliefs of an acutely personal nature operate at a more submerged level of the mind. Referring to the progression of an inferential process that commonly consists of sequential parts punctuated with decision junctures where the sequential progression branches out and one of the branches is chosen in a process similar to the adoption of a choice, the simple sequential progression is effected by means of inter-subjective rules, of which one has at least a minimal awareness.

Intuition is the name given to an inferential process that takes place entirely beneath the level of awareness and indicates that sound reasoning rules may be accessed and made use of even without a vestige of conscious control. It seems likely that this is achieved by means of heuristics [57] that have a solid content, many of which may have been acquired in the course of evolution — examples being sophisticated navigation techniques used by aquatic animals, insects, and birds, and even by humans under diverse conditions. We will not enter here into detailed considerations regarding heuristics having sound and highly effective content and how such heuristics may have been made part of the self of an individual, with a substantial role of unconscious inferential processes.

> There may be deep and subtle aspects to the acquisition and use of higher cognitive functions that operate in the unconscious hinterland of the mind. For instance, one possible mechanism may involve the so-called 'gut-brain interaction' (see, for instance, [92]), based on the enteric nervous system (ENS). In other words, our mind is a product not of our brain alone, but of the biological system as a whole.

Thus, intuition is 'intelligence of the unconscious', where we don't have a cue as to how we arrived at a correct inference. Insight and creativity share this one feature with intuition: these are overwhelmingly dependent on unconscious processes, with little realization on our part of the way these lead to success — at times quite awesome — in hitting the bull's eye in inference.

While inferences result in the setting up of correlations among concepts, most of those do so, in a sense, on a 'microscopic' scale where the concepts so correlated are not remote from one another. This statement makes sense when one has in mind some metric specifying the separation or distance among concepts in the conceptual space. We will not dwell on this difficult issue here, partly because no satisfactory metric appropriate for our present context is known (see [53] for a penetrating approach in this regard), but mostly because the structure of the conceptual space with all its tangled and nested features and all its mutual correlations at multitudes of levels is still not sufficiently clear for a quantitative description and analysis. Instead, we will assume in a qualitative sense that a notion of separation or distance between concepts and conceptual domains is defined, at least in some given context and within some given region in the conceptual space.

Recalling our earlier reference to networks (sec. 4.2.1), there exist notions of separation between a given pair of nodes in a network in terms of the number of links traversed in the shortest route from one of the nodes to the other [133]. However, there may exist a multitude of links of diverse types between nodes in a possible network representation of the conceptual space, owing to which the notion of distance in the conceptual space becomes context-dependent. Still, it makes sense to refer to sets of nodes as being sparsely or richly connected, where such an evaluation appears to be meaningful between domains of the conceptual space as well — two domains may be close to each other or remote, depending on how densely these are correlated and how stable these correlations are (recall that new correlations are constantly in the process of being established between concepts, and old ones get revised by means of inferential processes).

When an individual arrives at an inference intuitively, she does not know how that inference came to be formed, but such an inference does not commonly establish a correlation between *remote* regions of the conceptual space. More often than not, such an intuitive inference makes use of tacitly acquired cues that are made use of in the setting up of heuristics which then get involved in establishing the requisite correlations among concepts. For instance, *an experienced driver, on hearing a strange sound while driving a car, immediately locates the cause even without conscious thought, and reaches for his toolkit in order to set the offending machinery right.*

This is why all instances of intuition cannot be described as insights, since an insight is distinguished by the fact that it establishes correlations between *remote* concepts in some given context. Consider, for instance the following:

I lost my denture that I took out one day and kept on a shelf while cleaning my mouth and subsequently taking bath. Forgetful of setting it back where it belonged, I came out of the bathroom and became aware of having lost it a couple of hours later. I then spent half an hour in a vain search of it, wondering all the while where I could have lost it. Suddenly, in a flash, I recalled having taking it out in the bathroom and was certain that I must have absentmindedly put it in the medicine cabinet when I had taken an aspirin tablet from it and then put the denture there instead of the medicine bottle — and, sure enough, on going back to the bathroom, I found the bottle on the shelf and the denture in the cabinet.

Or again:

A patient suffers from strange fits of weakness and depression that a number of physicians are unable to diagnose and cure. One of them, after observing several episodes of the malady, one day orders a rare blood test that the other physicians — even renowned ones — did not even think of, and the cause of the affliction was immediately identified.

Insights may be more or less remarkable in their result (the medical diagnosis certainly more so than the retrieval of my lost denture) but all have the common feature that they establish correlations between remote concepts (denture and bottle of medicine, or fits of depression and rare blood condition). However, in most cases these correlations do not have a *cascading* effect, i.e., they remain confined to isolated sets of concepts and do not result in any major (or 'macroscopic') restructuring of the conceptual space.

Such major restructuring is, however, a characteristic feature of creativity (recall that our focus in this essay is on *scientific* creativity). In other words, an act of creativity does not remain confined to establishing correlations between a limited number of concepts in a limited region of the conceptual space. What may be initiated as an act of setting up of such limited correlations, may quickly result in a domino effect. To start with, the correlations set up initially give a sudden glimpse of an entire major region of the conceptual space that was hitherto unconnected with an existing region (a well connected cluster of concepts between which correlations had been set up in an earlier phase) from which the conceptual exploration got started. This is then followed up by numerous correlations being tried between the two conceptual domains (the previously explored domain and the one that is now accessed through new correlations), all of which are found to be relevant (and themselves productive of fresh correlations) in the context of the exploration. At this stage, the individual scientist engaged in the mental exploration experiences the much-discussed ‘aha!’-effect — a surge of startled ecstasy. The creative act is then followed up by groups of workers, subjecting it to exacting tests, and then giving it the stamp of a new and improved *theory*.

In other words, scientific creativity often results in partial or complete replacement of an existing theoretical framework in some domain of inquiry with a new and improved theory. This phenomenon is, at times, referred to as a ‘conceptual revolution’. A conceptual revolution often introduces a new mode of discourse in the domain in which it occurs. We will briefly touch upon this issue later in this chapter (sec. 4.5.8).

4.5.4 Affect and inference in creativity

All along the process of restructuring of the conceptual space, it is inference that constitutes the principal mechanism underlying creativity. As indicated earlier (sec. 3.2.2), inference is made possible with assistance from the affect system, where the latter acts as an *internal monitor* in the onward progression of the inferential process, with inter-subjective and personal beliefs establishing correlations between concepts in some domain of the conceptual space. What is special in an act of creativity is that remote domains get correlated and a ‘macroscopic’ restructuring sets in.

The importance of the ‘internal monitor’, even though invisible to the conscious mind, can hardly be overemphasized. The ever vigilant steering it provides forms the essential precondition of all goal-directed activities of the human mind. In particular, the evolution of the conceptual space that takes place by means of incessantly occurring inferences is made possible by the imperceptible and silent activity of the affect system where the evolution itself occurs in two discernibly distinct ways — one of a gradual change by means of inferences that establishes correlations only locally, and the other of a major restructuring that occurs through the setting up of remote correlations, not only between isolated concepts but between entire domains where a cascading effect takes place with one correlation almost spontaneously leading to numerous others, thereby opening up an entire new terrain in the conceptual space.

In order that such cascading may be made possible, the inferring mind, equipped with the unique ability of the affect system, has to *evaluate* the relevance of the conceptual correlations being set up in quick succession. *This* requires a trained mind, where the training consists of generating a vast repertoire of heuristics relevant to the task at hand — one of producing an explanation of an anomaly that defies a satisfactory resolution even when the resources inherent in an entire conceptual domain, richly endowed with correlations existing within it, are brought to bear upon the problem.

It may so happen that the 'problem' at hand is apparently rather a 'small' one — say, like the explanation of the spectral characteristics of atomic hydrogen or the explanation of the spectrum of black-body radiation, or even an anomaly in the expressed traits in samples produced in successive generations of a species of plant seeds. To start with, the problem remains hidden behind a big mass of 'success' of the existing theory within the domain under consideration. But it distinguishes itself by being strangely recalcitrant to the authority of that theory, even when the full power of the latter is brought to bear upon it. The hugely successful classical theory made up of Newtonian mechanics and Maxwell's theory of electromagnetism proved to be inadequate in explaining the hydrogen spectrum and the spectrum of black-body radiation.

As an analogy, one may think of the detection of a small error in the solution of a set of equations — let us say, a set of implicit equations in a large number (N) of variables. Commonly one attempts to find a solution by starting from a set of plausible values of the variables (in the so-called complex plane) and then initiating an iterative procedure that is expected to converge upon the exact solution. However, even when the initial values are chosen close to a purported solution (with only a small error showing up when the values are substituted in the equations), the iterative process, instead of converging, may begin to diverge with successively larger errors showing up in the process. This tells us that even a small error may hide a large discrepancy when compared to the actual and exact solution. On the other hand, when the iterative process is initiated from a set of assumed values in a distant domain of the complex plane, it may be found to converge quickly. In order to find an appropriate domain in which the initial set of values is to be chosen, the computation is to be carried out for a relatively large set of initial values (the process of ‘exploration’, where the store of heuristics assumes relevance) and every time one has to *check* whether the iteration is still diverging or is showing signs of convergence. In all likelihood, the right domain may be away by a large separation from the one from which the process was initiated to start with even when the initial values in that domain showed only a ‘small’ error on being substituted in the given set of equations. In the case of a set of equations involving a large number of variables, small errors may be particularly deceptive in the acceptance of a purported set of values of the variables as the true solution, which may lie in a domain remote from the initial one by a large measure. What is more, finding the true solution in this meticulous process of checking a large number of possibilities, may open the door to the discovery of a novel iterative method for the solution to a whole new *class* of sets of implicit equations.

In all this reference to sets of implicit equations in large numbers of variables, I only speak loosely, without sound mathematical theorems in my knowledge — I hope that this analogy with what can conceivably constitute to be a scenario involving implicit equations may be instructive in an attempt at understanding how the mind closes in on the solution to an anomaly that defies explanation in terms of the large number of correlated concepts within a given domain, and seeks out relevant correlations with concepts in a distant domain.

The basic idea in invoking the analogy with systems of equations in large numbers of variables is that the latter requires *a large number of consistency conditions to be satisfied simultaneously.* It is then conceivable that a small overall error (defined in some appropriate manner) detected in a purported solution may not be amenable to correction by making small changes in the values of the variables and such small changes may increase rather than decrease the overall error. In other words, the *basin of attraction* of the exact solution with reference to the iterative process of locating it may lie in some distant domain referred to the domain one starts from.

This, however, is difficult terrain in mathematics, and I am not competent to deal with these questions on rigorous basis. Only a loose (and, perhaps, illuminating) analogy is sought for.

Of vital relevance in the above computational approach to locating a solution to a set of equations is the process of continuously checking whether the intermediate succession of initial values in the various domains lead to a diverging or converging process of iteration. This requires a succession of *decisions* to be taken in executing the algorithm of the computational process, where each such decision depends on the output value of a certain number (with a magnitude and a sign) indicating the error in some stage or other of the iteration. *Additionally*, the algorithm is to compare successive error values to decide whether the iterative process is a diverging or a converging one. In other words, a *huge* set of decision processes (including the generation of error values and comparisons between successive errors) becomes necessary in order to relocate from one domain to another in the space of possible values of the relevant set of variables. In the case of a numerical algorithm under consideration this checking is performed in terms of the numerical value of an intermediate output that tells the program whether to go ahead or to adopt a different course altogether.

In the absence of such evaluation and monitoring, the computation may prove to be wholly *intractable* — more so when the solution to be arrived at is located in a remote region as compared to the starting point. A numerical algorithm often proceeds through an apparently random choice of successive initial points (the so-called *Monte-Carlo* approach) but even so, the continual checking for convergence is essential for the actual success of the process.

The inferring mind may conceivably adopt a 'Darwinian' approach of establishing random correlations and selecting out the ones that seem to be favorable to the solution of the problem at hand, but it appears that the mind lacks the wherewithal necessary for a truly random search ([70]) in the conceptual space.

However, this statement needs to be qualified. Creativity requires that the mind should go into *incubation* once it settles upon the job of solving some problem and realizes that the solution is not forthcoming when looked at from known angles and the exploration for the right concept needs to be carried out over a broader terrain. A relatively short incubation phase is also frequently involved in insightful inferences but such a phase of a longer duration features more prominently in an act of creativity. It is possible that some degree of spontaneous and uncorrelated neural activity may be involved during such incubation phase, but a more likely explanation of the necessity of incubation involves other essential aspects of creativity, as we see in sec. 4.5.5 below.

At the same time, it may be mentioned that spontaneous and uncorrelated activity may occur on two different *scales* — one of these involves *microscopic* fluctuations in the activity of single neurons or small groups of neurons in the brain, while the other refers to fluctuations in the correlated activity of large scale neural aggregates that determine the multitude of unconscious and conscious mental states and psychological functions of an individual — such fluctuations may be termed 'macroscopic' ones. Microscopic fluctuations occur incessantly within the neuronal assembly of the brain and have little psychological relevance, if any. Macroscopic fluctuations, on the other hand, may possibly be involved in the Darwinian process of exploration of the conceptual space referred to above.

Commonly one finds that reference to the affect system in the literature is confined to the role and significance of the 'aha'-effect as it is experienced in the process of conceptual restructuring in creativity. On the other hand, our present considerations gives center stage to the affect system in the entire inferential process leading to conceptual restructuring, as it occurs in creativity.

A correct inference has an obvious adaptive value — correct inferences, indeed, constitute the driving force of life. It is also not too far-fetched to conjecture that finding an unexpected short-cut was a rewarding experience for early humans as it is in modern man, and is possessed of adaptive value too. Taken together, the two may explain the conspicuous affect expressed in the form of the 'aha'-effect, where the associated emotions may partly be of evolutionary origin.

4.5.5 Constraint and freedom in creativity

An act of creativity always takes place in some *context* as does any and every act of inference — deductive or inductive.

Major acts of creativity may also occur serendipitously, to which the discussion in the present chapter applies only in parts. Serendipity and purposeful search for an explanation mostly go hand in hand, one or the other of the two being the dominant feature in any given situation. An individual or a group may be engaged in a research program with some purpose in mind, during which there takes place an evolution of their conceptual network, when a chance observation suddenly connects up hitherto uncorrelated conceptual domains, and a related but somewhat different issue gets flooded with illumination, almost too dazzling to bear.

Most inferences are of the inductive type where one has to go beyond evidence and choose between alternatives in which one is helped in a major way by the affect system, notably in the search for new domains in the conceptual space to which links are set up, to be pursued further on receiving 'approval' from the affect system. However, the process never assumes the form of a wild and blind search, as no inductive inference ever does — inferences are highly constrained by context and are domain-specific, since otherwise the mind would have to make a choice between too many alternatives and the process would be too heavily underdetermined — it is psychologically costly and unnecessary to conduct a search among chemical species while trying to solve a problem on gravitation. It would be equally out of the context to zero in on categories like 'bleen' and 'grue' (Goodman's paradox) while inferring on the color of emeralds from the observation of a number of green ones.

But too much of domain-specific constraint is suffocating and lethal where creativity is concerned — much like the case of the globe-trotter facing visa problems.

In order to explore distant domains and establish remote correlations, the mind needs to be set in the *default mode*, where it desists from outwardly directed purposive action and engages in intrinsic activity based on the default mode network (DMN) of the brain (see, for instance, [108]), congenial for long term planning and prediction, imagination, and *creativity*. This has been referred to as the *incubation* phase in reference to acts of insight and creativity and has been mentioned in introspective accounts of numerous original thinkers, notable among whom having been Henri Poincare ([14], chapter 2; [120], chapter 2). Such an incubation mode, often akin to 'day-dreaming', is perhaps even more of a necessity in the case of creativity in literature and the arts where domains are not always demarcated clearly and where internal, self-linked psychological resources are engaged not in establishing correlations between specific conceptual domains but in setting up associations between whole clusters of psychological ingredients mostly buried deeply in the unconscious.

Creativity in literature (including, specifically, poetry), music, painting, sculpture, and the like is related mostly to what have been referred to as the *psychic* or *repressed* components of the self earlier in this book (sec. 2.7.1), as distinct from the *cognitive-adaptive components*. While cognitive components principally involve beliefs of diverse types, the psychic components include such ingredients as drives, yearnings, cravings, fantasies, unfulfilled aspirations, repressed moral and ethical beliefs, poignant memories, and a host of similar other constituents of the psyche — ones that get addressed by the *soul*. Both the psychic and the cognitive components are, without exception, soaked to a greater or lesser extent in emotions and affect. It is affect that, in the ultimate analysis, steers us through the incredibly complex maze of life, at times leading us on to acts of creativity.

It is in the incubation phase that the domain-specific constraints on inference are released and the mind engages in the *free* exploration of the conceptual space, where freedom too is to be constrained appropriately. Creativity needs 'freedom' on two counts — one of these is the freedom to explore distant domains in the conceptual space while the other is the freedom from knowledge-based, socially induced and inter-subjective rules of inference so that the self-linked beliefs and heuristics may get into the act. The latter kind of freedom appears as a fundamental *non-determinism* in creativity since determinism is the name commonly given to *predictability* in terms of identifiable causal effects (refer to section 5.1 in chapter 5).

As indicated in sec. 4.3.2.1, a system is *self-determined* if there is no external system or agency that determines its behavior, and it evolves in accordance with rules that can be expressed in closed form, say in the form of a set of autonomous differential equations or an unambiguous code that does not need external inputs during run-time. Thus, for instance, a cellular automaton is self-determined even though it evolves according to a code set up by an external agency (an AI system or a human intelligence) — once the code is given, the behavior of the system made up of the automaton and the code does not depend on an external agency (save for the specification of the initial configuration of the automaton). The question, however, arises as to whether and to what extent such a self-determined system is *determinable*, i.e., how far its future behavior can be predicted in advance. The description of the rules of evolution of a *Turing machine* may be completely known, making it a determined system — nonetheless, given an arbitrary input string, it is not *decidable* whether the machine will ever halt. More generally, the behavior of the machine or of a self-determined system at an arbitrarily distant future may be indeterminate to a large extent. In a sense, predictability is the *foundational* issue behind all attempts at theory building. It underlies all our attempts at understanding *reality*.

Incidentally, the statement that reality is self-determined is a point of view and is not based on hard evidence. It implies that one can think of reality without a Creator. The idea of a creator once again rests on a point of view that one may or may not accept depending on how one interprets one's experience in life.

Creativity is intimately dependent on freedom from recognizable causal relations precisely because it is made possible by self-linked psychological resources — ones that remain hidden to an onlooker (even, to a large extent, to a psychologist) and mostly to the individual too who engages in the creative act. In this, creativity is analogous to the exercise of free will ([70]; see also 5.1 in this book) that appears to be non-determinable in a large measure.

At a fundamental level, the freedom from constraints in the exploration of distant domains in the conceptual space and the freedom from casual effects, defined in inter-subjective terms, in the making of a decision in an inferential process, are not too different from each other. Because, conceptual domains are, to a large extent, demarcated from one another by means of knowledge-based and inter-subjective categorization. In order that domain boundaries may be dissolved in the conceptual exploration, one has to suspend to a large extent the inter-subjective criteria underlying the demarcation of domains which is precisely why the incubation phase is so necessary in an act of creativity. What is more, the crossing over from one domain to another needs a choice to be made (and, equivalently, a decision to be adopted) in the same way as in an ordinary inferential process, where inter-subjective or shared rules are largely inoperative.

In other words, creativity is at heart a spectacular inferential act where the inference is freed from the constraints of some specific domain or other.

> In any given line of inquiry, concepts may have proliferated to such an extent that what was a conceptual domain yesterday is divided into sub-domains and even more restricted regions today. This is the natural course of evolution of the conceptual network — by the creation of new concepts, new correlations, and new clusters (densely correlated groups of concepts). Routine inferences take place in some tiny region within a larger domain. Creativity would then require that the constraints imposed by a fixed and limited region be relaxed and a conceptual exploration be carried out in some appropriately larger region.

In this context, it is important to appreciate the role of the *default mode activity* of the mind, based on the brain's default mode network (DMN), in creativity. As mentioned above, and as explained in brief in section 3.2.2.8, the default mode is the ideal and ever-active platform where inferential activities in general and creativity in particular are launched so as to avail of the benefits of affect and reason simultaneously and in an intimate mix. It remains an interesting and plausible conjecture that the default mode generates fruitful *analogies* essential in making possible the 'logical leaps' involved in inference-making and creativity (refer back to sections 3.2.2.5, 3.2.2.8).

4.5.6 Creativity: heuristics and non-determinism

We have made a qualified distinction between beliefs in general and heuristics (see sec. 2.8.2). Heuristics are beliefs of a special kind — they often have good credentials in being justified to a larger degree when compared to the majority of beliefs of other kinds that we hold, are more readily subjected to the test of evidence and revised accordingly, and have a better 'turnover rate' — being produced and discarded with greater facility, depending on their efficacy in the inferential process. While we may be aware of many of the heuristics held in the vast store of these tiny and active bits of belief lodged in our mind, many others remain hidden in the unconscious depths of it and presumably play a prodigious role in the conceptual exploration undertaken in acts of creativity. Heuristics, in particular, are generated in the very acts of inference-making where they prove to be of such great value.

As the examples of heuristics based on symmetry-related concepts (held by engineers, scientist and mathematicians) indicate, along with a huge number of others rendering good service to artisans, technicians, athletes, car drivers (men and women from almost every walk of life) — we are most of the time unconsciously guided by heuristics in all activities that have some degree of automaticity built into them. Most of these heuristics — though justified to no small degree — are included in our self-linked psychological resources since these are produced in a manner intimately related to our developmental history (including the history of inference-making in the past) and are associated to some extent with emotion and affect, though this feature of the heuristics may be less pronounced than what applies to other beliefs entrenched in the mind. In any case, an individual's store of heuristics is often specific to her and depends on her occupation, habits, temperament, and commitments in life. Above all, heuristics constitute a dynamic component of the mental world of a person and are highly active in establishing correlations among concepts, much like efficient enzymes in biochemical reactions.

It is the dynamics of the conceptual network, to which the fluidity of the heuristics contributes in a large measure, that is a major contributing reason behind the *non-determinism* inherent in creativity. The application of heuristics to the establishment of correlations among concepts occurs on such a short time scale compared to the time characterizing an act of creativity (including the incubation phase) that the latter effectively becomes *history-dependent* (for background on history-dependent processes in complex systems, see [133]; see also sec. 4.2.6), i.e., if the process is imagined to occur a number of times under identical conditions (i.e., with the context of the process remaining the same), the outcome would differ from one occurrence to another depending on the past history of the system under consideration, which determines the time sequence of generation and activation of heuristics in the mind, a process that occurs rapidly in the comparatively slow process of creative exploration of the conceptual space.

This, in a sense, distinguishes the non-determinism characterizing acts of free will ([70]; refer to section 5.1) and that relating to creativity. An exercise of free will depends in a major way on the activity of self-linked psychological resources that remain outside and beyond known causal relations, owing to which the act appears to defy the principle of determinism. In contrast, the heuristics that make a major contribution to an act of creativity are mostly ones based on justifiable principles (and hence appear to be inter-subjective ones to a greater extent) that can be identified as having contributed to the act, but only *after* it has occurred — the fact of its occurrence appears to be a random event owing to the exceptionally dynamic nature of the activity of the heuristics. Having said this, we have to add that the self-linked nature of the heuristics is also relevant in conferring a degree of non-determinism to acts of creativity (and, likewise, the fluidity of heuristics, and of thought process in general, must also be responsible in some measure for the non-determinism characterizing the exercise of free will). As we have seen, heuristics may have a good measure of justification in them but are still dependent on the details of the developmental history of an individual such as the level of rigorous practice in some particular domain that she goes through, the modes of thinking that she adopts in her various stages of development, the types of sources that she exposes herself to, and very many other factors of a similar nature. The items of knowledge and inter-subjective beliefs that a person acquires in the course of her life may be pretty much the same when compared to another individual with a similar background, but her store of heuristics may be vastly different.

1. As indicated in earlier sections of this book, *all* beliefs, including ones in the nature of heuristics, have shared and self-linked aspects to them — it is the one or the other of the two that acquires relevance in any given context. Our beliefs are generated by sets of experience specific to an individual and also by experience acquired in virtue of membership in larger social groups.
2. While determinism gets compromised in acts of free will and creativity, *causality* continues to be operative, though the exact manner it operates becomes indeterminate.

Heuristics are constantly being generated in our mind in the form of intermediate results in inferential acts, especially those that lead to successful conclusions. Recall that it is affect that acts as 'internal monitors' in inferential processes of the mind. In this the affect system essentially makes use of resources generated in earlier inferences of a successful nature, including the heuristics produced in those earlier acts where the heuristics, in turn, get lodged in the mind as self-linked resources associated with affect.

4.5.7 Analogy in creativity

The foundational ingredient of heuristics and beliefs — and indeed of all thought — is *analogy*. And, it is analogy that is primarily responsible for the triggering of a major restructuring of the conceptual space, so typical of an act of creativity.

The crucial relevance of analogy in our thought process in general, and in inference and creativity in particular, has been discussed by numerous authors (see [76], [14], [65], [67], [55], [142]) — making up a vast storehouse of ideas on how much of a subtle and universal role does analogy play in creative thought.

Analogy can truly be identified as the most effective means of establishing correlations among concepts. In broad terms, it can be described as a mapping from one concept (the 'source') to another (the 'target') and, as such, can indeed be looked upon as *the* universal means of correlation, since some mapping or other can always be defined regardless of what the two concepts (the source and the target) are. In this sense, analogies among concepts do not intrinsically reside in the concepts themselves but are generated and employed to relate them by the inferring mind. The setting up of an analogy is itself an act of inference where the former plays a role similar to a belief. However, this way of defining an analogy is too broad to be of use unless one takes into account the fact that an analogy can be either shallow or deep depending on whether or not it *simultaneously* tells us something of the way the two concepts are related to *other concepts* correlated with them. In other words, a deep analogy is one that tells us something interesting about the properties of the target-concept from those of the source-concept, thereby setting up a densely correlated *cluster* around the target concept in our mind. However, whether this cluster of concepts is 'interesting' or not depends delicately on the context. In this respect, a good analogy is like a metaphor or a joke that has to be tuned to the context to be of genuine interest since otherwise it falls flat. Just as a joke or a metaphor does not reside inertly and intrinsically in the situation it lights up but has to be discovered by an acute mind (though not invented by it), a fertile analogy is also one to be discovered within a given context to be useful in setting up new correlations. Analogy is a complex thing that results from a conjunction of the concepts involved (the source and the target), their context, and the store of ideas and heuristics in the exploring mind.

The important thing about analogies in the context of creativity is that their production and use is not directly dependent on the contiguity of the concepts involved. In this, an analogy once again resembles a good metaphor or a good joke, both of which are capable of connecting up *remote* ideas — the more distant these are, the more effective and potent the metaphor or the joke is. Run of the mill analogies depend on the contiguity of the source and target concepts that offer easy means of being correlated through their common or shared correlations. While two such concepts can be linked in an almost trivial manner, they can also be correlated through a distant route that can more often than not throw light on other concepts lying on that route, whereby the analogy becomes an 'interesting' one — a potent source for further correlations.

While an ordinary belief or heuristic correlates a source and a target concept by means of a relation of implication (where the implication may even be a vague one), analogies operate predominantly by means of association, owing to which these have an amazing power to link up remote concepts.

For instance, when one is asked the question '*in what respect is a glass analogous to a bottle*', he may respond by saying, '*both are made of glass*', which would be a rather uninteresting and flat observation. On the other hand a response such as '*both start flying in the late hours of a party*' would be more interesting in establishing a correlation through a rather long and circuitous route, thereby suggesting a revealing picture of a party in an advanced stage of dissolution — a picture impregnated with a host of associations of a telling nature.

Here we have left unspecified the precise definition of 'remoteness' of two or more concepts. The idea of remoteness rests on the numbers and types of links connecting them in the conceptual space. For instance, two concepts with no links connecting them would be taken to be almost infinitely remote. No two concepts can be literally infinitely distant from each other, since there must always exist some route, however circuitous, connecting them. In this context, we recall that there exist various different *types* of links between concepts. For instance, two concepts can be correlated by means of links based on their dictionary meaning and also by those of a mathematical nature. It would be the latter type that would be relevant in determining the distance between these in a mathematical exploration.

In other words, analogies generated by a fertile mind (equipped, let us say, with a huge store of relevant heuristics) can be effective in establishing *long range* correlations across domains, somewhat resembling the case of critical phenomena in *phase transitions* in the physical sciences.

1. The principal reason why we do not have a precise definition of the notion of distance between concepts is that concepts are often linked in a complex manner by correlations of multiple types. We hope that our tacitly held and common-sense notions of contiguity and separation between concepts are adequate to meet the requirements of the present set of considerations.
2. In a phase transition of a material from, say, the liquid to the solid phase, an analogous phenomenon is observed, where states of atoms at distant locations become correlated even though the interaction between any pair of atoms is of a much shorter range. In particular, a *critical state* observed in phase

transitions of diverse types, is characterized by correlations over an *infinite* range — it may be mentioned that phase transitions are possible, in principle, only in systems of infinitely large volume. How far the features of a 'phase transition' in the conceptual space resemble those in an infinitely extended material remains to be explored.

Phase transitions of a more general description are common in complex systems represented by networks [133]. One commonly observed scenario is that of *self-organized criticality* [133], where a slow *driving process* takes a system to a configuration on the borderline of stability when a cascading process starts, involving long range correlations among the components of a system, and a new stable configuration is arrived at. In the case of the conceptual space, during a relatively slow process driven by dynamically generated heuristics in the 'preparation' and incubation phases (the preparation phase is one where the mind absorbs ideas and perceptions from external sources and gets focused on the a problem that it is committed to solve, before going into incubation), both forward and backward inferences are set up (recall that 'backward' inferences are ones that proceed backwards from concepts supposed to be related to the purported solution to the problem at hand) when, at some point of criticality, a host of correlations are envisioned and a huge set of correlations fall in place as in a jigsaw puzzle. At this stage, some grand analogy becomes apparent, connecting two distant domains in the conceptual space.

The preparation phase within the mind of some single individual or of a small group of individuals may be preceded by a prolonged course of cumulative development in the conceptual framework of a larger community of scientists. An individual scientist, before entering into the incubation phase prior to an act of creativity, exposes her mind to all these cumulative developments and then isolates herself from the constraining effect imposed by external sources and lets her own self-linked and group-induced (sec. 2.7.8) psychological resources act as catalytic agents in the conceptual exploration.

Invoking deep analogies is a profoundly personal ability — presumably dependent on self-linked psychological resources (including the repertoire of heuristics at one's command), with roots extending to the unconscious level ([110]).

In closing this section on analogies in creativity, we repeat, from a slightly different angle, that analogies are not constrained by the terrain defined by shared rules in a domain of inquiry or, put differently, by explicitly formulated principles in the process of exploration of the conceptual space. In this, analogies are especially effective in exploring the infinitely extended and complex conceptual space since explicitly formulated rules operate sequentially and are effective in the exploration of limited (or densely connected) regions in the conceptual network, while analogies, which operate on an *implicit* basis make possible a more holistic exploration. For instance, analogies can link up various different levels of self-organized complexity without requiring the reduction of one level to its predecessor in the hierarchy of levels (for instance, the binding of the constituents in a nucleus is analogous to that of the elementary constituents in a proton). Indeed, it is doubtful if the role and efficacy of analogies can be explained in precise and explicit terms, since analogies are based on the ability of the mind to operate at an implicit level, based on affect and emotions — its pre-linguistic (or, more pertinently, *meta*-linguistic) ingredients. It may be noted that the affect system itself operates, generally speaking, by detecting analogies among various distinct episodes of experience.

4.5.8 Conceptual revolution

Scientific creativity, at times, gives rise to *conceptual revolutions* in the world of science. Such revolutions may lead to the emergence of a new theoretical framework in some domain of scientific inquiry — often expanding that domain almost beyond recognition — or may remain confined to some given domain, causing a major development within it, generating a fresh surge of ideas.

Such conceptual revolutions were highlighted by Thomas Kuhn [80] according to whom the historical course of scientific progress may be described in broad outline as an alternating succession of phases of so-called 'normal science' and 'scientific revolution'. Kuhn's description of normal science drew attention to a phase defined by a *paradigm*, which loses relevance as a 'revolution' breaks out.

Such a scheme of the course of scientific progress leaves many questions unanswered and even raises some doubts, but itself serves as a useful paradigm in looking at the historiography of science. I have referred to a number of such questions and doubts elsewhere ([82]; see sec. 4.3.10 for the relevance of Kuhn's scheme in *theory choice*) — even so, the paradigm of scientific progress that Kuhn outlined fits with the idea of self-organized criticality within the universe of scientific ideas which, in turn, dovetails with the scheme outlined above of the restructuring of the conceptual space within the mind of an individual.

> It has even been asked if Kuhn's scheme relating to theory revision runs counter to the very concept of scientific progress. Kuhn's idea relating to the acceptance (by members of a scientific community) of an emergent scientific theory over a pre-existing one has been seen as a brand of philosophy leaning towards what is referred to as *social constructivism*. I have reservations about the branding of a point of view with a name, sharply demarcating it from alternative and apparently contrary points of view, and giving it either a thumbs up or thumbs down sign which — though useful at times for the development of specific ideas — is, at the end of the day, uncongenial to progress as it is commonly interpreted.
>
> Terms like 'progress' or retrogression are value-loaded and are likely to be inimical to the spirit of inquiry. Reality knows no such thing as progress or retrogression — it knows only of *incessant change* — what appears as progress in some terrain is likely to be associated with 'retrogression' somewhere else.
>
> It is in this spirit that we include the next few lines in this essay. It may be mentioned, however, that such apparent value-neutrality is not inconsistent with the adoption of ethical and moral stands on specific issues relating to the practice of science.

As the conceptual space of an individual engaged in an act of creativity gets restructured, their occurs a major revision in her belief system when she sees things in a new light in something like a gestalt switch. Her account of things (within the expanded conceptual domain) changes, and she starts 'talking a new language'. This eventually culminates in the emergence of a new theory that an entire community of scientists accepts after adequate scrutiny and tests against hard evidence. Analogous to the case of the individual scientist seeing things in a new light, the entire fabric of the emerging new theory seems to differ in major ways from the earlier theory. Kuhn (and Feyerabend too; however, the two were not tuned to the same wavelength in their discourse) referred to this aspect of theory change as *incommensurability* (see [124] for background; see also sections 4.3.8, 4.4 in this chapter). The idea of incommensurability has been severely questioned on the ground that it reduces to denying the ground for a rational choice between theories succeeding one another and for the acceptance of one over the other as a better representation of reality.

> It may be noted that theory choice involves, literally, an element of *choice* or, equivalently, of decision-making. As we saw in section 3.1, decision-making is often not a 'rational' process. The same lack of rationality is involved in the matter of choice among succeeding theories. However, the 'irrational' factors involved in theory choice eventually give way to more rational factors getting into the act as a theory is examined by an entire epistemic community, when processes of critical justification attain relevance.

We repeat that the relation between succeeding theories is a *complex* one — it does not pay to be too hairsplitting and to sharply pose one viewpoint against another in the midst of complexity. What is often more rewarding is a genuine interpenetration of sets of ideas perceived to belong to such opposed viewpoints. It may be argued that the idea of incommensurablity certainly says something important about theory change, though it need not be taken in the literal sense to mean a fundamental incongruity — names are no replacements for *meanings*, and meanings in this complex world of ours interpenetrate one another to a degree that often makes a mockery of abounding philosophical disputes.

> *The contrary faces of scientific creativity:* Scientific creativity engenders the eternal conflict between *irratonality* and *objectivity* that leads to the generation of a huge tension within the viewpoint of *scientific realism*. The creative process in the individual is irrational in that it makes fundamental use of self-linked and implicitly operating psychological resources, but then, *simultaneously* it has to make a volte-face of sorts so as to generate an effective representation of reality — a sober 'scientific' representation, that is. Creativity in literature, arts or music, on the other hand, presents a different face.
>
> Incommensurability, in a sense, reflects this contrariness in scientific creativity.
>
> Here, incidentally, is a relevant observation by Baggott in [4]:
>
> *"Theorizing involves a deeply human act of creativity. And this, like humour, doesn't fare well under any kind of rational analysis."*

Incommensurability, indeed, *resides at the heart of creativity.*

We recapitulate here what was outlined in sec. 4.4. Incommensurability — as opposed to incompatibility — between a revised theory emerging from an earlier one through an act of creativity is understood by referring to the conceptual space which has a *multi-layered* structure (refer to [133]), where there ar multiple types of correlations between the concepts, in which the various different types can be imagined to be stacked one above another. For instance, the term 'malfunctioning of the circulatory system' contains a number of concepts connected by means of plain English language, overlaying which is a set of correlations having specific physiological connotations. Likewise, sets of scientific concepts are typically organized in a tightly knit layered structure. For instance, when one speaks of a 'particle', one can refer to a point-like mass or, alternatively, to a wave-packet in quantum mechanical theory, or even to a certain state of a quantum field, such as a photon in the context of the quantum mechanical electromagnetic field. All such multiple layers of interconnections result in an enormously complex structure of the conceptual network and, simultaneously, in a correspondingly stupendous complexity in the evolutionary dynamics of the conceptual space.

In an act of creativity, as distant domains get correlated, mostly by analogical heuristics, not all of the multiple layers of correlations get supplanted by newly emerging ones. Some of these layers are retained to a large extent, some are modified to a larger degree, while some new layers are formed, as a result of which the restructuring of the conceptual space assumes a complex character, and the newly emerging theory appears to be related to the earlier one in contrary ways. On the one hand, it shares to a large extent earlier layers of conceptual correlations and meanings characterizing the previous theory while, at the same time, novel layers of meaning make their appearance that give an entirely new texture to the revised theory. The earlier theory can be interpreted in terms of the emerging one by referring to the layers of conceptual correlation that are shared in common by the two theories, but the converse relation does not hold — essential aspects of the emerging theory cannot be interpreted by remaining confined to the concepts as correlated in the earlier theory.

This, in brief, is how incommensurability is inherent in the process of creativity.

In the present book, we mostly confine ourselves to considerations on creativity in science. In the case of fields such as literature, music, art and sculpture, incommensurability in creative work resides in the emergence of new genres that have different qualities, different structures, and different appeal to people, in whose minds different combinations of emotions and other psychic ingredients are evoked.

4.5.9 Creativity at all scales

Finally, we add the important observation that creativity can be identified *at all scales* in the dynamics of the conceptual space, which is inherently an exquisitely complex one. Thus, a creative act may connect distant conceptual domains and open up the possibility of a large number of meaningful correlations between remote concepts, thereby inaugurating a new theory supplanting an existing one, or else, may generate a whole new lot of correlations *within* a domain, thereby adding fresh layers in the conceptual network pertaining to that domain, giving it a new look wherein it is flooded with fresh ideas. In other words, a creative act may cause a remarkable enrichment of a conceptual domain rather than expand it in new directions. As an instance, the development of the theory of biological evolution by Darwin was without doubt a supremely productive creative act that not only changed the face of the science of biology but added a new dimension to the entire intellectual world of mankind. Compared to this, the development of the idea of evolution by genetic drift was an event of a smaller magnitude where the outlines of the theory of evolution were not altered in a major way, but led to a substantial enrichment of that theory, introducing a new perspective into it. We are all familiar with highly original solutions to ordinary-looking problems developed by friends and colleagues in workplaces and other commonly occurring situations — ones recognized as results of *insight*, a phenomenon that we discussed in sec. 4.5.3 earlier. In the present context we may draw a distinction between insights that generate new ideas from ones that fail to do so, and describe the former type as instances of creativity, though on a relatively limited scale.

Put differently, the creative process can take place within the confines of a conceptual domain of limited extent, in which fresh layers of correlations are established across sub-domains, signaling a conceptual restructuring within the confines of the domain, significantly altering its multi-layered structure.

Creativity, in other words, can be compared to a special type of evolution of a complex network where the creation of fresh links between clusters of nodes in the network results in further links being set up in and around these clusters and fresh layers of correlations being added to extant ones — it is this phenomenon that makes it resemble a process of self-organized criticality. Such self-organized criticality can occur *on all scales* in a complex network.

We now attempt to *sum up* all these various aspects of creativity that we have dwelt upon.

4.5.10 Summing up: affect and inference in creativity

In this part of the present chapter, we have had a close look at creativity, mostly in the context of *scientific* creativity, and have tried to put together a framework aimed at understanding and analyzing this elusive phenomenon.

At a fundamental level, creativity consists of a complex mix of *inferences* — one of a very special kind — involving a *restructuring in the conceptual space.* The latter has a tangled and nested complex structure that can be notionally represented as a network of concepts correlated with one another by means of a web of *beliefs*, including a huge repertoire of *heuristics*.

Beliefs and heuristics active in the mind of an individual have both a shared and a self-linked aspect, where the former refers to beliefs possessed in common by larger social communities while the latter are specifically associated with the unique developmental history of the person, making up a component of her psychological self.

Concepts are correlated with one another by means of these beliefs and heuristics, where the correlations keep on being built and rebuilt by means of an incessant succession of inferences, that results in an exquisitely complex evolution of the conceptual network. An inference, launched from one or more initial premises — each made up of a number of correlated concepts to start with — proceeds through a succession of consecutive correlations established by means of beliefs and heuristics, to finally end up in a conclusion, once again comprising of a number of correlated concepts. In the process, shared beliefs are made use of in setting up a simple chain of intermediate inferences, where the simple succession is punctuated with a number of decision junctures at each of which a choice is to be adopted among more than one disparate alternatives. This choice, and the entire chain of intermediate inferences is made possible by means of the affect system. Affect constitutes the core of emotions that plays the all-important role of reducing varied and diverse experiences to a 'common currency' corresponding to a single value-dimension, analogous to a number with either a positive or a negative sign, and a magnitude.

The reduction to a single value-dimension provides the pivot over which the entire complex of psychological processes of an individual keeps going smoothly, managing conflicts and clearing blockades. A disruption in this causes a major disintegration of the entire mental fabric of a person. The affect system, emerging from a long evolutionary process, works equally well for the shared as well as self-linked psychological resources of an individual, and gets involved at every stage in the making of an inference. In this context, we mention that creativity in such genres as poetry, music or art involve the so-called self-linked psychic ingredients (like, for instance, the drives, desires, aspirations and frustrations of an individual) in the place of the cognitive ones like beliefs and heuristics though, once again, emotions and affect are involved as essential components in the creative process.

Scientific creativity involves a restructuring of the conceptual network whereby a cluster of concepts and beliefs, all tightly correlated with one another so as to form a theory, gets linked with a *remote* set of concepts, where the latter centers around one or more phenomena not explained by the theory in question. As a result of the restructuring, there takes place an extensive belief revision whereby the correlations among the previously existing cluster and those linking the remote clusters get modified and a new set of beliefs appear, being consistent with the newly established correlations. The previously existing domain in the conceptual space, to which the earlier theory was applicable now expands into a larger domain, in respect of which the modified correlations along with the newly established ones constitute a revised theory.

An inference takes place in some *context*, where the latter constrains the process of setting up the chain of intermediate inferences leading to a conclusion. Without such constraint restricting the exploration of the conceptual space, the making of an inference would have been an intractable process. This is analogous to the set of constraints necessary to assemble a successful search program in computation. However, a creative exploration needs to connect remote domains in the conceptual space, for which the constraints are to be relaxed so that a more effective and free exploration may be possible. This phase of free exploration in a creative act is commonly referred to as 'incubation' where the mind is predominantly in the default mode. Such a phase of exploration in the default mode is also involved in what is referred to as 'insight' where, however, the remote concepts correlated in the process belong to a relatively restricted domain. What is more, creativity is characterized by a 'domino effect' where the links set up between remote concepts quickly lead to a further set of correlations and there finally appears a whole new domain to be linked up in a major restructuring, with the emergence of a revised and expanded theory. The revised theory is given a solid foundation by an entire community of scientists on the basis of critical examination against evidence, while eliminating possible inconsistencies.

Creativity involves an apparent *non-determinism* in that it cannot be explained rationally in terms of an accepted set of principles based on a theoretical framework existing prior to the conceptual restructuring. Such non-determinism, arising from a lack of predictability, characterizes acts of free will as well, where both are essentially based on self-linked psychological resources. We recall that these self-linked resources are specific to the developmental history of an individual and cannot be accessed by means of shared inter-subjective rules. Further, the involvement of heuristics in the conceptual exploration makes the latter a history-dependent one since the formation of new heuristics and their activity in an inferential process take place without the inertia characterizing the more emotion-laden beliefs. Such history-dependent evolution of a complex system is known to generate an effective randomness — one that characterizes an act of creativity.

1. In this context, we mention a revealing passage from [112] where the noted mathematician and mathematical physicist David Ruelle refers to the self-linked psychological resources of Newton and speculates on the role that these played in the great creative achievements of that intellectual colossus (italics ours):

   "The interplay between Newton's various intellectual interests is fascinating. These interests range from the greatest achievements in mathematics and physics to disreputable speculations (by present-day standards) about alchemy, history, and religion. It is tempting to apply censorship to Newton's intellectual production and decree that some is good and the rest better forgotten. If, however, we want to understand the process of intellectual creation in Newton's mind, we cannot forget his disreputable speculations. In his desire to grasp the meaning of the universe the research on the prophecies or alchemy was not less important than the work on gravitation or differential calculus. A lot, obviously, remains to be understood on how Newton's mind functioned."

2. The time evolution of a complex system is fundamentally history-dependent and unpredictable, where the causes underlying these features may be manifold. Generally speaking, a number of characteristic features of the time evolution of a complex system arise from *remote causes* (recall the fat-tailed phenomenon) — remote in space (the 'phase space', i.e., the space made up of the states of the subsystems) and *time*.

Of fundamental relevance in the process of correlation of remote concepts are heuristics based on *analogy*. In a broad sense, every inferential activity is in the nature of an analogy since, speaking in general terms, an analogy is a mapping from a source concept to a target concept. Analogies have the ability to connect remote concepts as much as they can connect contiguous ones in the conceptual space. Those connecting remote concepts appear as deep ones since these throw light upon additional clusters lying on the routes between the source and target concepts. Analogies, in other words, constitute the very essence of creativity, making possible major restructurings in the conceptual space by means of the cascade effect found in self-organized criticality.

Scientific creativity makes possible the emergence of a revised and expanded theory replacing a previously existing one that has failed to explain some particular set of facts by means of the conceptual correlations characterizing that earlier theory, where these correlations were confined to a restricted domain in the conceptual space. Such episodes of theory revision are referred to as conceptual revolutions and raise the question as to whether and in what sense the revised theory can be said to be *incommensurate* with respect to the earlier theory. As we have seen, incommensurability is related to the multi-layered structure of the conceptual space, where not all of the layers get modified to an equal degree in the conceptual restructuring while, additionally, new levels of correlations (and *meanings*) are set up among the concepts in an expanded conceptual domain.

Acts of creativity can occur on various *scales* in the conceptual space — not only connecting up distant domains in it, but connecting sub-domains within a domain, or enriching the correlations within a domain and generating a new set of concepts along with a new layer of correlations within it.

Chapter 5

The soul: examining and reconstructing the self

In this chapter we first look into the issue of *free will* where the role of self-linked psychological ingredients will once again be seen to acquire relevance. The fundamental idea, once again, is to distinguish between the necessary and the contingent — the former is predictable in terms of commonly shared rules, while the latter is not. Behavior determined by self-linked psychological resources falls into the latter category, i.e., appears to be contingent to us, since the causal links determining it remain unknown. This feature of non-determinability is shared by the exercise of free will with the making of inference (inductive inference in particular) where choices are to be made, based on self-linked beliefs and heuristics or, equivalently, with rules of various degrees of generality including, as a limiting case, ones specific to an individual.

Speaking of beliefs and rules specific to an individual, a case can be made that even these entail a certain *causal* operation, even though of a kind not to be captured by logical principles extrinsic to the individual concerned (ones invoked by a psychologist, for instance). Is free will completely free in the sense of being random? Even though such randomness is not to be found in the exercise of free will, we will see that the issue of 'freedom' is ultimately linked with the ability of an individual to look at and reconstruct her self, including her repertoire of self-linked psychological ingredients.

This brings us face to face with the *soul* (refer back to sec. 1.14 in the introductory chapter of the present book), where we will find the role of randomly operating factors in making an individual or a social group critically examine the self (either of the individual concerned or of the group, as the case may be) and reconstruct it.

5.1 Free will: the problem of unpredictability

The issue of free will is known for the huge volume of literature it has produced on the question of whether human will is truly 'free' or is under the shackles of 'determinism' (see [49], [71] for background). Is freedom of will *compatible* with determinism, or is it not? We will see that the issue of non-determinism in free will is intimately linked with that of *unpredictability*. To anticipate the considerations that follow, free will involves a contrast between causality and unpredictability that reside side by side in it.

In the course of the incessant mental activities of an individual, there arise junctures when a person seems to have before him alternative courses he can choose from. These alternatives can be a number of possible actions or of possible streams of mental activity as well.

A footballer, in the course of engaging in his daily fitness program, skips the gym and, instead, takes the option of doing a longer-than-usual stretch of jogging. He can be said to have *chosen* the option of jogging over the other option of stepping into the gym, and this can be cited as an instance of exercise of free will.

A group of students, on getting a day off from their classes, leaves their hostel building and heads to the nearby sea-beach for an extended session of beach volleyball, while one of them left the group and chooses to take a bus home to spend a quiet day with his parents and sister. He seems to have obviously made a choice, and has exercised his free will. As for the group of his friends headed to the beach, an individual member of the group cannot, perhaps, be said to have made a choice, though the group as whole has certainly made one since it could equally well have chosen to go to a movie, or to an open-air restaurant, or to indulge in any of a number of other possible group activities.

Sitting in my study on a rainy evening, I engage in anxious brooding over the possibility of my son meeting with some kind of a mishap during his cross-country drive in inclement weather in complete defiance to my earlier suggestion of desisting from the drive. I had started off with thinking about a political commentary that a local magazine had requested me to write for them, but then somewhere along the line I 'chose' to engage in the fitful brooding in preference to pursuing my political thoughts.

We will not continue with further instances of exercise of free will for obvious reasons, but will only note that, as the above examples illustrate, free will is commonly supposed to be 'free' in the sense of being associated with the exercise of *choice* from among alternatives. At times, the question of *intent* comes up — does a person, in the exercise of free will, in some sense *examine* the options before him and make a choice by intent? The instance of the homesick student is a case in point since it is likely that he chose to spend time with his family after deliberating upon the other option of accompanying his group of friends. The instance of the anxious father, on the other hand, is indicative of no such deliberation while that of the footballer doing the round of jogging may be supposed to have involved some marginal deliberation, if at all.

The matter of alternatives, the matter of choice, and the one of deliberation and intent — these are some of the supposed features associated with the exercise of free will. In the following, we will examine all these numerous features and will venture into a number of deeper issues that help us appreciate the specific sense in which free will can be said to be 'free', and the sense in which it can be said to involve a 'choice' from among 'alternatives', recalling the case of inductive inference where too an analogous question of choice comes up. In the process, the question of intent will get settled as being somewhat of a secondary one.

5.1.1 Alternatives: the yardstick of assessment

It might seem that the question of alternatives needs no clarification. That the exercise of free will involves a choice from among alternatives doesn't seem to be in doubt, and the common perception is that it is the *other* question of compatibility with determinism that deserves a greater degree of scrutiny. We will see in our own way what that other question involves, but will not let go of this issue of alternatives and the exercise of choice as one of lesser relevance.

The first question we will raise is, alternatives according to whose assessment? And, choice by whom, or by which entity? The latter question may sound silly, but we will see what sense we can impart to it. We turn, then, to the question of which agent identifies the possible alternatives involved in an instance of willing (we mention that this is the abbreviation for 'exercise of free will' that we will use for the sake of brevity)? Significantly, a person's action is often described as a choice from among possible alternatives not by the person herself, but by *others* looking at and assessing the action. True, the person herself may at times assess her own action as one involving a choice, but here again her own list of the alternatives involved would often not agree with the list that some other individual may come up with. A mother preparing to serve her child with dinner is unlikely to agree that her act is one of making a choice from among possible alternatives, but the child's father may think otherwise. He may seriously entertain the idea that a possible alternative would be for all three of them to sit watching a TV show that the child is fond of before she (the child) is made to dine and retire for the night.

It is significant indeed that different individuals are likely to come up with different lists of the possible alternatives from among which a person makes a choice while exercising her free will. Literally speaking, one can think up of an *infinite* number of alternatives to a supposed act of willing by an individual. When considered without reference to some given *context*, a person may be said to have made a choice from an infinity of possibilities. The loving mother *could*, for instance, go jogging in the night instead of serving dinner to her child. So, clearly, something more specific is involved in assessing the possible alternatives in an exercise of free will. As we understand it, that 'something' lies in the *context* in which the willing is enacted, and the *rules*, or the *standards* invoked in making an assessment of the possible alternatives. We will take up the question of standards first.

When a husband, after some deliberation, agrees with his wife's action of making their child sit for dinner and sending her to bed without further delay, he is clearly invoking the standards that apply to a child and not to an adult, since the standards by which he would assess his wife's decision (an act of free will on her part) regarding the time for the two of them (wife and husband) to sit for dinner, and the possible alternatives to her choice of the dinner time (sitting through a TV show first, for instance) would be quite different.

Invoking some standard of assessment, often involves the use of some 'rules', or some kind of 'logic' for the same. Suppose that, on questioning your son regarding an incident in the playing ground, you learn that he has beaten a friend of his, which makes you think as to whether to be strict or lenient in punishing him (your son, that is). The choice favoring a heavy punishment is possibly based on your earlier injunction to him that he should never hurt another person. On the other hand, a light punishment would be indicated on the 'logic' that this world is a tough place and everybody has the right to fend for himself when involved in a dispute. A completely different 'logic' points to a third and distinct possible course in the exercise of your free will, namely, the one that friendly but firm and frank discussion with a child is always to be preferred over punishment and, on adopting this point of view, you may then ponder over alternatives as to how best to carry out the discussion.

But, these are all in the nature of *possible* principles that might guide the mother of an errant child in dealing with him. An outsider looking at the situation can conceivably think of still other possibilities and still other 'principles' that she (the mother) can choose from in exercising her will. The relevant question that comes up here is, does the mother actually make a choice from among all these various relevant principles and then decide which course of action to adopt? It would be a very exceptional mother indeed who would deliberate over all these possible approaches and then decide upon how best to deal with her child. More likely than not, her action (the result of exercise of her free will) would involve only a minimal deliberative component and would come about, to all intents and purposes, almost spontaneously. However, though apparently spontaneous, it would at the same time involve a complex interplay of quite a few factors, which she would mostly not be aware of.

At this point, we will not pursue further how an act of willing may be interpreted by an onlooker as one involving a choice from alternatives and how the onlooker sets out to invoke rules of logic (not any strict or rigorous logic but some loosely patched up 'logic' that we commonly employ in daily acts of willing). Indeed, there can be thousands of possible interpretations of an act of willing when looked at from the point of view of an *external* agent. Put differently, an act of willing is done in some *context*, and the context is judged in various different ways by different persons.

The context in which an act of inferring or willing is done has an *external* and an *internal* aspect to it, where the former is, to some extent, independent of the person involved, being more or less the same for different individuals that may be imagined, each in turn, to be engaged in the act. The internal context, on the other hand, differs vastly from one individual to another. Having said this, I must hasten to added that the separation between the external and the internal does not come cut and dried, but this I am inclined to gloss over for now.

This is of fundamental relevance for our purpose *: given the same external context, it is almost certain that different individuals would exercise their will differently because of the stupendous difference in their internal psychological resources.* And, it is *this* difference that makes us say that an act of willing is a 'free' one, since *this* is really what makes the act of willing look like a 'freely adopted' choice from among alternatives.

We repeat that the internal psychological resources of an individual make up a system of great complexity. And here again, one part of the system can be said to be common to various different persons and groups while the *remaining* part is of crucial significance for our present purpose: it constitutes the *self* of an individual (refer back to sections 1.9, 2.7).

On a pleasant evening I find myself taking a walk along a crowded street and, on finding a bookstore I favor, enter into the display room for browsing through their latest acquisitions. My younger brother would, under similar circumstances, visit a nearby restaurant to relax and spend time. My adolescent son, on the other hand, would take the 'option' of strolling *to the riverside and meet his group of friends there.*

It is conceivable that I deliberated between the alternatives of the bookstore and the restaurant before opting for the former, and that my younger brother wold also similarly weigh his options before acting the way he is supposed to. However, none of us would even think of the other 'option' of walking off to the riverside. As we mentioned, there can conceivably be an infinite number of alternatives from among which one can be said to 'choose' while exercising his free will, but only so many of those are actually relevant in a given context.

For instance, in the case of me entering the bookstore, my act of willing can be described as one of making a choice between two (or more, including a few others that may have been available under the given circumstances) alternatives, and my brother's act would also be assessed similarly. But the possible alternative of strolling to the riverside (known to be frequented by the younger and more boisterous lot of the town) would not count in our case. We two being of similar general disposition (similar cultural and family background, similar age), the *actual* choices made by us are counted as possible alternatives from among which each of us would chose *individually*. What is important here is that different individuals with similar dispositions would act differently and *this* difference (among the actions of a number of individuals placed under similar circumstances) is what is said to constitute the set of possible alternatives. It does not matter in this description whether some particular individual, in exercising his will by acting in some particular way, actually weighs between these alternatives. This is why the question of intent may be assumed not to be greatly relevant in numerous acts of free will.

In other words, the matter of 'choice' is, in a manner of speaking, funny and strange. Suppose that, in a given external contest, three persons X, Y, and Z, of similar disposition, in exercising their will, perform acts A, B, and C. The question that comes up is, does X, in performing the act A, actually weigh the options A, B, and C as possible alternatives? Why should X be said to have 'chosen' between A, B, and C in actually performing the act A? What does it matter to her how some *other* individuals (Y, Z) might act under similar circumstances?

Looked at this way, the matter of ‘making a choice from among alternatives’ does indeed appear funny and strange. Such a ‘choice’ only *appears* to be incompatible with determinism (but may not be actually so) since the ‘alternatives’ cannot really be said to have featured in the mental processes of X: what X does as an individual can have nothing to do with what other individuals (Y, Z) *might* possibly do.

A deeper view to the determinism issue lies in observing that the other individuals Y, Z (and still many others) *potentially* reside in X herself. Put differently, the psychological resources of an individual that make up her internal context in the act of an inference or exercise of will can be looked upon as a composite whole of *two* parts — *one that she possesses in common with many others, such as Y, Z*, and *the other that is unique to her*.

If I care to examine my (unjustified) belief that women are inferior to men, I will find that many others in my community or in the world at large (unfortunately) share the same general belief. Be that as it may, my own belief is colored with other beliefs specifically my own, along with memories, experiences, emotions, and value judgments, all making up a complex network unique to myself. In other words, any and every psychological resource that I may have within me has a dual identity — one that is shared, and another that uniquely belongs to my own *self*. The shared part of my complex psychological system looks at things one way, while the other part that belongs to my own self ‘perceives’ otherwise. Of course, it is only notionally that the two ‘perceptions’ can be distinguished and, at the end of the day, there is one single perception that differs from person to person. The perceptions of X, Y, and Z looking at a tree certainly have something in common while each has memories and emotions uniquely her own that make one of the three pass by the tree without looking at it a second time, another to ponder briefly over its sublimity and grandeur, and the third to go sit under it, recalling nostalgic memories of sitting under a similar tree in bygone days, perhaps in agreeable company.

Significantly, it is the uniquely individual part of the internal mental resources that makes each person act the way that she actually does. Two distinct individuals might appear to act the same way — perhaps with both spending some time sitting under the aforementioned tree though, here again, the two may have entirely different trains of thought running in them.

In reality, the 'alternatives' exist only as abstract ones when one looks at some given external context from the perspective of the shared mental resources of an inter-subjective nature alone, with total disregard of the uniquely individual resources that make a person actually follow some course of action with all its associated idiosyncrasies. Looking at a tree standing in front of me I can think of many different things in virtue of my shared world-view such as how a tree is a gift of nature, or how green and densely grown its leaves are or, say, how much of cellulose there is in its trunk, but none of these in itself can be responsible for the particular way I respond to the sight of the tree. As we saw, even when two different individuals appear to respond similarly by taking time off to go sit under it, the *details* of their response are likely to differ markedly. Even if an individual does make some deliberations of an entirely interpersonal nature, relying only on what is known to be shared by others in the community, the specific aspects of her action cannot be accounted for in terms of those deliberations alone.

The actual action engaged in by an individual may have little to do with such abstract deliberations of an interpersonal nature, and is more likely to result from memories, beliefs, and propensities generated in the course of her own developmental history. As we mentioned, it is too much of an abstraction to even notionally isolate shared interpersonal aspects of an individual's mental resources from the uniquely individual ones and, in as much as one does engage in such an abstraction, one may come up with the possibility of deliberation among alternatives, but such deliberation can have only a marginal role in the actual realization of an action by way of exercise of will. In other words, willing is incorrigibly conditioned by the *self* of an individual. It then becomes necessary for us to examine as to what constitutes this strange and deeply enigmatic entity that we refer to as 'self'. This has already been done (within the limits of our understanding) in sections 1.9, 2.7 of this book — in the following we will recall that content as made necessary by the present context.

5.1.2 Constraint and freedom

When one speaks of 'free' will one actually means freedom from constraints. It needs to be appreciated that, significantly, the constraints are *not* those relating to mental processes groaning under deterministic shackle.

As a judge in a legal trial prepares to deliver her verdict, she feels heavily constrained, and she herself would be the first to concede that the constraint comes from the mass of evidence, the arguments put forward by the learned counsels, and the meticulously formulated clauses of the penal code, not to speak of the broader principles of jurisprudence and human justice. She is to refrain from basing her verdict on her *personal* beliefs and emotions as best as she can and has to go by the precisely drawn documents now piled up before her on her desk.

Contrast this with the preparations of the lawyer representing the accused in the case as the trial was about to begin. She was also constrained by a good number of things — by the clauses of the penal code, by her commitment to the accused, by her legal ethics, and so on. But she was a bit more free to make use of mental resources specific to her own self. She could, for instance, rely on her personal hunch to discard one prospective witness and ask a somewhat less reputable person to the witness box. What is more, she was, to begin with, free to make use of her reasoned considerations *and* her gut feelings based on her emotions even to deny her services to the client altogether. As the trial proceeds, everybody involved with it gets more and more constrained by the mounting evidence and arguments, and finally the judge finds herself before the mass of piled up documents heavily constraining her away from making use of her own personal feelings.

Of course, as in everything else in this world, constraint and freedom co-exist in real life, and neither of these two is to be found anywhere in pure form. The judge is more constrained (and less free) while drawing her verdict than the lawyer preparing for her defense of the accused, but even she is not absolutely constrained, simply because the mass of piled up evidence along with the clauses of the penal code and the principles of jurisprudence may not *unequivocally* lead to a verdict. There may still be some scope for *interpretation* where she can possibly invoke her own personal judgment to some marginal extent, which is why the case may eventually go to a higher court. Beginning from the time before the accused was committed to trial up to the settlement in the higher court, freedom of choice (in the context of the verdict) goes on getting reduced progressively.

A citizen is 'free' — unfortunately so — to indulge in drug abuse, a great deal of medical and social admonition notwithstanding, precisely because he ignores the latter and commits himself to his inner propensities. He can be said to have made a 'free' choice precisely because a second individual with similar background can possibly make an equally free choice of saying no to drugs. And the overriding fact is that, *both* are subject to the iron shackle of determinism, if that is how one chooses to describe the operations of a billion or so of neurons that make each of them do what he does.

Among two students of art college, one chooses a life of poverty and deprivation while committing himself to the practice of art as a means of human liberation. The other chooses to go in for commercial art in an ad agency so as to be reasonably certain of a life of affluence. Knowing that the two careers are *possibly* open to an art student and that some adopt one of these while some others adopt the other, it is commonly said that each of the students 'chooses' from among the two alternatives. In reality, however, there is no reason why the 'choice' of each of the students is anything but the outcome of a deterministic process based on neural interactions. The aspect of determinism eludes us because the workings of neural and psychological factors relating to the *selves* of individuals often go unacknowledged since these are hidden even from our own awareness. Significantly, the neuronal processes relating to the self of an individual do not take place in defiance of determinism or *causality*, but the associated mental processes cannot be determined by an external agent looking at and analyzing her action.

In summary once again, free will involves a 'free' choice only in the context of (and from the point of view of) inter-personal or universal standards of judgment while it is, at the same time, 'determined' by the operations of psychological resources hooked up to the self of an individual.

The rules of mathematical logic, along with the those of *set theory* are said to constitute a universally valid set of principles precisely because these are constructed so as to be independent of contingent standards of judgment. When a mathematician sets out to present the proof of a theorem she doesn't enjoy any leeway to deviate from these universally accepted rules. Likewise, the principles of evidence, the clauses of the carefully constructed penal code, and the principles of jurisprudence are supposed to constitute a similarly forbidding set of constraints within which a judge prepares her verdict in a trial. Even so, the rules of mathematics are more universal than those constraining a judge, precisely because the latter are designed to codify a set of principles relating to *human* behavior, while mathematics abstracts away from all contingent human factors.

Significantly, even mathematics is not as unrelenting as it is thought to be. There remain foundational questions in mathematics that indicate that no absolutely foolproof structure can ever be had in this world. Alternative approaches are possible within the world of mathematics concerning the interpretation of mathematical objects and mathematical truth (see, for instance, [56]).

No less significant is the fact that a mathematician in the process of *discovering* a mathematical theorem faces a completely different terrain as compared with her colleague who is preparing a proof of some other theorem for publication in a journal. Proof or justification is a matter distinct from discovery, because the latter is primarily an act of an individual in the capacity of her own self and is only ultimately one that is to conform to strictly laid down principles. Justification, on the other hand, is a process where, ideally, there is no leeway for the self to operate.

The process of discovery of a mathematical or scientific truth is referred to as one of abduction where a prospective theorem or scientific theory first appears as a conjecture and is then justified in terms of strictly laid down principles.

5.1.3 Discovery and justification

In this section we briefly recall the considerations made in chapters 3, 4 regarding the use of shared and self-linked mental resources in inductive inference, including the act of hypothesis formation in science. It is hypothesis formation that is basically of relevance in acts of (scientific) creativity. Inductive inference and hypothesis-making involve elements of 'choice' and decision, where it turns out that there is an underlying similarity with the 'choice' involved in acts of free will. In both the two cases (of inductive inference and exercise of free will), there arises a lack of predictability because of the role played by self-linked psychological resources. In neither of these, however, this lack of predictability is associated with any lapse in causality.

This section is mostly devoted to a recapitulation of ideas outlined in earlier sections (in particular, sections 3.2.1, 3.2.2) of the book.

The *context of discovery* is commonly distinguished from the *context of justification* [113], where the former is given a wide berth in the philosophy of science on the ground that it belongs to the murky terrain of psychology. Justification, on the other hand, is supposed to be a 'logically' well-defined process where there are strict standards involved. While mathematical justification is indeed a highly regimented exercise, the same cannot be said of theories in natural science, where one has to depend on confirmation by experimental observations coupled with 'internal consistency'. However, observation itself is *theory-laden* and the evidence of consistency is also not of absolute reliability since consistency is judged by existing standards, these being once again tied to current theoretical perception. In the end, scientific theories, however strongly justified, constitute intelligent *interpretations* of certain cross-sections of reality [82], which is precisely why these undergo spectacular revisions from time to time on the strength of 'evidence' from newly explored cross-sections and of revisions in the existing modes of thought.

But this is not the place to point a critical finger to possible fault lines in the justification of theories in the sciences because, going by current standards of justification, a finished and successful theory can indeed be accepted as one that, at least provisionally, has vindicated itself. A theory is preceded by one or more *conjectures* or hypotheses that acquire the stature of a theory only when they have successfully passed through the rigorous process of justification. Justification is usually a social process where a community of scientists made up of one or more working groups examine the hypotheses with critical eyes.

While the process of justification follows some strictly formulated grammar, that of *discovery*, in comparison, is shrouded in mystery. In contrast to the process of justification, a conjecture is formed more often than not in the mind of an *individual*. Even so, a conjecture is not plucked out of thin air. There occurs a two-fold process within the psychological terrain of the individual that can be sketched in broad outline even as that terrain is too misty to allow us a clear vision.

What is significant about the process of discovery, i.e., the one of formation of a conjecture, is that it must involve some measure of justification during this very process itself. A conjecture is formed by a process of inductive inference, and is essentially in the nature of a guess. But it is, at the same time, an *informed* guess. A scientist makes good use of the huge repertoire of the knowledge base stored in her mind in her endeavor to arrive at a reasonably good conjecture. But that, in itself, is not enough for arriving at the conjecture — otherwise, any and every reasonably well-informed colleague of hers could come out with it.

Along with the knowledge base, the scientist also makes use of an enormously large repertoire of *heuristics* [82] stored in her mind, where a heuristic is a fragmentary belief, of the nature of half-formed knowledge or some rule of thumb that has earlier been found to be efficacious in arriving at some sound conclusion. A good chess player makes copious use of such heuristics in addition to the set of memorized chess moves previously analyzed and commented upon by experts. The heuristics in her mind are made of vaguely formulated and elusive patterns of moves that perhaps she herself cannot set down clearly in writing but ones that fleetingly circulate in her mind, unconsciously driving her on to make some surprisingly spectacular move in a difficult situation. She, of course, does not make use of these half-baked patterns of move without due deliberation — once a move comes up in her mind as a result of processes she is only vaguely aware of, she sizes it up by looking hard at the current position of the board and gauging its consequences several moves ahead. This is how discovery and justification go hand in hand in the mind of the individual engaged in the making of a hypothesis.

And, this is where a champion chess player differs from her lesser opponents. While an opponent may have an equal or even greater command of published chess literature, the champion enjoys a clear edge in the profusion of *tacitly* operating processes involving vague and elusive patterns of chess moves — ones set in motion by the current challenge on the chess board in front of her. These tacitly operating resources, along with a huge store of chess heuristics she has amassed in her mind, are the ones that can be said to be specifically her own that makes possible the one spectacular move that her opponent could not even imagine being challenged with.

Given a challenging situation on the chess-board, a good number of parallel chess games get played out within the unconscious mind of the champion in which the heuristics are made use of — ones in which a good number of moves are tried and discarded, with one surprise move selected at the end which she now examines critically in the context of the actual position on the board. This unconscious process of sifting and selection is what the psychological *value-system* is so good at. The value-system operates in assigning positive and negative valuations (in terms of vague feelings of satisfaction and discomfort) as the mind engages in parallel processes of counterfactual thinking (see [64] for the neuropsychology of counterfactual thinking).

As mentioned earlier, the psychological value-system involves the activity of specialized neural assemblies aided by a number of chemical neurotranstters, and is of early evolutionary origin. It attaches a positive or a negative weight to the affects, emotions, and feelings associated with memories, thoughts, or perceptions of current experience, and is an integral part of the self-network of an individual. It is important to note that it is the value-system that is responsible for the ability of the mind to sift through alternatives when shared or explicit principles fail to do the job. It is precisely this gap, where the knowledge base of an individual or her store of inter-subjective beliefs does not help her reach a well-formed decision, that her self-based resources come forward, with her personal store of heuristics and beliefs, hooked on to her self-network by means of emotions, getting involved in the act of inference-making (refer back to sections 3.1, 3.2.2 in chapter 3) and making possible a logical leap when necessary. The self-linked beliefs, along with the positive or negative weights brought to play by the value-system, help her weigh between alternatives in terms of a common denominator — a task where the more 'logical' and justified of her beliefs may have failed her.

In summary, it is the set of self-based beliefs, emotions, and feelings that plays proxy to the reason of an individual helping her in making logical leaps in acts of inductive inference, in which the psychological value-system acts as an 'internal censor' ([82],[94]) in a continuing process of justification. Once the self-based psychological resources succeed in playing the role of a prop to her in crossing the logical gap where her knowledge base and her shared world-view were found lacking, she once again invokes these more worthy and proven resources, now in more focused awareness, to check and test the tentative hypothesis formed in the earlier act of the logical leap. If she now finds it lacking she discards it and starts the entire process afresh, this time armed with the negative lesson of the earlier failure.

This is how the cognitive act of the making of a scientific hypothesis can *possibly* occur in the human mind. All along, the scientist keeps on exploring larger and larger parts of her conceptual space ([14], [54]) and making new combinations and transformations in the conceptual elements in that space so as to arrive at some novel concept around which the hypothesis is formed. The knowledge base, the shared world-view, and the self-based psychological resources, all these are made use of in effecting a transformation that may, at the end of the day, prove to be of great value in the subsequent stage of theory building.

What is true of a momentous act of hypothesis formation in scientific exploration, is also true, in essence, in more mundane acts of inductive inference where, once again, clues acquired and retained tacitly by an individual in earlier experience, coupled with the operation of the psychological value-system, may prove to be effective in the formation of an inductive guess, to be subsequently tested by a more reasoned approach, perhaps based on more solidly grounded evidence. This, for instance, is how a physician, on examining a patient, often arrives at a tentative diagnosis which she subsequently tries to confirm by means of pathological tests of various kinds.

In other words cognitive acts such as a simple one of inductive inference or the formation of a novel scientific hypothesis make use of knowledge base and a world-view shared by a community or society at large but, at the same time, draw from self-based person-specific resources including the all-important value-system that acts as an internal means of justification called into play even as the act of guessing or forming a conjecture is carried out in the mind. It is the value-system that is ultimately responsible for whatever *choice* the inferring mind adopts from among possible alternatives.

5.1.4 Free will: the causal expression of self

Now is the time to spell it out: *free will is the causal expression of the self of an individual.*

And, *free will is free because it isn't constrained by shared principles, rules, and world-view.*

Free will is not in conflict with determinism, whatever that term is supposed to mean. In common parlance, determinism generally stands for the universality of the operation of causal principles. Looked at from this point of view, free will is not free of causal linkages, but it is free of the constraints caused by factors outside the private periphery of the self. To be sure, every action of ours, every thought that flows in our mind in some given external constraint, have both types of causal linkage — factors shared in common by a group, a community, or a society, and those that are specific to our self-network. Among these actions and thoughts are the ones we refer to as products of exercise of free will because as we attempt to trace back the causal links to these we seem to 'lose the trail' at some point, failing to explain their occurrence in terms of our known stock of knowledge, accepted principles and shared beliefs. The latter prove inadequate as explanatory factors precisely because various different persons under similar circumstances come up with different actions and different thoughts.

It has been observed [106] that free will is an *uncaused* decision by the self of an agent. In the present essay, we attempt to examine the *self* to uncover the causal roots of free will that is only apparently 'uncaused'.

5.1.5 Lingering questions: where to locate freedom

What we have said so far seems all very fine. But it gives the feeling that there is really no choice involved in the exercise of free will or, for that matter, in the making of an inductive inference either. All that is achieved in what we have said is to seemingly explain away the apparent incompatibility of free will with determinism by invoking the complex machinery of the self.

We thus seem to have restored the lost threads of determinism (or, more precisely, of causality) in the form of the workings of the self-network. The real question, however is, looking at the action or thought of a person under given circumstances, could that person have acted or thought *otherwise*? While all we have said above seems to engender an answer in the negative, we will find in the course of this essay, that there is *some* sense in which one can respond to this question in the affirmative too.

We thus ask the question — is there, then, *no* element of genuine choice involved in the exercise of free will? Let us first recall what the earlier pages of this book tell us in this regard, pruned of all background stuff.

5.1.6 Free will and inductive inference: recalling the basic ideas

This section, once again, is devoted to the recollection (for the sake of clarity of presentation) of material presented in earlier sections of the book (and, especially, in this chapter) along with a few clarifications.

We have referred to the knowledge base stored in our mind (*three plus three is six; all bodies attract by the force of gravitation; Shakespeare wrote Othello;* $E = mc^2$*; the doctor is scheduled to examine my child in the evening*; ...), our memories, our beliefs, emotions, and feelings, and our neuronal value-network. To this list we could add such things as our fantasies, desires, drives, and such other ingredients that make up our repressed self, though these will not feature explicitly in the present chapter.

Among these, the knowledge base is, to a large measure, detached from the self-network and, in particular, from our emotions.

However, even our knowledge base has links to our self-system since it depends on a person's level of commitment, her rigor and intensity of past training, her temperament, her tacitly acquired mode of thinking, and a host of such other factors associated with her self. Thus, the individual items of knowledge and, by the same token, the total store of knowledge, is tied to the personal biography of an individual. It may so happen, for instance, that a scientist has, through personal commitment, interest, and diligence, read up a huge mass of material relevant to her subject, far outstripping other colleagues of hers in this respect. However, when she recalls and makes use, for her research purpose, of some particular item belonging to her knowledge base her doing so may not evoke affects or emotions. Of course, a great deal of emotions is roused if the said act helps her solve some long-standing mystery in her subject area, but that is a different matter altogether.

Along with the knowledge base, there are shared beliefs that can also be assumed to be detached to a considerable degree from the self-network. Some beliefs and principles are almost universal (*everything is determined by the initial state of the universe; Einstein was a great scientist; one should try not to hurt others*), some are less so but still shared by large groups of people regardless of their truth value (*immorality is to be dealt with firmly; sexual excess will be followed by failing health; intelligence is linked to the genes*). On the other hand, there are beliefs and principles that are almost entirely of a personal nature (*there is insanity in my family and I will be overtaken by it some day; I have been wronged in life, and I will spare none; all my tormentors will be punished one day*). We speak of beliefs and principles in the same breath since beliefs are generally of the nature of rules and principles we go by in our life, the more so when a belief is hooked onto our self-network. The more personal a belief is, the more laden it is with a value-judgment, the more intense are the emotions attached to it, and the more insidiously does it influence our judgments and actions.

As we engage in an action or thought, some or many of the resources lodged in our mind get invoked by means of the ongoing interactions among the neural assemblies in the brain, along with the release of a number of neurotransmitters and their attachment to specific sites that activates the neural value-system, this last-named playing the role of a guiding spirit in our actions and thoughts by way of 'approving' or 'frowning upon' the bits of activity through which the mind proceeds in its business.

In an act of making an inference, the resources principally made use of by the inferring mind are the knowledge base and the beliefs — the more widely accepted or shared ones among the latter. The pieces of knowledge and the beliefs invoked are generally in the form of principles or rules ('if this then that', or 'since this hence that') by means of which the inference proceeds. At times, the inferential process reaches some specific conclusion (*'from all evidence, arguments, and counter-arguments, the accused is proven guilty beyond doubt'*). On the other hand, it is more generally the case that the inference remains inconclusive, with several alternative conclusions making their appearance in the realm of possibility (*my friend promised to come to-day — he must have fallen ill, or else must have had to leave station on urgent business*). As explained in [82] and in earlier sections of this book (see, in particular, section 3.2.2), one then has to *guess*, and to guess *in the right direction.* It is then that the inferring mind makes use of resources of dubious credentials — half-remembered snippets of knowledge, unfounded information, half-baked wisdom, beliefs of a personal nature, heuristics and hunches, vague clues, memory laden with emotion — in short, resources linked in some measure to the self of the inferring individual — resources associated with emotional valence.

This seems to be a *plausible description* of how inferential processes take place in the human mind. The actual neural processes, however, are completely out of the scope of this description since there is no clue as to how the various psychological resources and their dynamical interactions are *represented* at the neural level.

What is to be noted in the context of this description is that the possibility of 'alternatives' and a choice between those can be admitted with reference to the cognitive tasks of inference-making where there is considerable role of shared world-view, shared knowledge, and shared principles of inference.

On the other hand, it is equally likely that there occur mental process more heavily dependent on personal psychological resources where the possibility of alternatives appearing in virtue of inconclusive use of shared resources does not arise. While there may occur processes that are part cognitive and part self-psychological (I use the term 'self-psychological' here to refer to psychological processes in which, by definition, self-linked resources are made use of — the nomenclature is for the sake of adding emphasis), the other kind of a *predominantly self-psychological* nature appears to be relevant to a wide spectrum of actions and thoughts in which we engage, many of which appear as exercise of free will. The question of alternatives does not arise in these processes in the same way as in the other cognitive processes mentioned above. These are driven to a large extent by personal beliefs, emotions, and affects, and are not directed towards a 'conclusion' in the same way as the other cognitive processes are. Rather, starting from a certain external and internal context, these self-psychological processes proceed along some entangled causal chain depending on the workings of the self-network and end up in a 'result' that appears as an exercise of free will.

Two neighbors (call them A and B) are engaged in a bitter dispute over the possession of a small plot of land having a common boundary with the estate of each of them. In a fit of blinding rage A strikes a blow to B with a metal rod lying close at hand, and B is immediately rushed to the hospital. In this not-so-rare incident, A is said to have 'chosen' the 'option' of physical violence over several other 'options' of less violent resolution of the dispute. But a group of impartial men of the locality, when asked to give their opinion on the matter, inform an onlooker that A is not of a disposition to even think of these other options — he is ever bent upon using physical violence as the principal means of persuading an opponent. The interested onlooker, on further inquiry, learns that A has had an unfortunate childhood, and is currently having problems with his wife — that he does not have much of an educational or cultural background, routinely beats his children, and is devoted to drinking. B, on the other hand, is of a milder disposition, and 'in consequence', has suffered at the hands of A.

In what sense can the actions of A and B in the dispute be said to constitute instances of free will, when it appears that the outcome of the dispute was almost preordained by the very nature and disposition of the two disputants? Herein lies the central issue. The two have acted according to their own free will, *not* because of whether or not their actions are determined by the dispositions of the two and by the circumstances of the case. They have exercised their respective free wills precisely because two *other* persons (call them C and D) in the same situation would possibly have acted otherwise.

And it is here that a very big question keeps on bothering us. We feel in our guts that *A could really have done otherwise*. But, *in what sense*? That is what we will now turn to. Before that, we briefly look at the same question, but now in the context of a cognitive act of inductive inference.

Two friends (call them A and B once again) working in different laboratories are engaged in a friendly rivalry in their chosen task of determining the structure of a complex biological molecule. The race is exciting to both and each of them works hard in his own way. But it is B who succeeds by making good use of a novel hypothesis that later proves useful, even before A has barely got to the depth of the problem. A teacher of theirs from their earlier college days when both used to attend her class recalls an incident when both did ordinarily in a class test but B came back doing exceedingly well in the next test while A's performance was again quite ordinary. What this small incident tells us is that B engages in *self-examination* and *self-improvement*, and is a good *learner* from failures, but A is not.

Self-examination, self-improvement, and learning from one's own failures, herein lies the ultimate source of 'alternatives' in acts of cognition as also in the exercise of free will. As we have promised above, this is what we will now turn to.

5.1.7 Reasoned exercise of free will: the necessary and the contingent

One frequently associates reason and rationality with free will. This appears to be in conflict with what we have been saying in this chapter where we seem to have equated the exercise of free will with action and thought guided by emotion-laden psychological factors associated with the self of an individual. We accept that there is ground for such an impression from what has been written above. This section will, perhaps, set things in the right perspective while, at the same time, underlining the essential role of self-linked resources in the exercise of free will.

We have distinguished between shared mental resources such as the knowledge base, inter-subjective beliefs, and commonly accepted rules and principles on the one hand, and the self-linked ones involving emotional components on the other (where the unique developmental history of an individual assumes relevance), but such distinction is only notional. Everything in the mind is associated with everything else, which is why the mind is such a complex system, almost as complex and infinite as the universe out there. Mental processes cannot be classified in a dichotomous either-or manner, and any classification can only be in the sense of a spectrum, where one part of the spectrum is perceived to be distinct from another while there exists a complex interlocking between the two through a continuous transition from one to the other.

Imagine the judge preparing the verdict in a complicated case that was stuck in the court for long years. The judge is constrained by voluminous documents covering all legal aspects of the case including meticulously laid down clauses of the juridical code that may prove to be relevant to it. The judge exercises her reason and rationality to the utmost in interpreting all the documents now lying before her but still at the end of the day it will have to be *her interpretation* that will go into the judgment. There are *two* things here finely blended into one.

In arriving at an interpretation consistent with all the documents that have piled up, the judge makes use of her reason and rationality, calling into play a stupendous amount of legal knowledge, experience, and wisdom. However, when we look closely at the terms that we have just listed, one finds that there are certain fringe areas that raise the possibility of the interpretation of one judge differing subtly from that of another if she (the second judge, that is) were to go through the same exercise. Evidently, there is *some* scope of the self-linked factors of the judge being called into play. Which is why the judgment can go for review, though that will still not remove all possibility of an alternative interpretation — there frequently, if not always, remains the lingering doubt that a person accused of murder and sentenced to capital punishment did not actually commit the murder. But all these constitute *one* of the two aspects we set out to elaborate. We will now leave all these possible self-related issues of interpretation aside, and look at the one thing that cannot *possibly* depend on self-related factors of the judge — namely, the legal *knowledge* that she bases her verdict upon. Who can deny that knowledge, even legal knowledge, is, to all intents and purposes, universal? But here again, there remains a question mark that we have already hinted at. While individual items of knowledge are indeed of universal validity, the total *mass* of knowledge relevant to the case under consideration is specific to the judge, along with the way she makes use of it. One individual differs in thousand and one ways from others in her diligence, perseverance, commitment and qualities of perception that develop throughout her life and determine the mass of knowledge that will be in her command and the specific use she will be able to make of it. This indeed sets a judge apart from her colleagues and provides the basis of that extra respect that she commands in legal circles, apart from specifically self-related issues like her acumen, perspicacity, integrity, and independence.

We will not labor the point any more. The basic issue relating to free will that we perceive, reduces to whether and how we distinguish between the *necessary* (i.e., factors answering to reason and rationality) and the *contingent* (i.e., self-linked factors). Two leaves in a tree have the same general physical and biological structure characterizing them, but still they differ in a thousand and one small ways that may be described as contingent — ones that may be said to define their 'self'. Two individuals from the same family have many things in common, not only in virtue of both being human beings but additionally in virtue of being members of the same family — still they differ in their fingerprints that depend on *contingent* factors in early embryogenic history in spite of being guided by shared genetic characteristics.

While the fingerprints of an individual are contingent features determined in her early developmental history, innumerable other contingent factors accumulate during the entire course of her development, many of which disappear without leaving any trace while the remaining ones leave their imprint upon her, especially upon her mind. Broadly speaking, *all* these factors taken together constitute her *self*, though in the context of the present subject, we have earlier mentioned only a number of those that we have found to be of relatively greater relevance in characterizing free will as we see it.

Every act of exercise of free will is possessed of general or necessary features that can be *explained* in correspondingly general terms by invoking concepts such as reason, rationality, and knowledge base of the person concerned. But the same general terms of reference fail to explain *specific* aspects of the act since these specific aspects may be found to differ for different individuals under similar circumstances, because these depend on *contingent* factors relating to those specific individuals. It is *this* that is often interpreted as free will being 'free' of the constraints of determinism.

Leaving aside the ontological aspect of the tenet of determinism, I repeat that the epistemological aspect of the tenet is often taken to be what determinism means: what we commonly mean by the term determinism is that, given sufficient knowledge base, one can explain and predict everything there is to explain and predict. But, 'knowledge base' includes not only the mechanisms underlying all the relevant processes, but all the *contingent* factors that determine the *initial and boundary conditions* as well. Let us grant that it is possible to know all the relevant mechanisms (the 'laws', that is; as we have already stated, this is a *very big* supposition indeed), but then what remains is the hugely relevant, but *contingent* set of *initial and boundary conditions* on which the mechanisms operate. Unless ones takes into account all the contingent factors involved, predictability vanishes into thin air. This, precisely, is the case with free will where the contingent factors reside in the self of an individual.

The boundary conditions in a problem refer to its context. As explained earlier, problems usually pertain to *models* where only a part of the infinitely extended and complex reality is taken into account. The boundary conditions then denote the effects of those remote parts of reality that do not explicitly find a place in the model, but represent these in some idealized sense.

It is said that the laws of physics operate inexorably towards a future (thankfully, a distant one) of 'heat death', that is, an equilibrium configuration on a grand scale. But that is not what is relevant now and here. All our thoughts and activities are based on *non-equilibrium* configurations and processes where the detailed *structures* in space and time matter, and matter enormously because these include our very own life processes, our thoughts, aspirations, and hopes. And, *all* these configurations and processes depend on contingent factors along with the necessary ones — the latter pertaining to the 'laws'. The same apply, in particular, to the exercise of free will. An act of free will can be 'explained' only partially on the basis of what may be looked upon as 'necessary' aspects — the ones relating to reason and rationality. Recall that, in this sense, reason and rationality act as *constraints* rather than factors relating to freedom. What is really relevant to freedom are the *contingent* factors — all of those that relate to the self: the entire set of innumerable details relating to the developmental history of the individual — all those, that is, that leave an imprint upon her mind. What is relevant and interesting is that, our efforts at 'explaining' an act of free will flounder when confronted with these contingent factors in virtue of these being hidden to a large extent even from ourselves. The paradox of free will resides in this fact of all our actions and thoughts being stamped at once with necessity and contingency.

And, as we have stated repeatedly, this is precisely why free will appears to be free — appears to defy the iron grip of determinism. On the other hand, the 'freedom' is, in a manner of speaking, only an apparent one since free will is not free from the operation of correlations of a causal nature: the foundational position that we adopt is that all processes occur on the basis of causal correlations that can be determined by painstaking inquiry and investigation, at least partly and contextually, making possible successful predictions within limited horizons. In other words, freedom is not free from the dictates of causality but is free in the sense of freedom from the demands of *necessity*, being determined by *contingent* factors. The latter are responsible for the loss of *predictability* of behavior.

This brings up the question as to how determinism is related to causality and predictability. What is often confused in discourses relating to determinism is the complex relation between causality and predictability, because determinism is, to a large extent, co-extensive with causality. Is the state of a system in the immediate future dependent on its states in the immediate or the more distant past? This is the basic question that determinism and causality answer in the affirmative. But then comes the question, how far can this fact be relied upon in *predicting* the states of the system in the distant future? A naive answer to this question would again be an optimistic one. But then, this optimism holds for an idealized system where effects of pervasive complexity can be ignored. When we look at a chunk scooped out of a complex system, its behavior in the distant future depends *sensitively* on the initial and boundary conditions — the latter representing the context in which the system is probed and studied.

So, at the end of the day, is the question *finally* settled? Is free will *really* free? Is man genuinely free to exercise his will?

The answer to this question has to depend on what we mean to be free. The way we look at it in this book, the only real freedom lies in grabbing at chances to learn, to look at ourselves, to confront ourselves, and to enrich ourselves. Such chances do turn up. Whether or not we make use of those, depend on how we have learned to focus our awareness onto ourselves, to communicate with our very own selves, and to let reason work hand in hand with our emotions and value judgments. Setting up this communication is the ultimate thing in living an *open* life, in letting light shine on the twilight zone where our *soul* operates.

5.1.8 Free will: concluding words

First, a quick summing up (it has been a long monologue), and *then* a few generalizations and a few additional comments made in conclusion.

5.1.8.1 Induction and free will: summary

The following summary of the present section will also contain a number of additional remarks.

1. In this section, aimed primarily at assembling a point of view with re-

gard to free will, we have drawn a parallel with inductive inference since both inductive inference and free will are based on the use of *self-linked psychological resources* of an individual, namely, heuristics and personal beliefs (as distinct from shared ones in common with other people), affects, emotions, and feelings along with the all-important psychological value-system. These are associated with a number of other ingredients such as fantasies, desires and drives that we have not referred to explicitly.

2. We adopt the position that all these psychological resources and the processes they participate in are based on electrochemical events in the brain, involving neurons and their interactions, associated with numerous other physiological processes. We adopt the position that psychological processes are phenomenal appearances of underlying neural ones where the principal functional units are large neural assemblies rather than individual neurons, and where these large assemblies interact with one another by means of multitudes of neuronal pathways across which wavelike synchronous oscillations spread over various frequency bands. It is these complex waves that carry the 'information' produced by one functional unit to another, thereby making possible a global orchestration (without, however, any centralized control) that preserves the integrity of our mental world in spite of the enormous diversity of mental processes and conflicts residing therein.

3. Within the complex organization of the large scale neuronal networks, there is the *self-network* that is a complex system itself, made up of a number of large neuronal assemblies communicating with one another and with the rest of the brain systems — a network supporting dynamical excitation patterns where the two evolve throughout life. The functioning of the self-network involves the self-linked psychological resources mentioned above. According to the position adopted in this book, it is

the feature of *complexity* of the self-network and the interacting neural networks of the brain that is of central relevance relating to events and processes in our mental life.

4. While inductive inference (discussed more fully in [82], see also section 3.2.1) involves the *cognitive* face of the self, working in tandem with a commonly shared world-view made up of commonly accepted principles, rules, and logical presumptions, free will is linked to a greater extent with emotion-based processes that we refer to as 'self-psychological' ones. The dividing line between inductive inference and free will is not sharp, and there occur processes possessed of features of both.

5. It is said that both inductive inference and free will implies a 'choice' from among 'alternatives'. This chapter tries to explain whether and in exactly what sense such choice can be said to be involved. In the case of free will, the very existence of the 'choice' is supposed to pose a problem with the tenet of determinism. In this chapter, we have not engaged with the issue of determinism which we feel is too big to be of concrete relevance in our present context. On the other hand, we adopt the position that psychological processes have, associated with them, underlying neural ones that have causal links running between them — ones that can, to a certain extent, be identified and made the basis of predictions (in terms of the psychological processes themselves) within a limited horizon. As mentioned above, the human mind is a complex system, with the self-network functioning as a complex system itself. Complexity, indeed, *exists in all scales*, and *all* psychological processes, including the making of inductive inference and the exercise of free will, both based on interactions among complex systems. In particular, the external world is an enormously complex system itself, and psychological processes are, in one sense, the continuously occurring *response* of the mind (in association with the body) to the particular facet of the world that happens to be

presented to it at this point of time or that.

6. As the mind engages in the task of responding to a conjunction of internal and external contexts, it makes use of its shared world-view made up of its knowledge base, and of commonly accepted beliefs, rules, principles, and logical presumptions. It may so happen that an adequate response is formed solely on the basis of this shared world-view. More commonly, however, the shared world-view leads to no unique position, and the question of a choice from among alternatives arises. This, for instance, is the case with inductive inference which is a cognitive process involving the shared world-view of an individual as also her self-linked psychological resources. The latter are made use of in arriving at an inference which is not guaranteed to be correct. On the other hand, there occur processes where the use of the shared world-view is secondary to a relatively more predominant role of self-linked psychological resources. In such cases the question of choice from among alternatives does not arise in the way it is commonly understood to.

7. When we speak of the choice involved in an act of exercise of free will what we actually refer to is the commonly observed fact that various different individuals of a similar disposition respond differently when placed under similar circumstances. In other words, free will is 'free' precisely because it is not subject to constraints of a commonly accepted and shared set of principles, beliefs, and values. While it is possible that, in an exercise of free will, the self-linked psychological resources are brought into action when a number of alternatives are presented to the self by the operation of the ingredients of a commonly shared world-view, it seems likely that *that* set of alternatives is not of primary relevance in the final response that the mind produces, since the latter is primarily a product of the operation of self-linked psychological resources *regardless* of the commonly shared world-view. What is of greater relevance here is the operation

of emotion-driven processes, guided by the psychological value-system based on the activity of the so-called reward-punishment network (or, in an alternative description, of the *seeking* system, see sec. 2.4) in the brain. These processes lead the individual to a response that *appears* to be free from the shackles of determinism precisely because their mechanisms, which are hidden from us, do not conform to commonly accepted and shared rules and principles, while causal mechanisms keep on occurring all along within the mind of the individual concerned.

8. Thus, there is really no choice involved in any significant way in the exercise of free will when one looks at it in the context of the mental processes of a single individual, without comparing with how *other* individuals would respond under similar circumstances. Indeed, whatever happens in the mind of an individual, happens in accordance with the principle of causality. *On the other hand*, it still makes sense to ask whether or not *that same individual could do otherwise* than what she actually does, i.e., the way she actually makes a response to some specified situation. While this seems paradoxical, things clear up once one gets to comprehend the sense it is intended to carry.

9. This brings us to the issue of *self-examination* and the associated process of *self-improvement*, where the latter involves an enrichment of self-linked psychological resources in the form of revision of one or more of the beliefs of a deeply personal nature, and of an alteration, at least to some extent, of the ebb and flow of emotions in one's mental processes — of an alteration in the way the psychological value-system operates. This is brought about by way of an examination of the roots of one's own preferences, aversions, and negative judgments about the world. As one engages in self-examination at any given point of time, it results in the possibility of an alternative course of response to internal and external contexts arising at *later* times. In other words, as one engages in self-

examination *now*, one opens up the possibility of *acting otherwise later*, where the comparison is now with reference to the response that would result in case the self-examination were not to take place.

10. What determines whether an individual will engage in self-examination under given circumstances? Does it depend only on the disposition of the psychological resources of the individual at a given point of time? If so, the same old problem of determinism is ready to raise its head. However, self-examination constitutes a very special instance of *learning*. The role of learning is of great significance in the making of inferences, including ones of inductive inference. Learning expands and enriches the commonly shared pool of knowledge, beliefs, and inter-subjectively accepted logical principles. In case these ingredients fail to lead one to some specific response to a cognitive challenge, one is faced with the job of choosing between alternatives by bringing in one's self-linked psychological resources. But how and when does one learn? One does so under the impact of the world without, including instructions, suggestions, and various other types of messages from other members of the society. We learn from books, we learn from circumstances, we learn from exemplary behavior of people we admire, and we learn in thousand and one other ways.

    Self-examination, as the name implies, involves an appraisal of one's own self, and is a very special instance of learning under the impact of circumstances. Here the self, in some particular mode and disposition of the self-ingredients that happens to have materialized at some point of time in the course of a ceaseless dynamic process of evolution, comes in conjunction with some particular facet of the world that happens to be presented to it. It is this *conjunction* that leads to self-examination and self-enrichment which, therefore, is *probabilistic* in nature, though based on definite causal links that can be discerned within limits, where the limits arise because of the fact that one is now looking for causal links within *complex* systems. For instance, as an individual gets in touch with another person several times in succession, there may be just one occasion when the suggestions of that second person leads to an attempt of self-examination by the individual in question, because of some particular set of contingent circumstantial factors having come into play.

Some ten years back, my dad had called me to his bedside and had tried to impress upon me that I was then under the sway of a set of wrong perceptions about the world, which could inflict damage on me. I did not pay heed and had left his bedside with a sour face. Now, ten years hence, when he is no more, I suddenly recall his sadness at my refusal to confront myself and also recall how he had always been so very keen to help me in small ways and to ease the turmoil within me. A wave sweeps through my mind, and the same me who had been headstrong ten years back now find myself trying desperately to scratch the hard and roughened surface of my own mind so that I can communicate with my troubled soul.

Self-examination is akin to learning — but is learning of a very special kind. It is not borrowing a book and studying, nor looking at a protein molecule under the microscope. Here one looks at and studies *oneself*, questions one's own closely held beliefs, tries to divert the tide of one's own emotions, and *risks getting into a turmoil*. Is it logically possible for the mind to engage in self-reference so as to be able to do all this? As we see below, there is an element of chance involved in this. Circumstantial factors offer us a chance to confront ourselves — at times we grab that chance and at times we don't. Here the term 'circumstantial factors' refers to some conjunction of the state of the mind and the state of the world out there.

11. Continuing to summarize, it is self-introspection and self-enrichment that makes us ready to make adequate use of our reason and rationality. The exercise of free will covers a wide spectrum: from acts guided primarily by reason and rationality at one end, to those guided primarily by our self-linked psychological resources at the other. There is *no* act or

thought that is completely free of the self-resources. What we mean by determinism and what we mean by freedom are, in fact, the two faces of reality — the *necessary* and the *contingent*. Reason and rationality represent and express the 'laws' of this world — at least our *interpretation* of what we mean by laws — while what we mean by freedom resides in all the contingent factors of that same world. And, the *self* is contingent — contingent upon all the innumerable coincidences and conjunctions that occur during the lifetime of an individual and leave their imprint on her, especially on her mind. It behooves us to see and to comprehend that necessity and contingency are not two conflicting faces of nature. Nature is one single hugely complex entity, with myriads of parts correlated with one another in ways that we can only guess at. It is our way of viewing nature that makes us interpret it in terms of necessity and contingency. Necessity appears to us as determinism while contingency appears as freedom. This is how it is with our own acts and thoughts too, where necessity and contingency appear to be in conflict for ever.

12. Earlier sections of this book have dwelt at some length on how the self-linked psychological resources come to play a major role in our mental life. Some of these have been devoted to examining a number of features of complex systems including the ones of amplification leading to local instability and the counterposing influence of factors ensuring global stability. We have, in particular, underlined the role of emotions, aided by the psychological value-network, in leading to such instability and counter-stability, responsible for innumerable different *modes* of operation of the self-network, where such different modes of operation may be looked upon as *emergent* behavior in complex systems.

    In looking at complex systems, we have outlined how determinism and predictability are to be interpreted in a manner somewhat different from the commonly adopted view of these. In particular, we have indicated that causal factors never cease to operate in the workings of a complex system interacting with complex systems of other descriptions. However, these causal mechanisms imply a kind of predictability that is strictly limited in scope, the validity of which is lost beyond a time horizon depending on the nature of the interacting systems under consideration. In a sense, such limitations on our predictability of events and phenomena is ubiquitous. The inherent limits of predictability in respect of complex systems are reached in scientific investigations that have a tendency to push against these limits, break these open, and discover new causal links, only to be confronted with re-established limits of predictability making their appearance further away.

13. Let us allow ourselves to talk in symbolic terms here that may appear to be a bit abstract — ones that may nevertheless help to make things more comprehensible.

    States of our mind turn out to be constantly and incessantly in conjunction with states of the world at large — a conjunction of states of two complex systems. All our behavior, all our response to the world, are results arising from this conjunction, which determines how the two systems interact at any given point of time. However, because of their complex nature, states of either system are densely close to one another. For instance, a state, say, A, of the mind has a large number of other states (A', A'' $, \cdots$) close to it; we refer to these states collectively as, say, $\{A\}$. In a similar vein, a state B of the world at large (recall that we indulge for now in abstract considerations, only to make a point) has states B', B'' $, \cdots$ close to it; making up a set $\{B\}$. Because of the complex nature of the two systems, as a result of which their time evolution is possessed of complex features (refer back to section 4.2.1) it is essentially unpredictable as to which state belonging to $\{A\}$ happens to be in conjunction with which state of $\{B\}$ at any given instant, given our past readings of the behavior patterns of the two systems. For most situations this indeterminateness does not matter much in how our journey in this world unfolds, at least within limited time horizons. In some circumstances, however, the specific conjunction turns out to be of momentous significance in some time interval of practical relevance — which is how the phenomenon of *sensitivity to initial conditions* expresses itself.

Self-examination leading to *self-improvement* is precisely of such significant consequence.

Looking at free will as the operation of self-linked psychological resources operating causally, one has to admit that there is no escape here from causal and deterministic processes, though, as we have seen above, there is a sense in which free will is 'free', namely, in being free of causal mechanisms when explanations based on shared rules and principles are sought for and comparisons are made between individuals. However, the question mooted above relates to 'freedom' of a different kind — partial freedom from the dictates of the affect-driven self. It is *this* freedom that attempts at self-examination and self-improvement offer us.

14. From a broad point of view, the interaction between two vastly complex systems — the mind on the one hand and the external reality on the other (we ignore for the time being the fact that parts of the mind itself is incessantly mapped by other parts) lies at the root of of *all* our thoughts and behavior, which is thus stamped with a measure of inherent randomness and unpredictability. In this book we have not had much occasion to refer to this foundational fact of our mental life. It is this unpredictability that characterizes the way we face our moments of self-examination or, on the other hand, shy away from it.

5.1.8.2 From individuals to societies... and on to mankind

Each of the large scale neuronal assemblies in the brain has its own built-in functionality involving the processing of incoming information by the mass of the interconnected neurons it is made up of. The results of this processing are then exchanged between the various assemblies by means of wavelike interactions of a complex nature among the various assemblies.

As repeatedly stressed in this book (see, in particular, section 2.7 and also sec 5.2.8 below), this has a remarkable resemblance with how larger groups of men and women and entire human societies operate. A society of men and women is usually organized into units or groups that communicate and interact among themselves by various means, constituting functional units of diverse descriptions. Thus, there are children belonging to a class, teachers in a school, workers in a factory, the managerial and administrative units in a business organization, the local council in a city, the federal government in a country, a political party, a religious denomination, and so on, each operating in its own unique way. The individual persons belonging to all these groups and collections have similar constitutions in respect of their bodily structure, anatomy, physiology, and mental capacities while, at the same time, they participate in the common functionality of the respective groups.

> We focus here on the common features of the individuals of a group while, in reality, each individual has his or her uniquely distinct personal identity too.

Transcending the level of individual existence and functioning, and also the level of functional groupings, the society as a whole is marked by its own existential features and characteristics. In other words, looking at the individuals as the 'neurons', and the various groups and organizations as the 'neuronal assemblies', one can, in a manner of speaking, look at a large social group or a society as a whole as the collection of these assemblies where, at each level, one has a complex system belonging to another complex system at a 'higher level'. Finally, the society as a whole interacts with other societies of human beings.

1. Relegating individuals to the status of 'neurons' in a society looked upon as a 'brain' raises the specter of science fiction horrors. But that, obviously, is farthest from the spirit of all that we venture to write here.

2. When one places the idea of complexity to the fore, the classical mode of description in terms of computational models of individuals and societies is avoided, though the classical computational models are important in their own right.

Analogous to the beliefs of an individual, a social group possesses its own cultural features, its own preferences, and its own dispositions. A social group also possesses 'emotions' — those internal modes of communication that amplify a response in the face of an 'input' from external sources. For instance, an event having a major political impact raises the passions of various sections of the citizens of a society. The preferences and 'emotions' generate the 'self' of the group.

We will pursue this line of discourse in later sections of the book while, for now, we will come to what we propose to say in this concluding part of this section on free will.

Just as in the case of an individual, a group of individuals can also be said to engage in acts of 'free will' where such an act distinguishes it from other similar groups under similar circumstances. We do not usually describe such activity as instances of free will, but every group does have its own culture, its own specific response to situations, and its very own ethos.

Several groups of young college students are sitting with snacks and coffee in a cafeteria, members of each group exchanging pleasantries among themselves. An old man in tattered rags is found to collapse in front of the shop, gasping for breath. Most of the groups carry on their loud laughter and conversation, paying little attention to the incident. But one group of young persons immediately rush to the help of the hapless old man and get him first aid, and then further medical attention. Clearly, this particular group among all those sitting in the coffee shop has a culture and identity of its own. Given a situation, various different groups of men respond differently, all those different responses being alternative ways from which each particular group may, on the face of it, be said to 'choose' some particular alternative or other. But such 'choice' is no more than a convenient way of describing the specific response of one group as distinct from those of other similarly constituted groups. For instance, the group of students rushing forward to the help of the sick old man did not spend time in weighing between the relative merits of continuing with their pleasant gossiping and rising to provide urgently needed support — they 'spontaneously' rushed out in response to some tacitly felt common impulse without pausing to 'choose'.

As with a group of men, so with a larger society of men or an entire nation. Looking at the ways of governments and nation states, one finds infinitely many differences in the way they set up their relations with one another. Is it possible to speak of the cultural ethos of a nation while recognizing that the innumerable individuals and groups of men and women that it is composed of think and act differently? As I understand it, one can indeed speak of the specific characteristics of a nation, though in a qualified sense, just as one can speak of the character trait of a single individual in spite of the innumerable pulls and pushes operating within his mind that cause him to have various distinct and conflicting tendencies, even though his outward behavior pattern at a given point of time is a resultant of sorts of all this multitude of disparate propensities. The idea of a nation or a large society of men engaging in self-introspection is not inconceivable though, understandably, such self-introspection is rare indeed, being possible only under exceptional circumstances and by exceptional means. It can happen, for instance, if one person or an influential group of persons of great stature, or even a government in power or, say, an opposition group, rises above others in a time of crisis and places before the people a set of novel ideas with some stirring message, around which the nation rallies — not that such passion always produces the best of results.

And finally,..... on to *mankind*. Is it possible for mankind to self-examine and to turn over a new leaf? Is it conceivable that we, men and women of this world, are capable of looking collectively and meaningfully at the destruction and corrosion that is eating away at our soul? Is there yet a 'free will' left in mankind?

5.2 The soul: reconstructing the self

We begin by recalling and clarifying a number of foundational concepts outlined in earlier sections of this book, some of which have been repeated many times over.

The soul is interpreted in many different ways by many people. In this book we have defined the soul as the capacity of the conscious mind to examine the self-system and to keep on reconstructing it (section 1.14). Moments and opportunities present themselves when the conscious mind can be directed inward and the self can be examined at least to some extent, but we let those moments slip by and recede from us, only to be sadly reminded later of lost opportunities when life seems vacuous and unbearably burdensome — when we turn away from life and let it crumble us to pieces. Oftentimes, this inward disintegration doesn't 'show up' — life, apparently, continues; we get on with our denials. But the rot spreads. It is not even possible to 'prove' unambiguously that the rot is connected underneath with our inner debacle. One doesn't know if that is the way the world is destined to 'progress'.

5.2.1 Goal-directedness in nature: driven systems

It all begins with the mind working incessantly with a purpose — the purpose of looking for and appropriating happiness, contentment, and satisfaction. But there begins a story riddled with conflicts and frustrations, ending in despair.

Mental processes occur in virtue of an innumerable variety of neuronal excitation patterns generated in interactions between neuronal assemblies in the brain. These processes lead to further processes of excitation on the one hand and neuro-motor interactions on the other, the latter generating an infinite diversity of behavior of men and women engaged in social interactions.

A foundational organizing principle underlying all these neuronal and social processes is *satisfaction of needs and preferences*, which makes them purposive and goal-oriented. Such fundamentally purposive processes of the mind of an individual is analogous to the apparent goal-directedness of biological evolution itself where populations of entire species evolve by means of adaptive interactions with their environment whereby some survive and proliferate while some others become extinct.

The goal-directed evolution of biological populations is based on genetic transmission of traits, while pre-biotic evolution can also be said to be a 'goal-directed' process, that 'goal' being the generation of self-replicating molecules in an appropriate environment. The takeaway from all these background considerations is that goal-directedness appears as an emergent phenomenon in the time-evolution of complex systems that involves self-organized complexity, and does not necessarily reflect any grand design in Nature.

A huge colllection of randomly moving electrons, protons and neutrons within a finite volume is likely to end up in the production an ensemble of atoms and molecules. Even this can be described as a 'goal-directed' process. In other words, goal-directedness or purposive time-evolution appears to occur on a 'macroscopic' scale from underlying 'microscopic' dynamics of large systems of entities, and depends on our perception and interpretation of the associated processes. An alternative description of 'macroscopic' and 'microscopic' assemblies of entities is in terms of 'levels' in a hierarchy of complexity.

An enormously large assembly of entities interacting with one another passes through phases of self-organized complexity in the course of approaching thermodynamic equilibrium. In the process, various structures representing local equilibria appear within it, where these local equilibria are then destroyed on a longer time scale by means of fluctuations within the system. Even a global equilibrium is destroyed on an enormously long time scale by processes of fluctuation. In other words, goal-directedness and purposive evolution are features that appear at various spatial and temporal scales in the time evolution of complex systems.

Refer back to sec. 4.2 for a brief and sketchy background.

Goal-directedness is a feature that is routinely generated in *driven* systems where processes occurring in a fixed direction are maintained by a supply of energy and matter from outside. For instance, a chemical reactor produces a constant flux of output by means of supply of heat and input material that maintain chemical reactions within it at a fixed rate and, moreover, stabilize the steady state against perturbations.

The neurons in the brain work on an analogous principle. A neuron would be inert after generating an action spike unless a myriad of non-equilibrium processes were maintained in it by supply of energy and chemicals from its environment. Such supply maintains the incessant generation of excitation patterns in the neuron and also in a neuronal assembly. In other words, the neuron is also a goal-directed system in virtue of appropriate supply of energy and chemicals to it — its 'purpose' is to keep on generating action spikes. Of course, for the neuron or for the brain to operate in a goal-directed manner, the biological system within which it is embedded and from which it receives its supply of energy and matter has to be goal-directed too (such goal-directed flux corresponds to a flow of 'information', when the term 'information' is defined in an appropriate manner), though the two 'goals' (those of the driving and the driven systems) need not be the same.

Driven systems have the interesting property that they not only maintain processes occurring in a fixed direction (the 'goal'), but also maintain the stability of those processes — a small deviation from the 'goal' is compensated for.

> This is a result in non-equilibrium thermodynamics (stability of steady states close to an equilibrium) and is essentially an extension of the second law of thermodynamics.

This is analogous to what happens in the case of a living organism as well, though a living organism and its subsystems are maintained in a dynamically stable non-equilibrium state maintained *far* from equilibrium. Here the results on close-to-equilibrium steady states do not apply in a rigorous sense but all the life processes are organized in such a manner that the system as a whole and all its parts operate in a similarly goal-directed manner (i.e., in a fixed direction) and deviations from the 'goal' are compensated for so that, on the whole, there is ensured a stability of the states of the living system and its parts.

This, then, is the takeaway: *processes in a living system occur in a fixed direction (i.e., with energy, matter, and 'information' flowing in a fixed manner) and are stable, with deviations compensated for by means of flows generated in opposite directions.*

Similar features are observed in many instances in the non-living systems as well (air currents, oceanic flows, atmospheric systems) where the specific details may differ but the above feature of stability of steady states holds.

There are two important qualifications, though: (a) while we have used the term 'steady state', in reality, the state may be (and often is) a non-steady one (for instance, a periodically varying state) — but on the whole, the course of the process is maintained against small fluctuations; and (b) there do occur secular changes, by means of which the system eventually moves away from the steady state on a longer time scale.

All these observations do not have solid and rigorous theoretical foundations but have been observed in simulations of driven *complex* systems and their subsystems, where the 'driving' is done by systems in the 'environments' of the ones under consideration (recall the features of self-organized complexity and self-organized criticality, outlined in sections 4.2.1, 4.2.6).

In this book, we take this as the fundamental point of departure of all our considerations: living organisms share with a host of other natural systems the property of goal-directedness and stability under driving by other systems with which they interact, and deviate from the 'goal' on longer time scales, when a new set of processes appear, with new features that may or may not answer to the description of goal-directed ones.

What distinguishes a living organism from other systems with similar features is the specific set of interactions with the driving systems, the specific physical parameters characterizing the states of the organism, and its specific modes of behavior — the *general* features of goal-directedness and stability are shared with many other systems in nature.

> The stable time-dependent states of living organisms were brought under systematic discourse by C.H. Waddington, a visionary scientist who pioneered numerous fields in biology, including those of developmental biology and epigenetics. Waddington referred to the phenomenon of maintenance of a stable time-dependent process in a living organism as *homeorhesis*, which constitutes a generalization of *homeostasis* — the maintenance of steady states — in various parts of an organism.

The human body is a driven complex system maintained in a homeorhetic state that we call life. It is made up of parts (organs), all of which are complex systems themselves, all in respective homeorhetic conditions. Among these the brain, made up of neurons aided with neurotransmitter systems, is the seat of the mind. The brain is a major driving system (and a driven system too) responsible for keeping the body in a homeorhetic condition by making the body respond in appropriate ways to its environment. The *mind* is a complex set of processes in the brain, making its presence felt as a *phenomenal* perception, that can be described as an emergent phenomenon based on the incessant and complex dynamics of the neuronal system of the brain.

The mind, in turn, is a driven and goal-directed complex system. Its 'goal' is perceived phenomenally as a most curious one — the *satisfaction of a huge set of preferences* that keeps on evolving during one's lifetime. For this, the mind is served by a remarkable device — the *affect system*, based on processes in a set of neuronal assemblies in the brain, that generates affective valuation, aiding in the generation of emotions. All our experiences are either directly or indirectly stamped with affective valuations and emotions in complex combinations. This, precisely, is how the self gets assembled in the mind and evolves in a life-long developmental process.

5.2.2 The affect system: how are preferences generated

The human mind operates on the basis of a vast number of preferences, where these preferences depend on the developmental history of an individual. How are these preferences generated in the mind? Do these, in any way, reflect the goals set by the process of biological evolution?

> The term 'preferences' is used in this book in an inclusive sense so as to stand for 'preferences and aversions' of the mind — aversions, blended with negative emotions, are produced in virtue of negative valence generated by the affect system.

It is doubtful if an universally acceptable answer can be given to this last question. However, one can with greater assurance indicate the *beginnings* of the process of generation of preferences in an individual. Stated simply, the affect system in the newborn operates to guarantee its well-being as a *biological* organism, as it does in the case of other animals in the biological world. The newborn baby cries out in pain when suffering a stomach cramp, protests vehemently when the bottle of milk is taken away from it, complains bitterly when left hungry for a long time — all this with the aid of its affect system at a stage when it does not possess the benefits of linguistic communication. Affect, indeed, is the non-linguistic mode of expression and communication innate to an individual, inherited in a long evolutionary process.

Affect, moreover, assumes a set of unique features in the case of humans in virtue of their social nature that they share with other primates (and with many other groups of animals). In the human world, a baby requires a great deal of care and nurture before it can meaningfully participate in the complex social and physical environment it finds itself in. Even after that, a child needs protection and care in such basic needs as feeding and clothing, for which she needs to be in constant dependence on her caregivers, especially on her mother, father and a number of other caring persons. In all this, the child primarily depends on affective expression and communication to have her biological and social needs met with. What is more, she quickly develops a set of *psychological* needs that she feels to be essential in getting the biological needs satisfied — attention, reassurance, sharing of joy, initiation into new adventures .... all these blend with her biological needs into a complex maze signaling the beginnings of her self-system, based on a burgeoning set of preferences, not solely biological.

Once initiated, the process of generation of preferences assumes curious and even fantastic proportions — growing by means of *association*. What at initial stages of life were associations of a direct nature for getting the biological needs satisfied (*excited at the familiar sight of mother approaching, with anticipation of being fed*), now become more and more indirect and circumstantial (*taking a fancy for blue color owing to mother wearing blue dress* — a fancy destined to last the lifetime even without obvious biological benefit; *child taking fancy to a particular tune that the father hums while cuddling her* — gradually being transformed into a liking for a particular genre of music). The process proliferates into curious and strange channels quite unconnected with the 'logic' of conferring direct benefits to an individual (*an adolescent imitating the hairstyle of an icon whom he adores*, or *generating an aversion to persons with a certain physical build resembling that of a hated character (villain) frequently appearing on the TV screen*, or even *fascination for an addictive drug acquired from anecdotes circulating about the heroic figure of a student* *leader universally adored in the campus*).

> Associations are established between percepts by the simple expedient of juxtaposition and concomitance, and not by any other rule or 'logic' beyond these. In consequence, preferences generated by means of association are not subject to requirements of consistency beyond the ones embedded in juxtaposition and concomitance. For instance, *mother's smiling face* and *fancy for blue color* have a measure of consistency between them because of concomitance, but *adoration for a campus hero* or *fancy for an addictive drug*, or even (fascination for a currently raging fashion in dress), are not related by similarly discernible consistency. This is the reason why the set of preferences lodged in the mind of an individual assume fantastic patterns as he progresses in life.

In other words, originating in biological needs that require attention as much in the case an adult (*a persistent back-ache*, for instance, *generating acute anxiety so as to lead to a negative affect even in the case of a slight back-ache* long afterward) as for a child, preferences generated in the mind proliferate in avalanche-like rapidity with advancing age of an individual, creating a widely flung and complex web that influences our life in major ways. Preferences, in other words, are generated *iteratively* by the interaction of new experiences with the existing set of preferred entities and ideas.

Thus, preferences are generated in interactions with individuals by emulation and empathy (likewise, aversions are produced by disgust generated by individuals), and *most importantly*, by social and cultural means. What is more, a complex set of preferences are generated from *correct inferences and decisions* since such inferences and decisions are of great adaptive value in a complex and uncertain social and physical environment (likewise, negative affect is generated by inferential failures). Finally, a complex set of preferences are generated in the conscious mind as *moral* and *ethical* ones.

1. As mentioned, preferences are generated *iteratively* where, at any point of time, new preferences are produced on the basis of the existing ones by means of interaction of these existing preferences with the varied signals presented to the mind by experiences of diverse types. In this, the building up of preferences (and aversions too) closely resembles the development of the self-system, as indeed it should.
2. While the affect system responsible for the generation of preferences operates in the unconscious layer of the mind, preferences, beliefs, and dispositions can be consciously felt and invoked. Moral and ethical beliefs and dispositions, generated entirely by means of social interactions, add a hugely important dimension to our conscious being.

3. As we have seen in section 3.2, affect plays a significant but covert role in inference. At the same time, inference itself generates affect that plays the role of an 'internal monitor' in subsequent acts of inference. Thus if an inference proves to be erroneous when confronted with reality, a tacitly felt warning is generated when a similar error is repeated in a subsequent inferential act — likewise, a correct inference generates a positive affect and produces a 'feeling of encouragement' when repeated subsequently.
4. Indeed, we have an inbuilt urge to explore the world and to learn about it, that manifests itself in the form of the emotion of *curiosity*. This lies underneath the affect generated in a successful or unsuccessful act of inference.

The web of preferences quickly turns out to be a powerful factor in our mental life as the mind generates a hugely complex set of *emotions* that get associated with all our experiences, and finally becomes organically intertwined with an equally pervasive web of *beliefs* so as to take almost total control of our life.

The psychological value system that generates the preferences of the mind, the affect-based and cognitive mechanisms that produce the emotions, and the beliefs — all these get entangled to generate a vast and intricate set of *dispositions* in an individual, too complex and intractable to be described in clear and explicit terms. The dispositions define how a person responds to specific situations and experiences, i.e., her *behavior*.

Dispositions are varied and diverse — some of these operate unconsciously in a covert manner and some explicitly as consciously expressed inclinations and intentions; some get fixated in the behavior of the individual concerned in the form of 'character traits' and some remain fluid; some get lodged in the mind as *repressed* tendencies, and some find discernible expressions in social behavior.

Preferences, emotions, beliefs, dispositions — all these taken together form a vast and enormously complex system, operating within an equally complex social environment. And, as in the dynamical evolution of all complex systems, these generate an immense set of *conflicts* within the mind — conflicts that ultimately play out in the open arena of the society as well.

Beyond all these diverse sources responsible for the generation of preferences and dispositions, there remains a distinct mechanism — one based on *reason* operating within the folds of the conscious mind. It is reason that plays the dual role of generating a large body of preferences and dispositions and, at the same time, of moderating the conflict-ridden chaos that owes its origin to affect-linked preferences. This latter role of the conscience mind of bringing order to the unruly mass of affect-based preferences, beliefs, and dispositions finds its ultimate expression in the activity of the soul.

Preferences, beliefs, and dispositions go a long way to make the operations of the mind 'goal-directed' or 'purposeful'. In so far as the fulfillment of the goal does not generate conflicts, one can say that the mind is being 'adaptive' in going after its goals. The adaptation can be either social or an adaptation to our physical environment, though the two are intimately linked. A part of our adaptation to the physical environment has been achieved by means of genetically transmitted traits. Another type of adaptation of great importance takes place by the agency of the mind in virtue of its capacity to decide and to infer correctly in understanding the ways of Nature. This is why we use the term adaptation (or 'adaptive') in this book in a broad sense.

5.2.3 The mind: conflicts and their regulation

States and processes in a complex system are burdened with an enormous number of *conflicts*, and the mind is no exception. Processes in the mind are principally directed towards the satisfaction of the huge set of preferences lodged in it — ones based on the incessant and infinitely complex neuronal processes in the brain. In keeping with what has been said earlier, mental processes correspond to interactions among large neuronal assemblies (or *modules*), where it is these interactions that play out in the mind toward the satisfaction of all its preferences. Preferences are produced in neuronal assemblies in the brain, all modulated by the affect generated in specific sub-cortical and cortical regions, and most of these preferences are, to begin with, *independent* of one another. Interaction among neuronal assemblies find expression in the form of conflicts and (also) harmonies among these preferences.

An instance of the role of conflicts in the physical sciences is to be found in *disordered systems* in the form of *frustrations* that stand in the way of co-operative interactions among the units making up a system, leading to *order-disorder transitions* in the latter. In virtue of the pervasive feature of frustration, the order-disorder transitions in disordered media acquire a complex set of novel features when compared with similar transitions in *ordered* systems. In the case of model systems referred to as 'spin glass', these complex features find expression in the form of *Parisi order parameter* (see, for instance, [130]).

In the present context, we confine ourselves to the consideration of conflicts and harmony among the preferences and dispositions in the human mind, without overt reference to interactions among underlying neuronal assemblies, mostly because the correspondence between neuronal and mental processes is largely a matter of guesswork.

In virtue of the fact that preferences are generated in the mind primarily by means of associations among percepts, without any overt requirement of consistency and harmony imposed on their formation, there arise all-round conflicts and inconsistencies among these. For instance:

My son has developed a great fascination for mathematics but doesn't have an appetite for problems in geometry. His mathematics teacher, without trying to make him attracted to the subject of geometry by gentler means, forces him to devote an inordinately long time every day to geometry problems. In consequence, he is torn between opposing tendencies in respect of his involvement with mathematics.

Or, again:

I have a great liking for musical concerts but in recent years, have developed an equally strong reservation against the expression of loud and vociferous appreciation by young people attending concerts in our town.

It is no use citing further examples of conflicts in our mental life and behavior, but it is of overriding importance to take note of the fact that conflicts occur *at all scales* in our mind — right from the level of our preferences for smell and color and contrasting tastes in food items, up to our commitments and dispositions in social, political, cultural, and religious affairs. Such conflicts develop disordering effects in an individual and in society at large — recall that social groups have 'minds' and 'self's of their own (refer back to section 1.15; see also sec. 5.2.8 below), with their characteristic preferences and dispositions. The fact that, in spite of these, the minds of individuals and social groups at various scales of aggregation retain a measure of integrity and coherence owes its origin to the operation of pervasive processes of *regulation* within the mind and within social formations.

Processes leading to inconsistencies and conflicts on the one hand, and those of a regulatory and integrative nature on the other, occur in profusion in the human mind and in social formations of diverse descriptions. In the case of the mind of an individual, these are based on interactions among neuronal assemblies while in the case of a social formation, the corresponding interactions are among sub-groups and dominant individuals within the said formations.

> One has to distinguish between integrative and regulatory activities of the mind. The former is necessary in order that localized excitation patterns can be assembled together so as to give rise to coherent mental ingredients such as concepts, beliefs, and dispositions in the mind. In this, the mind acts as a huge and complex factory with a large number of production shops, each producing intermediate products, all of which are to be appropriately processed and combined so as to yield finished products of various descriptions.
>
> Regulatory activities, on the other hand, are necessary for the resolution of mutual incompatibilities. For instance, the production plan of some intermediate item may not be consistent with that of some other in view of the fact that the demand of some particular finished product has suddenly dropped in the market. This calls for a swift order from administration for revised production schedule in the production shops concerned. If the regulation fails to get implemented, the running of the factory becomes uncertain, with a host of problems raising their head.

When considered in purely mental and psychological terms and invoking a mathematical analogy, all these processes of contrasting types make for complex trajectories described in the 'phase space' of the mind. Continuing to refer to the same mathematical analogy (section 1.16), the evolution of the mind by means of thought processes occurs in a 'phase space' of vast dimensions and with local growth rates ('Lyapunov exponents') spanning a wide spectrum of positive and negative values, that changes from one region of the phase space to another — the growth rates, indicative of the sensitivity of the mind to perturbations arising from various causes, arise due to the operation of affect and emotions activated under various contexts. As mentioned earlier, the conflicts between preferences and dispositions, and the integrative and regulatory processes exerting their moderating effects make the mind proceed through complex sequences of local instabilities and stable regimes, generating an infinite diversity of behavior of an individual.

The regulatory activities do not all occur within the conscious layer of the mind. Affect and emotions exert their influence in a major way in unconscious regulation too. This is apparent in the way the unconscious mind generates decisions in the face of disparate choices facing an individual (section 3.1). However, some of the conflicts do not find resolution at either the unconscious or the conscious level, thereby threatening the integrity of the mental set-up. This is when the mind forcibly *represses* its own dispositions in a ceaseless effort at maintaining an over-all coherence and *status quo*.

Similar suppression of dispositions is common in the context of social regulation of conflicts, where administration, law and order institutions, juridical institutions, *power relations* based on sheer force, and rules imposed by social and cultural norms, all act together to generate harmony, regulate conflicts, and to repress numerous social trends arising from conflicts.

Some of the regulatory activity of the mind occurs at the unconscious level, such as putting a leash on incompatible emotions operating unconsciously, or long-term repression of a preference that the unconscious mind assesses as deleterious. On the other hand, a great deal of regulation is done at the conscious level too. Indeed, the conscious mind is specifically suited to assess the consequences of harmful dispositions, especially those with social implications, and routinely keeps those on check. For instance:

I am strongly inclined to make a showdown with my friend who has recently cheated me in a business deal. However, I curb my impulse by mentally visualizing the implications this may have in respect of my long term business interests.

What is more, the conscious mind is a great device to repress our emotions and dispositions on a long-term basis and to resolve conflicts by planning and control. For instance,

My financial failures in the recent past has led me to bouts of excessive drinking and occasional foul behavior with my wife. I have a strong feeling that these are sure to bring disaster to my life in the near future. I exert supreme efforts to repress my frustration and keep a calm exterior. However, I also feel that this forced calmness is showing up in untold small ways in my social behavior with respect to my children, friends and acquaintances. I take a vow to reason with myself and decide to visit a counselor for advice, if necessary.

> Introspection and self-examination requires an awareness of one's own vulnerabilities and lapses, and are comparatively rare occurrences in a person's psychological life; the default mode network (DMN) provides a platform where our consciousness meets our self, though an exploration of the self for the purpose of re-structuring it requires a motivated self-examination of greater intensity; if one's own self is inaccessible to an individual, it may, in a sense, be less so to others — others may read us better than we ourselves can, which is why one needs to look at oneself from the viewpoint of others (others often know us better than we ourselves do); for this, however, one has to cease judging others.

Occasionally, however, *we fail to resolve conflicts*, mostly those that we are not aware of. This leads to repression of dispositions that eat at our heart. Of equal relevance are conflicts that we diagnose ourselves with and try to resolve where, however, the process of resolution leads to the generation of unconscious behavioral issues that we fail to recognize — these cause us to feel resentful and isolated but we fail to recognize that the reason lies more within us than in others, whom we implicitly keep on blaming.

While the mind operates within the confines of the body, it expresses itself in the form of *behavior*, which brings the individual face to face with other human beings — with the rest of the *society*. All the conflicts of the mind, all the repressed dispositions, eventually find expression in the form of behavior, where the context now shifts to interactions among human beings belonging to larger social groups, starting from the family up to ever larger groups and social formations, ending up with the entire mankind. The behavior patterns of an individual assumes relevance in the context of those social groups whose members interact with him directly, but the relevance spreads further by means of indirect interactions. The aggressive attitude of an individual, for instance, can spread far beyond the confines of the group of people in direct interaction with him, just as the message of peace and harmony spread by one can propagate far and wide.

The unresolved conflicts within a person are expressed in the form of *traits*, *attitudes*, and *behavior patterns*. Typical of such traits are anxiety, boredom and impatience, resentfulness, secretiveness, aggressiveness, possessiveness, and a host of other characteristics that generate conflict and mutual antipathy among fellow human beings. Traits, attitudes, and behavior patterns are all associated with the self of an individual. And conflicts in society are consequences of unresolved conflicts within the selves of individuals where one has to recall that the self of an individual and the group-selves of larger social groups to which he belongs are tied to one another.

> Recall from sections 2.7.6, 1.15, 2.7.8 (see also sec. 5.2.8 below) that a social group has a 'mind' and a 'self' of its own, analogous to what one finds in an individual. That self, including the preferences and dispositions of the social group (we refer to it as the 'group-self') is generated by interactions among individuals and subgroups within it and, in turn, the group itself induces a set of preferences and dispositions in an individual belonging to it that we have referred to as the 'group-induced self' — it is the group-induced self that is indicative of the continuity between the self and the non-self. Recall further that it is the set of preferences and dispositions that define the self — either of an individual or of a group.

Social groups have their own repressed emotions and dispositions that often break out in the open. Whole nations, for instance, are known to have been possessed of grotesquely aggressive fervor to 'regain' lost glory by conquering neighboring land and subjugating innocent people. Diverse social formations have their own 'character traits', as distinct from one another as those of individuals. And, just as the mind of an individual has regulatory mechanisms operative within it, so are social groups equipped with more or less elaborate regulatory mechanisms to resolve and to manage conflicts and contrary tendencies. The most elaborate of the regulatory institutions is the *state* that has the explicitly formulated job of maintaining the status quo in the society ruled by it and to cause 'advancements' without disturbing the status quo in any major way. The society may harbor within it a number of conflicts that cannot be eliminated without causing a radical disintegration. The state institutes mechanisms such that life may go on without such disintegration taking place just as the mind of an individual represses a number of elemental conflicts that would otherwise threaten the very fabric of existence. At the same time, the state equips itself with the means and institutions (the administration, the parliament, the judiciary) by means of which the myriads of lesser conflicts can be resolved without snowballing insanity developing within society.

All this pervasive conflict and all this pervasive mechanism of regulation results in the need for self-examination for individuals and also on the part of social formations, in the absence of which regulation is overpowered by conflict. This is where the *soul* comes in.

5.2.4 The soul: the highest level of regulation in the mind

The soul is where, in a manner of speaking, *reason meets affect* by intent. We have, in earlier sections of this book, spoken of reason and affect being intertwined in an inseparable tangle, but that entanglement is in the very nature of things without there being special mechanisms where reason is made to *intervene* on affect. We will speak of the mind of an individual first.

The unconscious and conscious layers of the mind are incessantly engaged in resolving myriads of conflicts that arise in virtue of incompatible preferences having been generated in it. Recall that preferences are generated not in accordance with any comprehensive scheme or principle in operation but by the simple expedient of association: whatever is juxtaposed to or concomitant with an item of preference is itself acceptable as a new item of preference, though the state of the mind at the relevant point of time plays a role in this as a filter of sorts — among a host of perceived items in mutual association, only one or a few are selected out as new items of preference. We also recall that the psychological value system that generates preferences works in a complex manner (in association with psychological processes based on diverse neuronal assemblies) to associate emotions with percepts and that, further, affect and emotions are blended with beliefs of diverse descriptions to produce dispositions of the mind that have no in-built reason to be compatible with each other. This is analogous to personal propensities and inclinations of members of a religious community being mutually incompatible in spite of the over-all unity of their religious orientation as members of the community.

However, the mind finds a *modus vivendi* in spite of all these conflicts among contrary preferences and dispositions, based on interactions among neuronal assemblies established by means of neuronal circuits and pathways. We recall that neural connectivity is experience-dependent and is possessed of *plasticity* that, more likely than not, generates counter-currents in the mind in a kind of self-organizing process whereby most of the conflicts are rendered innocuous so that the mind can operate on the whole as a coherent system. This is the basic regulatory process whereby conflicts are managed at the 'grass-roots', so to say. Added to this are the regulatory processes set in motion by *reason* operating in the conscious mind. A huge set of preferences are generated by reason-based activity in the conscious mind. Reason-based regulation reaches its culmination in the activity of the soul.

Notably, regulatory processes occur *at all levels*, playing out in a hierarchy. What is left unresolved at one level and threatens the integrity of mental processes is subjected to the next level of regulation. Such higher-level regulation, for instance, is made to operate on an individual by *social* means, where a huge hierarchy restrains contrary preferences of an individual to be expressed in his behavior. Morality and ethics, social ethos, cultural traits, religious tenets, all are transmitted on to the mind of an individual so as to operate in a vast and invisible hierarchy preventing mutually incompatible preferences and dispositions to compromise the coherent activity of the mind and to prevent disruptive behavior.

In all this, *the soul appears as the highest level of regulation* operating on the self of an individual.

As we see it in this book, the soul is not anything transcendental, other-worldly, or eternal. It is as earthly as the conscious mind making a critical examination of the self and its intransigent dispositions, mostly by means of *inferences* and *decisions*.

> We recall here the sense in which we use the term 'disposition' in this book. Dispositions are built up on the basis of preferences and beliefs, aspirations (wish to fulfill a preference), unfulfilled wishes, and so on, and determine the way an individual responds to specific situations and experiences. For instance, someone may be unusually aggressive in facing an adverse situation, or may be found to be too suspicious in dealing with fellow beings. Our dispositions determine to a large extent our behavior.

The conscious mind operates largely on the basis of reason, i.e., rules of appreciable generality, based on perceived relations of various kinds among concepts. In other words, reason is a way to establish correlations between concepts without overt influence of affect. In looking at percepts on the basis of mutual relations between concepts, the mind *objectifies* whatever it perceives so that the relations of the latter with other concepts in the conceptual universe can be established and reflected upon. In this, the conscious mind has the capacity to objectify the *self* as well, when the self becomes an object of reflection and deliberation. This is the point of departure for the soul to operate.

As mentioned above, there are myriads of ways in which the unconscious and the conscious mind routinely and incessantly resolves conflicts generated between preferences and dispositions of ours, and thereby provide a *modus vivendi* for the mind to operate as a coherent whole. However, there are repressed dispositions that remain inaccessible to the unconscious mind and, to a large extent, to the conscious mind too, that evade resolution for too long. We have already mentioned that the ability to repress unresolved conflicts may, up to a certain extent, be looked upon as an *adaptive* one, since such conflicts may otherwise have disruptive effects on the mind. However, repressed or not, conflicts always tend to be potentially subversive. Many of those, including the repressed ones, leave their mark on our behavior that often estrange us in some measure from our fellowmen even though we ourselves may not be aware of the fact and the cause of such estrangement. If and when we become dimly aware of it, we initially tend to blame persons around us for their 'inability to understand' our mind.

> Recall that our social cognition possesses the feature of being mostly affect-driven, in contrast to our cognition of the physical reality. This is in consequence to the fact that, in the social context, we often have to form inferences on the basis of very limited evidence, and don't have the advantage of repeated experimentation. In consequence, our inferences in the social context are heavily underdetermined by evidence, where affect rather than reason assumes the upper hand. Recall, moreover, that the self is inordinately sensitive to the social emotions of shame and pride (section 2.7.4).

In consequence of all this, our affect-driven self slowly but surely becomes a liability to our continued social existence and for the coherent functioning of the social groups to which we belong (*the CEO becoming a liability to the business enterprise in virtue of his uncontrolled hubris*).

This, precisely, is where the soul is required to take over. The soul is synonymous with the conscious mind being aware of continuing troubles and turmoils within the mind at large, looking closely at the self and restructuring the latter, at least partly, so as to relieve it of tension and to make the self confront the world without generating further tensions in the latter. In this, the conscious mind examines the self as an object, by reflecting upon its relation to a host of entities in the social world, and to ingredients residing within the mind itself. This is no ordinary job that the conscious mind undertakes to perform, since it requires it to access, in particular, the *unconscious mind*, which seems to be a contradiction in terms and poses a major challenge. We now take up in succession three basic issues relating to the possibility of conscious mind examining the self *with the intent* of restructuring it: (a) how the operation of the soul, where reason is supposed to meet the self, differs from other instances of reason and affect routinely working hand in hand, and of conflicts being routinely managed in the unconscious and the conscious mind, (b) how, in deciding to examine the self with a view to restructuring it, the mind has to escape from the 'bootstrapping' effect of self-reference, and (c) how the conscious mind does the job of 'accessing' the unconscious mind.

Self-restructuring and self-improvement is, in a sense, close to what Robert Kane calls 'self-forming action' (see [49]), to which he attaches great relevance in the context of free will.

Before we address these questions, we underline that the soul constitutes the highest level of regulation of conflicts raging within the mind, where processes of regulation occur hierarchically at very many levels. Thus, there are the unconscious processes of regulation where the cognitive-adaptive unconscious tries to ensure that the affect-ridden unconscious does not make life difficult for it. Even as the unconscious mind mostly runs in parallel, without any central organizing process operating in it, it is still a complex system with some minimum of interaction between modules running in parallel (recall that interactions between neuronal assemblies are among the fundamental principles on which the mind is constituted) and, in consequence, has to have some degree of self-organized complexity characterizing its activity — self-organized complexity is a routine matter in complex systems even without any sign of life, and without the presence of any centralizing principles involved. Indeed, even the perceived distinction between the cognitive-adaptive unconscious and the affect-ridden unconscious can be looked at as a reference to self-organized complexity, telling us that there is *structure* at every level in a complex system.

1. Of course, the unconscious mind (or the mind at large) has structures in it not solely in virtue of it being a complex system — a protracted process of evolution generates genetically controlled activity that makes the brain a structured collection of neuronal assemblies. Subsequent processes of experience-dependent plasticity, including epigenetic processes, make the brain generate (in a manner that truly defies comprehension) the mind with all its infinitely complex structure. It is the latter set of processes, built upon evolution-generated modules and capacities of the brain, that can be described as being the result of the brain's self-organized complexity, where we leave as implied the role of a host of associated physiological systems of an individual, including the neurotransmitter systems.
2. While acknowledging that the neuronal processes in the brain occur within the framework of the structural features imparted by the process of evolution, the latter itself, from a broad-based point of view, can be looked upon as an instance of self-organized complexity in the realm of the biological world.
3. Structures arise by self-organized complexity in a system in apparent defiance of the tendency of all macroscopic system to approach a state of thermodynamic equilibrium. However, fluctuations arise on all scales over sufficiently long time ranges, and generate structures compatible with the relevant boundary conditions representing the effect of systems with which a given complex system interacts. Moreover, the boundary conditions themselves may vary in time, resulting in local structures representing space-time limited regimes of stability. The very term stability implies that corrective processes are set in motion — also in virtue of the complexity of the system concerned — counter-acting the effects of

local fluctuations (over spatial and temporal scales smaller than the scales characterizing a structure) that tend to disturb the stability. In other words, not only are stable structures within limited space-time regions possible in a complex system but corrective processes are also ensured in virtue of complexity. There may arise special instances of structures with a more extensive regime of stability, such as the stationary states close to equilibrium, or time-varying states further away where stability is once again ensured by means of corrective processes. Particular instances of these are the homeorhetic states that Waddington spoke of.

Ontologically speaking, processes based on neuronal interactions generate the various levels of structure in which conflicts and their regulatory mechanisms operate, the lowest among these being those in the unconscious mind. At higher levels are conflicts that are more or less directly perceived — one whose existence and consequences the mind becomes aware of in the course of its routine inferences and social interactions. For instance:

On my way home from office at the end of a day's work, I feel like having a cup of coffee at some roadside cafe; but then I recall how thrilled my wife feels when I take her out for an evening stroll and we have coffee together; I immediately change my plan of having a relaxed cup of coffee on my own and hurry home so as to take my wife out for coffee to be followed by a film show later in the evening. Sitting in a restaurant all by oneself, relaxing and enjoying loneliness in a crowd is certainly an attractive proposition, but in the present instance, its attraction is overridden by the prospect of an evening out with a radiantly smiling wife. While this is one instance where a conflict between preferences generated in one's mind is rather easily resolved, there are myriads of instances in on's social interactions where conflicts require some degree of deliberation to get resolved, over and above spontaneous and affect-based processes of choosing between disparate alternatives involving contrary preferences. Life is full of instances when the conscious mind has to take a long-term view and employ reasoning and deliberation, even overriding affect-driven propensities operating on a shorter time scale. Many are the occasions when we give in to more elemental inclinations, forsaking the path of prudence and deliberate intent, simply because prudence is psychologically costly. However, in most of such occasions, we turn back to reason and reflection as we face unpleasant consequences.

Highest in the scale of difficulty in diagnosing problems within ourselves and sorting those out are the ones where conflicts among affect-generated preferences are not perceived as such by the conscious mind and give rise to dispositions hidden even from ourselves — ones that nevertheless exert cumulative influence on our mental and social life. These are the situations when one gets progressively 'estranged from oneself' because of a lack of comprehension of the contrary forces raging within the self. Not being able to examine the self with sufficient clarity while feeling vague but mounting uneasiness in our social interactions, we get into a more or less prolonged phase of *denial*, all the while blaming people around us for our troubled existence. Desperation, despondency, and resentment at everything around us starts eating at our heart. And then it spills out to our social interactions, polluting and infecting the 'social psyche' too.

> In other words, the conflicts within the self of an individual, now get expressed as conflicts in the social milieu. Considering a social group to which an individual may belong, preferences and dispositions of the group as a whole are formed by means of interactions among individuals and among sub-groups, and come in conflict with one another just as the dispositions of an individual do. Such dispositions produced in a social group generate the 'group-self'. Repressed and hidden conflicts within a person often give rise to similarly repressed and simmering conflicts in the group-self, of which more below.

It is here that the soul is to be set in operation, thereby constituting the highest level of regulation within the mind. In this, the soul can make use of ideals, morality, ethics, spirituality, and religion, all of which the society makes available to it. However, these are still not sufficient for the soul to begin operation, because it is a question of *intent* that is still to be settled. The soul has to make the mind engage in an unpleasant task, because the mind has already formed an aversion toward looking at its own repressed conflicts and is in denial — one now has to deliberately overcome that aversion. In other words, there are two major aspects to the soul being set in action — the problem of being aware of dispositions that need to be set right, and the one of forming the intent to engage in the job. Both of these involve the seemingly intractable problem of self-reference and 'bootstrapping'.

One cannot draw a complete picture of oneself because that involves an infinite regress. And one cannot lift oneself by pulling at one's bootstraps. How can the self, which is in turmoil and in denial, form an intent to confront itself? When we speak of the conscious mind looking at repressed conflicts, it is actually the conscious part of the self that is to confront a deeply hidden recess of the self-system which, though a difficult proposition, is not an impossible task. It is the soul that alone can pull it off.

1. In the operations of the soul, affect and emotions play a dual role: on the one hand, these are responsible for the conflicts that need to be resolved and, on the other, these very resources help in forming the resolve and commitment of the mind whereby the latter generates the intent of looking at the self and restructuring it. For instance, morality and ethics generate a set of preferences of a very special kind that override the barrier against self-examination and make possible the process of restructuring of the self. Thus, in brief, affect and emotions are responsible for the generation of conflicts at one scale and for their regulation at another. Indeed, it is the proliferation of *scales* that make possible their dual role. At the same time, it is *reason* — shared preferences and beliefs such as morality and ethical principles — that is fundamental to the activity of the soul. This, once again, is indicative of the strange interaction between affect and reason that characterizes the mind in all its activities.
2. **Morality and ethics.** Morality and ethics are instances of shared principles that constitute the *reason* by means of which the soul resolves repressed conflicts. It may be noted that a repressed conflict is often one between a set of preferences of the mind and a set of social norms, as a result of which the preferences are kept hidden in the mind — hidden in the sense that they do not get expressed in behavior. It requires socially induced reason such as those encoded in principles of morality and ethics to examine and expose the roots of those preferences and to deactivate them from being expressed in behavior. Spirituality, scriptures, religion, *and* common wisdom, all indicate that this is important, but individuals and societies have to do it their own way — there is no scientific treatise ever written on how to do it.

   At times, 'logic' is supposed to imply that morality and ethics have no 'basis' in the same way as science has a basis in reason and experimental evidence. But then, 'fashion' has likewise no basis, faith has no basis, religion has no basis. 'Logic' does not explain how, in spite of huge achievements having been made by science, fashion, faith and religion continue to have increasing following in men. It takes some depth and delicacy to realize that morality and ethics have very real relevance to human existence, regardless of the fact that there are no ultimate standards set for these, that these differ from time to time, from place to place, and from one culture to another. It takes some acuity to realize that religion and spirituality have real relevance to human existence, regardless of the fact that there exists a perfectly tenable point of view that science can do without a creator.

   We don't live by reason alone — we live by affect *and* reason. That is precisely why morality, ethics, spirituality, and religion have such huge relevance in our life — a journey through an impossibly complex reality.

   But this book has to draw a line here since further deliberations on morality, ethics, spirituality and religion lie entirely beyond its purview.

5.2.5 The soul: when reason meets affect by choice

In the operation of the soul, reason can be said to meet the affect-ridden self *by choice* or, equivalently, by *intent*. This makes the reason-affect interaction very distinct indeed.

We have remarked time and again in this book that reason and affect are notionally distinct but not operationally so. We acquire reason by rising above affect so as to look at percepts not in terms of innately generated responses to inputs received from the world but by learning to focus on relations between stimuli of various different types, between percepts, and between concepts built out of those. But we do not rise above affect by divine decree, we do so bit by bit, by learning to look at the world not solely through our own eyes, but through the eyes of others. This impresses on us, over and above our own personal preferences and emotions, socially acquired preferences and emotions as well. In consequence, the reasons we follow are no 'universal' ones, but those that give new dimensions to our innately operative 'logic' based on affect — the logic that develops solely on the basis of associations acquired in personal experience. But there is no absolute or sharp dividing line between what is innate and what is acquired — a newborn is open to social influence from the moment it appears in this world — the influence exerted by persons in close relation. This is why affect and reason are inextricably blended with each other — the two reflect distinct *styles* of functioning of the mind that the latter makes use of in ways solely depending on the needs of situations it faces from one point of time to the next.

The relation between reason and affect is strange and curious indeed. Even in the workings of the unconscious mind, the affect-ridden unconscious and the so-called cognitive-adaptive unconscious illustrate this relation in a nascent form. We have seen earlier that even the unconscious mind has an innate ability to respond to statistical regularities in the inputs it receives from the world — indeed, this ability itself is involved in the workings of the affect system. In other words, even the unconscious mind operates in two styles that we find in clearer contrast in the workings of the conscious mind when the latter joins forces with the unconscious.

Of course, this 'contrast' — the dichotomy — is not 'really' there in the mind; the mind is happily ignorant of all this talk about 'affect' and 'reason' except when it is called upon to describe its own workings in explicit terms. The fundamental takeaway from all this is that a system (say, the mind) in itself is one thing, and a description of it in verbal or explicit terms is quite another.

This, fundamentally, is at the root of the distinction between the noumenal and the phenomenal reality that we talked of in chapter 4 — in comprehending and describing the noumenal reality we do so in terms of *categories* born out of our *interpretations* of it.

The next 'higher' level in which this intertwined relation between reason and affect gets expressed consists of the making of myriads of inferences and decisions that we routinely perform in our daily life without which life would be impossible to go through. This is what we saw in chapter 3 when we attempted to see how affect is implicitly involved in decisions and inferences made by the reasoning mind, mostly as a matter of course.

There is a spectrum of still higher 'levels' in which reason is brought to bear upon affect in our journey through life. For instance:

I must not indulge myself with unplanned quantity and quality of food — it will increase my sugar level, which is already quite high. This provides an instance where a piece of evidence-based reasoning is employed to resolve the conflict between indulging in preferred food items and an affect-based predilection to stay fit and healthy. Or again:

You must not talk of others in abusive terms — this is not decent and will create enemies all around you. Here the intervention of reason on affect is in deontic terms — once again, the basic idea is to resolve the conflict between affects (a predilection for abusive outpouring and the desire to avoid social disharmony).

In other words the interaction between affect and reason occurs at various levels, where there is some degree of regulation involved at each level and where the extent of regulation varies across the levels. In the above paragraphs we have spoken of the level within the unconscious layer of the mind, the level involving our decision-making and our inferential activities which is partly unconscious and partly conscious, the level of our beliefs (doxastic), and the level of our obligations (deontic). The soul combines in itself and supersedes all these various levels in regulating our self.

At the highest level of all the regulatory processes in the mind is the soul in operation. The soul once again tries to make the mind respond to the appeal of reason where now the reason is mostly of the deontic type — moral, spiritual, religious, and above all *plain sanity*:

There must be hidden some innate aggression somewhere inside me — people are getting wary of me and are avoiding my company — I must get hold of myself and see what lies underneath this feeling of aggression. This is plain sanity and, at the same time, a genuinely deep realization — a rare admission of truth. Our deepest realizations of ourselves are simple ones — mankind does not need formulas of abstract algebra to realize and to act upon its deep-going blunders. Still, simple realizations are the hardest to come by.

The spectrum of 'levels' of regulation (all talk of 'levels' is, once again, of epistemic origin) is characterized by differing *time scales* over which reason interacts with affect. For instance, unconscious regulation of affect occurs almost instantaneously — 'on the fly', so to say. The ceaseless making of decisions and inferences in our daily life is not much different since reason and affect have to work hand in hand in these processes. The deontic regulation of our mental life needs more deliberation in getting the mind prepared. And, beyond all this, the reconstruction of the self that the soul undertakes is contingent upon a great many factors — sometimes it never happens at all. The choice and the challenge of confronting the self may — sadly — never be taken up in a lifetime.

In closing this section we note that the reason — mostly deontic in nature — that is brought to bear on affect in the operation of the soul is not itself divested of affect. It is one part of the self that takes a look at another, and the self is ultimately an affect-driven entity. The conscious part of the self has the special characteristic that it makes use of reason while being rooted in affect.

5.2.6 The soul: evading the 'bootstrap'

The operation of the soul amounts to willing the mind to do something for which the will is not there — somewhat like lifting oneself by pulling at one's bootstraps. It mostly concerns a set of unfulfilled preferences repressed in the mind because the fulfillment of these would either cause unacceptable clash with the *status quo* or with one's self-image based on pride, bringing shame and humiliation in its wake. This is a conflict that the self shrinks away from. Maybe a person tries it once or twice and then desists on having a glimpse of what it may involve — loss of 'peace of mind'. Perhaps it involves radically revising a set of foundational beliefs, ones that may be perceived as being able to put the entire existence at stake.

However, we will steer clear of the *logical* problem of self-reference and will focus on how self-examination and self-improvement occurs in situations occurring *in practice* in our life.

> The logical problem assumes significance when an entity attempts at describing and modifying itself in every part. This is not really a problem with the mind where one part makes a map of another by means of interactions through signals propagating along neuronal pathways. There also remains the possibility of one part of a neuronal assembly *sampling* itself by means of re-entrant loops.

The 'bootstrap' issue retains a measure of practical relevance because, as mentioned above, the will has to struggle here against the force of repression where the latter thwarts the will itself. This is where the mind has to resolve or smooth a conflict *by intent* — going against its own inclinations, as it were, and it is no easy matter trying to understand how that can be done. Self-examination and self-reconstruction is in the nature of a learning process — one of a special kind where one has to *set* the mind on to the course of learning. This is like the case of the reluctant student being made interested in his studies, where he has to make up for a lot of lost ground.

This, precisely, is why the process of confronting and restructuring the self is, in a very real sense, a relatively *rare process* — it doesn't occur every day and is contingent upon a set of circumstances that don't come about automatically. Indeed, one may even say that there is an element of *randomness* in a person being presented a chance to self-examine — much like the reluctant student meeting a friendly teacher who opens a window to his mind and helps him undergo a radical transformation. Once the crucial window is opened, the student would need no further spurring or cajoling and would continue with his studies with a frenetic zeal. Self-examination, however, differs from this to some extent since it is not a one-off affair but needs sustained motivation in order to be effective. And it often needs contingent circumstances coming one's way repeatedly, like a run of good hands in a game of cards.

That self-examination is learning of a very special kind may be seen in the case of a student interested in many subjects, repeatedly rushing to the library and immersing himself in piles of books, but still falling short of taking a hard look at *himself*, failing to examine his telling shortcomings in social life, and to lock horns with his own deeply held beliefs.

So, what does it take to be able to engage in self-examination and self-improvement? There is no set recipe for this since one needs a set of contingent circumstances for this to happen.

Fundamentally, there occurs a competition between a set of hidden preferences denied fulfillment, most of which are repressed in the mind and, preferences calling for continuation of the status quo, involving aversions against upsetting the self-image based on an existing balance between shame and pride. As in every such competition, it is not only the identity of the opposing factors that is relevant, but their *strengths* too, because it is a change in the relative strengths of these factors that initiates the process of their realignment — a process where major changes in the life-pattern of a person may be involved.

The mind and its various components are in a process of incessant dynamical evolution as a complex system, where ebbs and tides occur on all scales due to the interactions between an inordinately diverse set of factors and due to the changing context set by external factors as well.

> We repeat that the term 'external' is to be taken in an inclusive sense and not literally in the sense of the world outside the mind. Thus, in the evolution of one part of the mind, some other parts may act as part of the context. The distinction between the internal factors in a dynamics and the context is itself not absolute and depends on circumstances.

It is this complex dynamics of factors that, at any point of time, determines the relative strengths of the competing forces involved in the process of self-examination. More precisely, it is the *conjunction* between the state of the relevant parts of the mind and the context set by state of the relevant parts of the external reality that assumes significance in initiating the process of self-examination and of restructuring the self. This conjunction is, in a sense, *fortuitous* and largely unpredictable. A person may go on committing offense after offense against the society of his fellowmen without feeling remorse or the necessity of reorienting his behavior pattern when a relatively insignificant event may, in a strange way, set him thinking. One never knows when one's mind becomes receptive of the idea of a changeover, more so when the relevant competing factors are already in a meta-stable configuration in the course of their complex dynamics so that a small pull here or a small push there may change the entire balance of forces — much like the change in the course of a protracted warfare where hidden factors (the supply of food, the availability of arms, the morale of the fighting forces, environmental and geographical factors, and a host of other complex circumstances) operating in unpredictable ways may create just such a meta-stable configuration.

In brief, the process of self-examination and the restructuring of the self may have strange and indeterminate beginnings. It is the same indeterminacy that governs much of our mental life, our behavior pattern, and our interaction with our fellowmen in society. What appears to be determinate within limited domains of space and time, acquires strange aspects when viewed in the context of wider domains.

I was suffering for long from a vague sense of inner turmoil and was feeling lost under a series of financial blows when one day I saw a little girl building up a sand pile on a sea-

beach that fell away repeatedly. The calm demeanor of the girl who went on adding sand patiently and was not flustered when the pile fell away, added to the serene atmosphere on the beach, hit a chord somewhere within me, and I suddenly saw what was wrong with me and my mental make-up. I discovered that I had been guilty of gross misbehavior with my wife and had gone on ignoring her sound advice regarding my business affairs, relying instead on a friend who had been systematically undermining my trust. There followed a protracted process of re-assembling of large parts of my mental and social life.

Two young individuals happen to meet an elderly man who narrates to them the story of his life, in which he underlines how he confronted himself on several occasions at critical junctures and, in consequence, gained in courage, conviction and inner tranquility. Of the two young persons, one is deeply moved on hearing this and starts looking at his own beliefs, personal preferences, and drives, thereafter delving ever deeper into his own developmental history and gradually coming to understand and engage with the conflicts and turmoils within him. But his friend remains unmoved and goes on living the same old life without self-doubt bothering him in the least.

Inscrutable indeed are the ways of the soul.

5.2.7 The soul: accessing the unconscious

In examining and reconstructing the self, we often have to discover our repressed and hidden dispositions, including our own preferences and deeply held beliefs, where we are required to discover the way our emotions work. As we mentioned in sec. 5.2.6, it is not an easy and routine matter that such a process can at all be initiated in a person. Equally important is how the process is to be continued successfully. This is what is commonly referred to as 'discovering oneself'.

We discover our own dispositions and emotions fundamentally *by inference*. We look at our own behavior and try to infer what lies underneath. We look at ourselves through the eyes of others around us, trying to guess what they think and feel about the way we are constituted. As we infer how our emotions affect our behavior, we consciously modify the behavior pattern by force of will, thereby hoping to turn the tide of emotions. There are age-old pieces of wisdom how all this can be done, to which we add our own keenly felt need to come out of our emotion-ridden behavior pattern and our own observations and learned truths about ourselves.

In my efforts at re-assembling my own self, I found out that whenever something goes wrong in my life, I 'discover' elaborate cause-effect relations involving the motives of people around me and set out to 'correct' those people. I have now taken a vow to observe reticence when I catch myself teaching others, including my wife and my children. As I achieve partial success in this regard, I find that I can now look more meaningfully at the root causes of problems arising in my life and deal with those.

There are no text-books devoted to the subject of how the soul is to be made to operate because that is a question as deep and as complex as life itself. There are age-old pieces of wisdom, anecdotes involving the life of others, sound advice of people who have gone through and endured life silently beyond the limelight of success and glamour — all of which we ignore dismissively as we go about own own business seeking success, achievements, and glamour ourselves. We go about achieving success, and not fulfillment in life, taking no heed of the fact that it is fulfillment and not success where our emotions find themselves with minimal conflict with the world and society around us. We go after knowledge, dismissing realization and wisdom. We go on judging and blaming people without trying to understand them. We categorize the world in binaries, without being aware of the infinitely rich and extended spectrum spanning the two poles of a binary. We exert great efforts to be precise and to be explicit in our understanding, without devoting time to what we feel implicitly, because the implicit does not speak of an 'analytical' mind. We analyze without pausing to look at the whole and without comprehending how the whole has complex structures generated in it. We try to analyze reality into simple parts so as to get a firm hold on those, without pausing to adapt ourselves with the complex whole. We try to get hold of truth without caring that every truth has an obverse — truth relates to only some part of a complex and infinitely extended reality.

Life needs success *and* fulfillment, knowledge *and* wisdom, the explicit *and* the implicit, judging *and* understanding, focusing on the poles of a binary *and* looking at the spectrum connecting those. These are no textbook things. These we pick up from folklore, from anecdotes, and from pieces of wisdom coming from our next-door neighbor rather than from the lectures of a university professor. We discover ourselves as much from the lectures of the professor as from the simple anecdotes of the decrepit old man residing next door — none is to be given precedence to the other.

We are keen on discovering the mechanisms of Nature, ignoring our own selves hidden in the world, because the self is not captured by mathematical formulas and statistical data sheets. We are keen on discovering the ultimate, the one single 'law' that will tell us everything, not pausing to think as to whether it will give us knowledge of ourselves. Meanwhile, persecution, torture, murder, ethnic cleansing and genocide goes on, simply because we do not know our own dispositions that won't stay repressed.

5.2.8 The social 'mind', the social 'self', and the social 'soul'

The most remarkable thing about neurons, neuronal assemblies and brains is that they do not operate solely within the confines of our body. Brains operate in collection in social groups and in society at large.

The brain, in association with other physiological systems including the neurotransmitter systems, produce the mind, the mind produces *thoughts and behavior*, and these leads to *social interactions*. The social interactions, in turn, generate sets of complex experience in individuals, which they interpret, generating meaningful thoughts in their own minds. Experience, interpretations, and thoughts, these imply the activation of neurons and neuronal assemblies — dynamical excitation patterns are generated, neuronal pathways get modified in virtue of plasticity, and the mind keeps on producing behavior on the basis of complex excitation patterns.

This is how the circle gets completed and brains interact in society — not, to our knowledge, by emitting and receiving electromagnetic waves, nor by means of mysterious extra-sensory signals, but through social behavior. And this is how social formations and society at large come to have 'minds' of their own. In the following, as elsewhere in this book, when we speak of social groups we mean families, schools, workplaces, farms, business enterprises, welfare organizations, trades unions, cultural organizations, political parties, religious sects and communities, governments, states, and entire societies within countries. We include even mankind at large, but that as a special mention since mankind constitutes a closed formation, without other social groups to interact with.

Every social group has some *aim* that constitutes its identity and its *raison d'être*. As mentioned earlier at several places in this book, one can imagine the individuals within the group as being analogous to neurons in a brain and sub-groups within the social groups as neuronal assemblies.

> We repeat that this analogy does not take anything away from the distinct identities of individuals but only underlines the fact that they communicate and interact between them as the neurons in a brain do.

Communications between the members of a social group occur at all levels. A social group includes within itself numerous sub-groups formed by means of regulations and rules (teachers in a school on the one hand and students in the various grades on the other; officers, clerks and workmen, and the various departments in a business enterprise) or by personal affinities and antipathies (groups of friends, rival groups within a political party). The subgroups communicate with one another over and above the communication that goes on between individuals. As we consider larger and larger social groups, the communications among the subgroups assume greater relevance in the functioning of the group as a whole. In this sense, the sub-groups resemble the neuronal assemblies that, by means of their interactions, generate mental processes and the mind based on a brain. The overall functioning and behavior of a social group, in interaction with groups of various other descriptions, can be imagined to be the produced by 'mental' processes based on the communications between the subgroups.

This analogy between the mind of an individual and the 'mind' of a social group (refer back to section 2.7.7) is a fertile one — more fertile, perhaps, than it at first appears to be. Neither of the two has a material existence, and both are phenomenal entities generating behavior. Their difference lies in the fact that one has a phenomenal perception of the workings of one's mind while the 'mind' of a group is not phenomenally perceived in a similar manner. An infant develops a theory of mind (ToM) early in life in virtue of innately developed capacities, which gets enriched and expanded subsequently in life by means of inference and learning, but the mind of a social group is not understood by means of a theory having such innate origin. The social 'mind' is nevertheless a useful idea since the behavior of the group is mostly not a random or haphazard one, and has a coherence stamped on it that is conveniently described in terms of an underlying consistency that we take as the coherence of the social 'mind', just as in the case of an individual.

One can extend the analogy and imagine 'unconscious' and 'conscious' mental processes in a social group. The former are the ones that are not articulated and expressed with sufficient coherence in the collective thinking of the group — these are mostly based on thoughts and actions of individuals and small subgroups that do not find overt expression within the group, while the latter (the 'conscious' processes in the present analogy) are precisely the ones that depend on the interactions between subgroups (recall that these are analogous to the neuronal assemblies in the brain responsible for mental processes) that find overt expression as 'thought processes' of the group (the circulars, office memos, mails, financial plans, personnel policy and similar other internal and external communications in the case of a business enterprise; explicitly articulated ideas of the group as a whole; the cultural views of a society).

This brings us to processes of integration and regulation in the 'mind' of a social group. Such processes are many and varied. For instance, the elaborate plans for mutual compatibility of the inputs and outputs relating to various workshops within a large factory whereby a fixed volume of its final output can be produced every day is an example of a process of integration. On the other hand, corrective measures are always necessary to prevent disruptions of the functioning of a group when mutual consistency among its subgroups is compromised (strictures and penalties imposed on the employees or legal action against the trade union in the case of a factory; or, say, the rules and regulations formulated for the smooth working of a welfare society) — such processes are of a regulatory nature.

The analogy is further extended so as to include the 'self' of a social group. The 'self' is made up of the preferences, beliefs, and dispositions inherent in a group. For instance, a political party has a disposition to oppose any and every move of rival political formations and to grab representations in the local councils in a community. Or, say, the deeply ingrained belief of a religious sect that the observance of certain rituals would lead to liberation from earthly distress. It is the 'self' of a social group that is mostly responsible for its behavior in relation to other groups and within the society at large, analogous to the self of an individual. What is more, there arise conflicts between the preferences of a group (the inclination of a political party to grab power, and the one to ensure a democratic set-up). Many of the preferences of a group are 'repressed' by various means — for instance, doubts and reservations may be entertained by certain individuals and minority groups within a political party regarding the policy followed by it, but the majority group in power may prohibit the expression of such reservations, which results in simmering discontent within the party.

This, finally, brings us to the 'soul' of a social group. The 'soul' examines the preferences and dispositions of the 'self' by intent, i.e., as a separately accepted obligation, so as to result in a reorientation of these dispositions. In this, the social group institutes processes that stems the tide of emotions sweeping through it, where the 'emotions' of a group are rapidly transmitted but covert (i.e, not systematic or cogent, but implicit response to situations) communications among individuals and among subgroups. Thus, the underground political literature criticizing a dictatorial state machinery is an instance of group-'emotion'. Imagine a cultural organization where a feeling is generated and left unattended for a long time that it has been deviating from the cultivation of healthy cultural values due to propensities and activities of a section of influential members. A day of reckoning arrives when a number of members requisition a general body meeting where all the past mistakes and deviations are to be openly discussed and a new course set for the entire group.

Just as in the case of an individual, a social group keeps on fighting shy of looking at its own self and reconstructing it. Entire societies are known to turn away from self-examination indefinitely, effectively avoiding it for ever. Once again, it is a set of contingent circumstances that conspire to re-align the strengths of opposing propensities and to initiate a process of soul-searching by the social group under consideration. For instance, an imminent stock market crash may send the members of the board of directors of a corporate firm hurrying towards the board-room so as to steer the company away from grossly wrong path of having deceived the public for too long.

Only when a social group reconstructs its ‘self’ to a discernible degree can it be be said to have realized its capacity to exert *free will* in the true sense. As in the case of an individual, this is a phenomenon of rather rare occurrence. On the other hand, what commonly appears to be an act of free will is nothing but self-based activity that differs for different groups facing similar circumstances, and is not explicable in terms of principles or rules of some degree of generality, common across groups — in the ultimate analysis, it is generated causally by factors residing in the developmental history specific to the group and result from the operation of a unique set of intrinsic ‘self’-based ‘rules’.

5.2.9 Power relations: the great obstacle

Power relations pervade our life and our mind. These constitute a vast and convoluted web — sticky and barbed — in which we all are caught like insects in a spider’s net. We lament over how we are helpless prey to power dominating and restraining us, and we gloss over how we ourselves hold others in our own privately flung web of power. We cut ourselves free from this or that particular thread of power, only to get caught in some other treacherous strand. In this infinitely extended and infinitely fine maze of power relations we all are both prey and predator, and it is this dual role of power relations that ultimately stands in the way of the soul restructuring our self.

Power, like belief, is all-pervasive. You engage in a lot of struggle to eliminate one false belief of yours when, unknown to you, some other belief insidiously makes its way to the inner recesses of your mind. Power relations are no different. They hold us — as both predator and prey — in their fold and never let us go: we claw our way out of one pit, only to be sucked into another.

Power relations are both ubiquitous and deceptive. They appear in myriads of forms and are often not recognized as such. Power is often established without explicit awareness of it. It may even assume the form of benevolence or, say, pity. It is only when we are aware of being dominated under power that we are seized by a feeling of being oppressed. Likewise, we may exert power over others without getting drunk by the sense of it. Thus, power relations penetrate every pore of our existence, and we seldom recognize how far they have sent their roots into our mind and our life. Power, in other words, has infinite dimensions to it, and we are seldom aware of more than a few of these dimensions, while the remaining infinitely many dimensions continue to hold us into their grip, sucking into our spirits. Power is the ultimate obstacle to the soul.

We often talk of political power and economic power being exercised by the powerful over the powerless. We are also aware of military power, or even cultural power. But these are only the visible dimensions to power, where it is exercised by few over many. In any case, we are aware of power relations when these assume social proportions. But the infinitely many dimensions of power include those that remain in apparently innocuous guise in relations between smaller groups of men and even in deeply personal ones. Power relations, in other words, may be explicit or implicit — there exists a wide spectrum relating to how implicit a power relation is.

> It is no exaggeration to say that *every* human relation is, at the same time, a power relation.

What, then, *is* a power relation? This is a question to which no satisfactory answer can be given, precisely because power is infinite dimensional and can be deeply implicit. One aspect of a power relation is an *asymmetry*. But then, there can be asymmetry in so many relations in so many ways — the relation between a teacher and a student, for instance. Or, one between a husband and a wife. One can be more specific by stating that a power relation involves a *constraint* on one of the two parties involved in it, as compared to the other. Even this is a vague and ambiguous statement, especially when applied to a relation in which there is no overt awareness of power being exercised.

There have been attempts to define power and even to have an estimate of the extent of power in the context of politics, economics, and military conflict, and many of the considerations there can be extended to power relations involving individuals as well. But no definition can avoid the problem that power is multidimensional. Thus, power may involve coercion, influence, authority, force, and manipulation, each of these being distinct from others [88]. A working definition, consistent to some extent with the considerations in this book is the following: two individuals (or social groups, or an individual and a group), say, A and B, will be said to be in a power relation if a set of preferences of one, say B, are constrained from being fulfilled in virtue of the relation, without there being corresponding constraints on the preferences of A. Admitting the incompleteness of this definition (or of others of a similar nature), we will now make the following important statement: *power is inimical to the soul*.

A power relation mostly involves two parties, each of which may be either an individual or a social group. One of the two is the party wielding power and the other is the one that is subjected to power and constrained by it. We will denote these two as A and B (or the A-party and the B-party) below and will mostly focus on the A-party, unless stated otherwise.

> The B-party is also of great interest in respect of power relations and their dynamic. The B-party, while being subjected to constraints by the power relation, *internalizes* numerous aspects of how domination is exercised and is likely to be found to act as the A-party in its own right in some *other* power relation. At the same time, as and when the power relation between A and B is overturned, the B-party often assumes the role of the A-party in a *succeeding* power relation. As a government is overthrown on charges of corruption, the party in opposition, on forming the next government is, on many occasions, subsequently found to nurture corruption within its own fold. With this understood, we will, for the sake of brevity, focus on the A-party in most of our considerations below.

Power relations affect our empathy system and our social cognition. Moreover, like substance-abuse, power is addictive, since it offers deliverance from the existential burden arising from the multitude of bonds tying an individual to reality and demands on him imposed by the latter. It is very common to find persons wielding power (the A-party in our terminology) to develop strong preferences for various associated aspects to power such as others' attention, subservience, and awe. It is this addictive aspect of power that is most relevant in respect of the operation of the soul. It is commonly assumed that an individual subjugated under an overt power relation such as being a member of the work-force in a large industrial establishment, is likely to be in possession of his own self, i.e., can come out of negative influences with relative ease. This, however, ignores the subtlety and ubiquity of power relations, in virtue of which that same individual is likely to be the dominating party in some other power relation, perhaps a person-to-person one of an implicit nature. Or that same individual (or social group) is under the spell of an aggressive instinct, in virtue of which he intends to inflict damage to some other person or some other social group.

From an *existentialist* point of view, power relations are of foundational relevance to the existence of man, related to such other existential issues as anxiety, boredom, restlessness, self-love, yearning for social company, and aggression where, once again, there are all-round conflicts in these existential urges. Existential issues are, in the ultimate analysis, affect-linked. When deep-rooted preferences and drives of an individual (or of a social group) fail to be fulfilled in the course of his developmental history, his existential anxiety and boredom gets hold of him, causing a 'mental vacuum' when he develops such drives as aggression [50] or a quest for power.

> The fundamental existential disposition of the mind is to keep on generating preferences and concepts where the latter are related to other concepts by the great web of the conceptual network lodged in our mind, which generates reason. If by some circumstances one (or both) of these foundational necessities of the mind gets thwarted, then existential issues come up, mostly in the form of anxiety and boredom. This is the 'mental vacuum' referred to above, leading to restlessness and drive for power and aggression.

What is more, all the existential issues we face are related to consciousness too. The unconscious mind is ever-active and does not suffer from 'mental vacuum', while the conscious mind does. Thus, the operation of the soul brings to the fore an exquisitely convoluted problem: while the conscious mind driven by reason needs to modulate affect generated unconsciously, the existential power relations, associated with the other existential issues, cast a shadow over the conscious mind and generate very strong addictive preferences that act against any meaningful modulation of affect.

We will not enter into further considerations relating to these issues since that is way outside the scope of this little book of ours and will involve too much of speculation. Instead, we close this section and this chapter with just the observation that power relations constitute a great obstacle to the operation of the soul and to the realization of an 'open life' on the part of man. How that obstacle is to be overcome is the great unsolved problem looming before mankind.

Chapter 6

What, if anything, does this book tell us?

Section numbers pertaining to previous chapters (chapter 1 to chapter 5) of the book have been mentioned below at places where these have been deemed to be specifically relevant. Most of the topics in these chapters have been touched upon in multiple sections. In any case, the present chapter is not a section by section summary of the book, and is more a summary of how I look at its spirit — what *my* takeaway from it is. I have made no attempt here to maintain the sequence of the presentation followed in the earlier chapters. In other words, the present chapter is more a condensed re-interpretation than a summary.

The brain, the mind, and the world.

In this book we have been mostly concerned with the *human mind*. The mind operates as an intermediary between the neuronal set-up in the brain and the world we find ourselves in. Each of the three is a *complex* system interacting with the other two. The mind, in particular, is an emergent phenomenon produced by the brain and guides our response to signals of diverse types received from the world. In turn, our experiences of the world, registered in the mind, react back on the brain by means of neuronal *plasticty*. This book tells the story of how the mind is constituted and how it responds to the world.

1. When we speak of the brain as giving rise to mental states and processes we actually mean the brain along with the associated physiological systems (including the ones involving neurotransmitters) that regulate and modulate its activities.
2. In the context of the operation of the mind, the term 'world' includes both the 'external' and 'internal' worlds, where the latter refers to the mind itself since various parts of the mind send signals to various other parts, and activate those — the mind responds to these internally generated signals as it does to signals received from the external world.
3. The mind does not have a material existence but is *real* nevertheless. From the ontological point of view, it can be looked upon as a *property* of the brain and is phenomenologically perceived by us.

The mind is how the brain expresses itself in the world. As we have seen in chapter 4, a property of a material entity can be only notionally isolated from the entity itself since the very definition of the latter remains incomplete unless we specify how it behaves in this world of ours.

In other words, the entire spectrum of behavior patterns of an entity, in association with all other possible entities in Nature, is needed for a *complete* specification of that entity itself — anything else is at most a partial and contextual definition that can get altered drastically as the context changes. From a fundamental point of view, this is because of the pervasive complexity of the world.

This implies that, defining an entity in isolation from the rest of nature is fundamentally an act of violence. However, various parts of nature appear to have an existence of their own in our phenomenal world, with their inalienable links to the rest of the world appearing as their *properties* in our phenomenal perception. This is essentially what we saw in section 4.3. It is with this qualification that we speak of the brain and the mind as being notionally distinct, with the latter appearing as a spectrum of properties of the former. In describing and explaining the behavior patterns of the brain, we often use expressions to the effect that 'the mind does this' or 'the mind does that'.

The fundamental property of the mind is to *perceive* signals from the world, to *interpret* those by integration with earlier experience, and to generate *response* to the signals, the response being geared to a huge set of preferences (and aversions) that the mind generates in the course of its ceaseless interactions in the world. In this sense, the mind operates in a 'purposeful' or a 'goal-directed' manner, its 'purpose' being the satisfaction of the preferences. Preferences and aversions in the mind are generated by the *affect system*.

The goal-directed processes in the mind make it an *active* device (and not merely a passive one) since its 'output' is seldom in proportion to the 'input' it receives — it adds to or subtracts from the latter, depending on what is already stored in it (mostly in the form of preferences, beliefs, and dispositions) in virtue of *past experience*. This active alteration of signals received from the world appears, in the final analysis, as the interpretation of the world that the mind generates. It is the interpretation that constitutes our sole takeaway from reality and acts as the basis of our response to the latter.

The generation of preferences and dispositions in the mind by the affect system relates to one of the two major areas of discourse in this little book of ours.

Complexity of reality. How the mind is structured.

The *reality* in which the mind is embedded and with which it interacts by way of receiving signals of diverse types and generating appropriate responses, is an endlessly *complex* system, owing to which the brain and the mind are enormously complex systems too. The brain has evolved into an amazingly complex system so as to be able to make us survive and thrive within a complex reality. In consequence, there have developed complex *structures* in the brain and in the mind too. The neuro-anatomy, developmental biology, and neuro-psychology of the brain are way beyond the scope of this book. As for the mind, we have briefly outlined some of the structures and processes in it, indicating (a) how it operates in two streams comprised of the unconscious and conscious processes that are analogous in numerous respects but have distinct styles of functioning, and (b) how it accommodates the *self*-system beside the non-self aspects of the phenomenal world, and how the self, with *affect* (section 2.4) as its axis of development, is further constituted of the adaptive-cognitive and the *repressed* (or psychic) components.

The self (section 2.7) is mostly made up of a person's own affectively linked preferences, beliefs, and dispositions, and associated memories.

The complexity of the world and of the mind (section 4.3) is the second of the two major areas of discourse in this book.

As mentioned above, the fundamental job of the mind is to *perceive* the world and to *respond* to it in ways compatible with its goals and purposes.

However, the complexity of reality makes all perception and knowledge necessarily fragmented and incomplete. The mind can access and understand the world only in patches, and that too only by integrating the signals it receives with ones already accumulated from past experience. It is this process of integration that gives *meaning* to the multitude of signals the mind receives incessantly. This means that the only awareness and comprehension of the world that the mind can have is by means of *interpretations* formed in specific *contexts*. In other words, the vast and infinitely complex reality that exists independently of the mind is not comprehended *as it is* and *in totality*, but only by patching up our interpretations. It is in this manner that we assemble bit by little bit the *phenomenal* reality that differs essentially from the *noumenal* or 'real' reality out there. All the takeaway that the mind acquires from the noumenal reality is in the form of its interpretations that, in the ultimate analysis, generate the phenomenal reality. We *decide* and we *infer* by making use of the interpretations that generate our preferences, beliefs and dispositions, and we then go on to make *theories* of reality that we find ourselves in. All these produce anticipations and predictions about the reality and, finally, generate our *behavior* — the latter constitutes our *response* to signals received from the world. This constitutes the foundational position that forms the backdrop to all our considerations in this book.

The perception and understanding of the reality that the mind gains by means of its incessant interpretation of signals received from the world, constitutes *cognition* — a process that can be a conscious or an unconscious one.

In order that perception of and response to reality may be possible the organism housing the mind is to continue its existence, which is ensured by the brain engaged in controlling a host of physiological regulatory functions that make the body survive and thrive — ones that are not commonly counted as operations of the mind. The mind, to be specific, is ceaselessly engaged in *self-linked* and non-self linked processes, each of which occurs within the folds of both the unconscious and conscious layers of it — all in the pursuit of perception and response that constitute its fundamental job. For instance, a host of adaptive-cognitive processes routinely operate as unconscious ones ([147]), illustrating the phenomenon of 'unconscious cognition', and occur mostly outside the ambit of affect. In contrast, the self-linked processes, whether conscious or unconscious, are associated to a greater extent with affect. Finally, processes in the conscious mind — ones we are aware of — are perceived as *thoughts*, regardless of whether these are self-linked or not.

The *self* is the driving engine of the enormously complex and goal-directed activity of the mind and, in a sense, orchestrates the mind as a whole. Mental processes in which the self plays a pivotal role are referred to as *psychological* ones. The self-system (commonly referred to as the 'ontological self', in contrast to the 'subjective self' where the self acts as the knower of the world) organizes the activity of the mind around the axis provided by affect, while the non-self part of the mind operates to a greater extent on the basis of relations between *concepts*. Concepts, indeed, are the constituents of thought that arise in the conscious mind (section 2.3).

The self, incidentally, is generated iteratively in the developmental history of an individual where layers are assembled one above another. In other words, affect-linked experiences at any given point of time lead to the accretion of a fresh layer over the already existing layers that constitute the extant self-system — the initial development of the self system being entirely affective. This has been referred to as the 'self-principle' in this book (section 1.9).

The non-self part of the unconscious mind makes use of innately perceived regularities in perceptual inputs and operates on the basis of *proto-concepts* that have affective links on the one hand and are based on strictly limited relations between percepts on the other. The non-self part of the conscious mind, in contrast, operates on the basis of concepts that capture relations between percepts through a widely flung network, though the affective cores of the concepts still remain meaningful.

Representations and excitation patterns: ingredients of the mind.

As the brain receives the myriads of signals of diverse types from the inner and outer worlds, it generates *representations* of these signals in the form of spatio-temporal excitation patterns in neuronal assemblies (section 1.8), ones that are then *integrated* by means of interactions between these assemblies, and *interpreted* in terms of past experience (once again, gained primarily in the form of signals) by being associated with patterns generated in the past, so as to form the various *ingredients* of the mind, the latter being self-linked and non-self linked or *shared* ones (shared in the sense of being common to members of some social group or other to which an individual may belong). Among the ingredients, we have focused in this book on preferences, personal beliefs and, more generally, the self-linked *dispositions* of the mind (including the *repressed* ones) on the one hand, and concepts, shared beliefs, and *reason* on the other, where these latter constitute, in general terms, the ingredients of the conscious mind, including its non-self part. These ingredients are mostly lodged in the various memory systems where, additionally, diverse events experienced in the course of life are stored for later recall, mostly colored with affect and emotions. The non-self part of the unconscious mind is formed of proto-concepts having limited relational associations and relatively weak affective links, and also of implicit memories of various kinds.

1. *Beliefs* lodged in the unconscious and conscious layers of the mind constitute the fundamental means by which an individual (or a social group) navigates through life. This is made possible by correlations between proto-concepts and concepts set up by beliefs acting as links. Beliefs, moreover, are often tied with affect and emotions and may be self-linked to a considerable extent.
2. The shared beliefs span a wide spectrum from the point of view of justification and credibility. At one end of the spectrum are to be found the huge store of items of knowledge, made up of justified and true beliefs. However, the knowledge base of a person, along with the store of heuristics that have proved themselves useful and justified in the past, have a self-linked aspect to them in the manner these are acquired, made use of, and integrated with the rest of the self-linked ingredients of the mind. This apparently contrary fact of a set of ingredients being affect-linked (recall that the self system of an individual is assembled around the axis provided by affect) in spite of being justified against the yardstick of widely shared beliefs explains why many of our inductive inferences pertaining to the physical world turn out to be valid and more or less accurate.

The structures in the mind and the 'ingredients' (or 'resources') thereof are outlined in chapter 2

The response of the mind to the world is founded on *interpretations*, *decisions* and *inferences*, based on which the mind produces *behavior* of diverse types. The behavior of an individual then leads to *social interactions* generating, in turn, experience of people belonging to social groups, thereby completing the circle: experience – interpretation – ingredients of the mind – decisions and inferences – thoughts and behavior – social interactions – further experience – experience-dependent modifications in the brain. It is in this way that the brain and the mind of an individual get connected and related with the brains and minds of other people belonging to various social groups (refer to section 2.9).

Affect and reason: the two modes of 'logic of the mind'.

The response of the mind to the world occurs fundamentally in two modes, based respectively on *affect* and *reason*. Affect is rooted in experience — it is experience that generates the preferences, personal beliefs and other self-linked dispositions in the mind. As such, affect has only an indirect link with how the world around us is organized, and it is precisely this latter aspect — complementary in nature to what affect recognizes — that provides the foundation of reason.

Reason (section 2.3.3) encapsulates the relation of implication (among several other types such as the one of causation) among concepts, where the latter result from proto-concepts in a process of objectification — entities appear as objects of contemplation in our mind as their mutual relations in the phenomenal world become salient, over and above their affective meaning. In a manner of speaking, reason is also generated by experience — the shared experience of individuals organized in various social groups. It is shared experience that makes the mind cognizant of the mutual relations among entities in the world, beyond the affect that they generate in our minds. In brief, *affect attaches personal meaning to experience while reason attaches shared meaning.* Among the two, the latter is one that is the more 'objective' in nature, i.e., in greater conformity with correlations between entities existing in the world.

Affective meaning stands for the way the preference-linked valuation and emotional color of an experience is integrated with diverse other perceptions so as to relate to the overall context in which an individual is situated.

Affect and emotions (commonly referred to as the 'affect system') serves the important purpose of generating a classification of the entities of the world (in *implicit* terms) so that the classification is retained in memory and comes in handy in making sense of and responding to signals registered in the mind in subsequent experience. In other words, the affect system assigns an order of *salience* to diverse sequences of disparate experience gained in the world, without which the latter would remain fundamentally *incomprehensible*. In order to classify and assign an order of salience to entities and events, the affect system *amplifies* certain inputs to the mind compared to others, which is always done on the basis of past experiences and future anticipations, the latter generated by the former. This amplifying action of the affect system generates an abundance of local instabilities in the processes of the mind that can be compared to instabilities in the *phase space* of a complex dynamical system. As in the case of local instabilities in a dynamical system, these lead to transitions between a succession of stable *modes* of thought and behavior of an individual (sections 1.16, 4.2.1).

However, if it were only affect that generates our preferences, creates the order of salience among experienced entities of the world, and takes up the job of adapting our existence to the complex reality out there, all its efforts would be woefully inadequate. Adaptivity requires reason as well — that second mode of logic of the mind. Conversely, reason and consciousness would fail to comprehend reality all by themselves. The very complexity of reality dictates that the mind confronts it with a blend of affect and reason — one made up of the *implicit* and the *explicit* modes of perception and comprehension. No part of reality, no experience in our life can ever be described in finite and explicit terms precisely because of the fact that infinitely many aspects of reality relevant to the description remain necessarily beyond the scope of such a description. Affect (and its associated capacity of making analogies) makes up for the deficiency to some extent by supplying an implicitly defined qualitative feeling that is specific to an individual — empathy and affect-based communication between individuals attempts to transmit such feelings from one person to another.

Reason: the spectrum of varying generality.

Reason, contrary to common perception, is not absolute or universal, but can be graded, with varying degrees of generality. At one end of the spectrum of generality we find mathematical reasoning which, to all intents and purpose, can be treated as universal, while reason of intermediate generality gets involved when an individual makes use of beliefs acquired in virtue of membership in some social group or other. The reasoning used in the natural sciences comes close to being general and universal, but often deviates from it to some extent, where informed *guessing* has to be resorted to — that guess, in certain cases, may even be traced back to personally held beliefs.

Affect and reason: uniqueness and generality. The irrational and the rational.

Affect and reason, in other words, are the two fundamental modes of 'logic' that the mind resorts to in formulating its response to the world, with affect tied to the unique developmental history of an individual and reason of various degrees of generality to experience shared with various social groups to which the individual may belong (or, which she may have knowledge of). The 'logic' based on affect therefore seems 'irrational' when seen in the perspective of reason, the reason based on mathematical logic being perceived as the most rational of all. This is why the psychological tests of rationality — a trend initiated in the nineteen sixties to see how normative the thought processes of people were when confronted with various tasks requiring inferential ability — were by and large seen to be indicative of a low degree of rationality of individuals from various walks of life since the norms of rationality invoked in such tests were close to being ones of the mathematical nature.

If this book has one message to impart, it is that the self is based on affect, and *is unique to an individual*, while reason or shared views make up a chunk of our conscious mind, partly distinct from the self. Our journey in life is a strange interaction between the two — affect and reason — where the two are blended together in such an intimate manner that their distinction can be only a notional one. The interaction between affect and reason assumes an ultimate form in the operation of the *soul* (section 5.2).

Decisions and inferences.

Beliefs of various shades (including ones of universal validity) are employed as *rules of inference* when we engage in the making of decisions and inferences. Among the latter are the deductive and inductive inferences (section 3.2.1), the two being notionally of distinct types (just as reason and affect are), while being blended together beyond recognition in actual acts of inference. Most inferences in real life are of the predominantly inductive type (mathematical derivations and many of those made in the natural sciences excepted) where there is an essential element of choice and decision-making involved at intermediate stages of an inference when the mind makes use of beliefs of even low degrees of generality in making guesses while exploring the conceptual space. In other words, affect-driven and self-linked ingredients of the mind get mixed with reason and shared ingredients at various stages of inference and decision-making. Among the former, a huge store of *heuristics* — often ones of substantial credibility — lodged in the mind of an individual is invoked in inferences relating to the physical reality, while personal beliefs and dispositions are mostly used in social inference.

The making of inferences involves an exploration of the 'conceptual space' in which myriads of concepts are linked with one another in a widely flung network, where the 'exploration' consists of establishing new and meaningful correlations among unconnected concepts or groups of concepts. The process of exploration occurs sequentially by means of reason, punctuated with 'logical leaps' (section 3.2.2.5) where self-linked rules of various types are made use of. In this, analogy often plays a seminal role.

Exploration of the conceptual space: the powerful tool of analogy.

It is in the process of exploration of the conceptual space by means of informed guessing ('logical leaps') that *analogy* (sections 3.2.2.8, 4.5.7) makes its worth felt, proving itself to be of supreme value in an act of (scientific) *creativity* that can be described as a glorious act of inference-making. Generally speaking, inference-making involves connections being sequentially set up between closely related and contiguous concepts in the conceptual space (by means of either shared or self-linked 'rules') in a process that can be likened to the exploration of a landscape with the aim of discovering, say, a certain landmark, that can be compared to the intended conclusion in an inferential act. In this, one often also needs to establish a connection between relatively remote concepts in the conceptual space in an act of inference where the application of rules, whether of general or of restricted validity, does not prove effective. In this context, analogy constitutes a type of 'reasoning' that is at once self-linked and sensitive to subtle and delicate relations between concepts — relations that are exposed and acquire significance by means of affect. In other words, analogy exposes relations between concepts with reference to the huge canvass of the developmental history of an individual. When expressed in conceptual terms, the developmental histories of two individuals can be compared with each other, but the developmental history of either individual involves affect-based associations that cannot be compared in terms of common referents.

This is why analogies establish relations between concepts but are peculiarly elusive — establishing a genuinely eloquent analogy is a deeply personal achievement on the part of an individual that can be appreciated by another person without, however, the latter having been able to 'discover' it. And this is also why analogies establish relations between concepts *remote* from one another in the conceptual space. Here the term 'remote' signifies 'sparsely linked', and retains relevance for concepts within some given conceptual domain as well as ones belonging to distinct domains with only few connections linking them. Invoking analogies within the confines of some given domain leads to insightful solutions to problems encountered more or less frequently in life. Analogies establishing correlations between distant conceptual domains, on the other hand, leads to episodes of creativity (section 4.5) where a single analogy may unearth a multitude of correlations that light up entire terrains previously unexplored. This is analogous to critical phenomena in phase transitions in a physical system where constituents of the system get correlated at all length scales.

The mind confronts the world.

The mind, in a sense, *confronts* reality while being itself a part of reality at large. It is important that we realize what this confrontation consists of. The mind is housed in a body — a biological organism that has needs of its own. The mind incessantly receives signals of diverse types from the world around it, *and* from various parts of the organism, including signals from parts of itself. Its job is to respond to these signals and to help satisfy the 'needs' of the organism, in accordance with its *interpretations* of those. All the response of the mind is based on interpretations of the reality it faces, and all the interpretation is done by the mind with reference to *past experience*. Whatever the mind feels as having been 'useful' and 'congenial' in past experience, it retains in the form of beliefs, preferences, and dispositions, to be invoked subsequently in its response to the world. Preferences and dispositions, indeed, are generated in a strange and complex process (section 5.2.2). The only adequate description of this process is that it proceeds iteratively, *by association* with the already generated preferences (analogous to the development of the self-system). Thus, remarkably, there exists no superior principle regulating the generation of preferences and dispositions. *This explains why there exists pervasive conflicts within the preferences of the mind.*

> The response of the mind is not of a simple or 'mechanical' type and includes a generous dose of *anticipating* and *planning ahead*. Based on past experience, the mind incessantly anticipates by making guesses, and constantly generates 'error messages' (based, to a large extent, on the affect system) encoding the deviation of the plan from currently received signals. These error messages help the mind to plan ahead, which is essential if it is to do its job in an uncertain world around it.

The world is fundamentally a complex one. All the information that the mind can have about it is partial, fragmentary, and woefully paltry for getting the needs of an organism satisfied adequately. Faced with this, the mind interprets the world and fashions its response in accordance with that interpretation. In other words, the mind resides in an interpreted world — the phenomenal reality generated from the noumenal one (section 1.12) that is too vast, too complex, and too extended to be comprehended *as it is*, in totality.

Social reality and physical reality.

The phenomenal reality that the mind finds itself in is made up of the social reality (including the reality of the *self*) *and* the physical reality — both are far too complex for the mind to respond to without making up *models* born out of an effort at interpretation. The models may be made up either explicitly or implicitly, and inferences are made within the limits set by the models. Inference and decisions are essential in order that responses be formulated with any degree of efficacy. And, in order to infer and to decide, 'rules' of inference are to be invoked. The huge repertoire of preferences, beliefs, heuristics, and dispositions come in useful here, playing the role of *if-then* type rules. The vast set of rules that the mind invokes, including the rules dictated by *reason*, is highly inhomogeneous, fuzzy, and *inconsistent*, and is often given coherence by means of *emotional* ties.

As mentioned repeatedly in earlier chapters of this book, the 'logic' that the mind makes use of in the form of rules of inference has two distinct aspects to it — the logic of affect and the logic of reason (section 3.3). The former is predominantly operative in *social cognition*, while the latter is needed in a greater measure in the *cognition of physical reality*. All the while, however, the two brands of logic operate hand in hand, inextricably and inexorably blended with each other. Their mutual entanglement renders possible the making of *inductive inference* in social cognition and in the cognition of physical reality.

The affect-laden social cognition is fundamentally judgmental and value-ridden, involves the imputation of hidden motives to people as a general principle, and is conditioned by prior preferences and aversions to them — it abounds in adulations and resentments since it is vastly underdetermined by evidence. This is because people are *in fact* driven in their social behavior by drives, dispositions, and motives, much of which they themselves are unaware of. In comprehending and responding to social reality, the mind of an individual relies on the *theory of mind* that it develops from days right after birth and keeps on reconstructing throughout life.

As for the cognition of the physical reality, the mind makes overt use of reason while affect keeps on operating subtly and covertly. It is of vital importance that the mind forms a reasonably accurate picture of the physical reality in order that the organism — and the entire society — may survive and thrive. But once again, the very complexity of reality stands in the way, and makes it imperative that interpretations be formed, with a more or less heavy dose of compromise between the accuracy and efficacy on the one hand, and a sense of satisfaction with *theories* on the other.

Theories in the sciences are born out of the necessity of explaining evidence and are more or less limited in scope, depending on the *models* (section 4.3.5) these are applicable to. Nature is fundamentally incomprehensible in virtue of its infinitude and complexity. In its continuing attempt at comprehending nature, the mind of an individual and that of society as a whole makes use of models and of interpretations in the form of theories formulated within the limits of these models. Theories do not apply to nature as a whole but to the models, where a model refers to some chunk scooped out of reality, with the effect of the rest of reality (and of its past history) represented approximately and provisionally in the form of initial- and boundary conditions — ones that may or may not be formulated explicitly and precisely.

Theories act as *codes* (section 4.3.2.3) with reference to the models, and are to be unpacked, with underlying acknowledgment of the initial and the boundary conditions — only then do these result in *predictions* that can be used to respond to and to *act on* nature. Moreover, theories have to conform to evidence — a requirement that demands great care in the construction of models. On the one hand, a model has to be sufficiently *simple* so as to be amenable to an approximate solution as regards its future behavior and, on the other, it is to represent with sufficient accuracy real-life systems under realizable conditions of observation. Given these two requirements, theories have proved themselves to be remarkably efficacious in predicting the behavior of systems, at times almost miraculously so.

1. Models are, essentially, relatively *simple* systems scooped out of an enormously complex reality, as a result of which *deductive reasoning*, based on rules of relatively large degree of credibility can be employed to understand those, while logical leaps based on beliefs of a lesser degree of credibility continue to be

relevant as well.

2. Theories can be looked upon as chunks of *information* that provide us with condensed description of parts of nature perceived in our experience (see [141] for an interesting point of view).

In this context, quantum electrodynamics and the standard model can both be counted as theories applying to 'simple' systems since these represent only a few fields (basic entities in the quantum theory at very high energies) in interaction, with precisely formulated rules of calculation leading to results that can be tested in the laboratory by means of set-ups where there is little interference from extraneous systems. Compared to these, the theory describing a solid is beset with huge problems, as a result of which there is no overarching theory of solids — the behavior of a solid, as discovered by probes of diverse descriptions can be constructed only in patches, where the spectral characteristics in the various frequency regions are worked out in terms of effective Hamiltonians involving 'quasi-particles' independently of one another, even when the total Hamiltonian of the solid as a whole can be written down in principle (and that too approximately).

In any case, one does not even hope to work out the theory of a solid solely in terms of the standard model, precisely because the solid is much too complex a system for that. Results in the standard model are tested not in the context of a solid or a liquid as found in real life, nor for a living organism, but for two or at most three particles colliding at a time in a highly evacuated enclosure in the presence of strong accelerating fields — such systems are 'simple' indeed, though requiring great talent in deciphering them, and a fabulous investment too in a world ridden with poverty and distress. Disregarding the aspects of distress and poverty, as one tries to go to higher collision energies at exponentially escalating costs, the 'system' becomes more and more complex as strong interactions and gravitation become relevant for theoretical predictions to work, and known theoretical schemes, instead of yielding convergent results, become divergent and useless.

Decisions and inferences in the description and explanation of experience often involve the making of informed guesses (the hallmark of inductive inference), since our experience in real life is, generally speaking, not amenable to explanation solely by means of shared beliefs or rules based on reason. In the case of inferences in the social domain, these guesses are mostly based on self-linked beliefs and preferences of an individual, while in the case of scientific inference and decision-making in the explanation of physical reality, where self-linked rules of a higher degree of credibility are needed, it is the vast repertoire of heuristics (justified to some extent in earlier stages of inference making) stored in the mind that is commonly invoked, in addition to the knowledge base stored in the mind. Heuristics are rules of thumb, are often generated in a hierarchical manner in processes of inference themselves, where successes and failures in the inferential acts generate self-linked rules ('such-and-such is or is not a good way to proceed in such-and-such situations'). The linkages to self occurs through the action of affect, since the mind always finds success pleasurable and failure aversive.

What is most remarkable is how affect, generated in past acts of inference, plays the role of an 'internal monitor' (or of internal 'censor') steering our inferential processes (section 3.2.2.3), where it keeps on generating error messages telling us whether we are on the right track in an inferential act or diverging away from it. Without this silent guidance from the affect system, no act of decision and no process of inference in life could ever be carried to successful conclusion. This is where affect and reason, in a manner of speaking, hold hands in an intriguing partnership that can only evoke amazement.

The mind as a complex system.

We have spoken of decomposability of complex systems (section 4.2.1), whereby the constituents of a complex system can be said to enjoy a measure of individuality and autonomy. However, decomposability can vary widely in gradation (section 4.2.1). If the constituents of a system interact weakly among themselves, where the degree of strength or weakness of interaction is to be defined contextually, then the decomposability becomes manifest, while complexity makes itself felt mostly through the effect of what has been referred to earlier as 'remote causes'. If, on the other hand, the interactions are sufficiently pronounced, then the decomposability of the constituents also gets compromised and complexity assumes even more daunting proportions.

The mind, in several respects, is a complex system of the latter kind, where it is a more or less tightly organized system of interacting neurons, and communications among neuronal assemblies makes it problematic to explain mental processes in terms of activities of decomposable units. While this is not a precise statement, it gains a measure of significance when our attempts at deciphering the mind of an individual is compared with those relating to the description and explanation of physical reality, since major chunks of physical reality (especially those included in numerous studies in physics and chemistry) belong to the weakly interacting category and are, in a large measure, decomposable which, precisely, is why such systems can be studied with reference to idealized and simple models.

In contrast, simplified models of the mind are farther from reality and, more often than not, theories of mind prove to be blatantly inadequate. Such theories are heavily underdetermined by evidence and we are, in the course of our social interactions and interactions with ourselves, frequently forced to renew them, even to the extent of discovering entirely novel aspects of a person even under apparently small changes in context. The newly emerging theories, on the other hand, remain almost equally speculative, and our picture of the mind keeps on changing kaleidoscopically in relatively rapid succession. The time scale in which the mind keeps on evolving through stages of self-organized complexity is comparable to typical time scale relevant in social interactions.

Starting with physics and chemistry, one can go on along a progressively decreasing scale of decomposability — not a strictly defined one — as we engage with geology, ecology, climate studies, weather science, public administration, economics, politics, and so on, so as to finally end up with the science of living organisms, social psychology, and the psychology of individuals. The mind, in other words, is way more inscrutable than what is generally revealed of the physical reality in our studies in the physical sciences.

Reality is one vast and complex system, in which all other systems are included. The above considerations only tell us that all the subsystems of the widely flung reality do not interact equally and homogeneously among themselves, and there is a wide variability of decomposability across various cross-sections of reality. This is a feature that applies, in a sense, to all complex systems since the latter routinely involve several *types* of interactions among subsystems. Even the mind includes chunks where the interactions among subsystems is relatively weak — for instance, those generating unconscious processes — though we do not know if some of these same subsystems are *also* involved in strong interactions with *other* subsystems in it.

Is the mind ruled by random factors?

The mind easily gives the impression of being a system ruled by random processes, leading to similarly random patterns of behavior that it generates. For instance, most of our *thoughts* (processes in the conscious mind that we can be aware of) appear not to follow any set pattern and seem to go this way or that with few, if any, causal factors in sight. More precisely, the mind appears to follow a strange combination of discernible patterns (*getting tensed up when humiliated*, or *feeling elated at solving a difficult problem that defies my peers*) and ramblings of a random nature.

All complex systems have random features in their behavior patterns, symptomatic of the fact that no such system can be known completely and precisely, in virtue of which their behavior is governed by a great many unknown and even counter-intuitive factors. In this book, we have not spoken much of the randomness that can be discerned in the workings of the mind, mostly because considerations of randomness lie, in the main, beyond our concerns in the present discourse. All the same, randomness in the workings of the mind continues to be a relevant feature though its roots and its exact nature require special considerations before their role in human psychology and human behavior can be adequately understood. A notable exception in this regard is constituted by the exercise of free will. We have spoken briefly of the apparent randomness in the exercise of choice among alternatives in an act of inductive inference and also of an apparent randomness in the exercise of free will. However, within the limits of what this book considers to be relevant, we have identified the role of *self-linked* *psychological resources* as being of greater significance in understanding a number of aspects of inductive inference and of the exercise of free will. We have also implied that the randomness of mental processes is not something that is antithetical to the pervasive operation of *causal* factors whose roots lie in remote terrains within the folds of complexity.

The only relevance of randomness pertinent to our considerations in the present book concerns the occurrence of events of *self-examination* that lead to a restructuring of the self of an individual, while similar considerations apply to the possibility of self-examination by social groups as well. Such self-examination takes place when there is a favorable conjunction (sections 5.1.8.1, 5.2.6) between the social context and the affect-based configuration of the mind so that the soul can evade the bootstrapping effect in confronting the self whose deeply held beliefs and prejudices are thereby opened up and are accessed by the conscious mind guided by reason. The reasoning capacity of the conscious mind in this case commonly takes the form of making use of doxastic and deontic reasoning associated with ethics, morality, conscience, and philosophical principles.

More generally, randomly operating factors abound in the activities of the brain and human mind, starting right from the level of the individual neuron (section 4.3.11.2; in this context, refer also to sections 4.2.1, 4.2.1) up to and including the complex ebb and flow of emotions and the meandering thought process in the conscious mind. The significance of these random processes with reference to causality, determinism, and predictability in the context of the human mind constitutes a subject beyond the limits of the present book.

The human mind is truly inscrutable and, from a computational point of view, intractable too. Of course, in keeping with the general characteristics of complex systems, there are islands of simplicity and regularity strewn here and there — these constitute the terrain in which our theories of the mind are built up and applied, being important tools in our attempts to understand human behavior. It is mostly with reference to this terrain that the considerations of the present book acquire relevance.

Theory revision.

As a theory pertaining to a model of sufficient simplicity is compared with available evidence, one finds large areas of agreement and, *at the same time*, anomalies that remain out of sight to start with, these anomalies being tell-tale signs of the vast span of complexity that has been left out of consideration in the setting up of the model. However, as the testing of the theory is made more and more demanding, some of the anomalies assume painfully glaring proportions, and the existing theory proves to be inadequate to match their refusal to get explained. This is when the hunt for a new theory becomes imperative. That hunt takes the form of a renewed exploration of the conceptual space over an expanded terrain in that space including remote domains, which is where the capacity of the mind to invoke *analogies* makes itself felt. The beginnings of a new theory (section 4.3.8.3) become visible in the form of novel *hypotheses* that now appear on the horizon, with new layers of correlations connecting remote concepts over and above the layers that are retained (perhaps with modifications) in the multi-layered conceptual network. The revised theory that now gets formulated with progressively greater precision stands in a strange relation with the earlier one, a relation of *incommensurability* (section 4.4), where the two theories have common referents but are still on different grounds as regards their basic concepts — the earlier theory can be understood from the point of view of the later but the converse claim does not hold.

Theory revision involves a kaleidoscopic change of the world view — there can be no ultimate theory of Nature since every view of it is contextual. We employ reason to comprehend what we find of nature within given contexts — in which affect plays a covert but crucial role in the making of decisions and inferences — and in creating new hypotheses but, at the end of the day, reality 'in itself' remains as elusive as ever. Beyond every view and every theory of reality, there remains something underneath.

Man in isolation and in society is just as inscrutable, and an ultimate theory of man and society is just as unattainable. We are bound to ourselves and to society by ties of affect and we comprehend both by means of affect, aided with reason. In this, our efforts are of a complementary nature to how we try to comprehend physical reality.

In both, we attain only a fleeting, contextual and fragile understanding but one that is useful nevertheless. Human reality and physical reality are complex entities both, and in their course of development by means of self-organized complexity, offer us islands of stability and comprehensibility, where we find regularities that we grasp and make use of.

In brief, in our journey in life, we need to comprehend and to engage with the physical reality, with the human society around us, and *also with ourselves*. All three are exquisitely complex systems, and we need a great deal of *interpretation* and *theory* in this. The theories we invoke are context dependent and are underdetermined by evidence, and as the context changes, our theories need changes that are radical and spectacular. All three are fundamentally inaccessible in their totality — one can only have mutually incommensurate views from various perspectives and one has to patch together all these views as we engage in the act of living our lives.

The human mind: conflicts and their regulation. The soul.

We need to understand Man and Nature because both are full of contrary aspects that make for a great deal of uncertainty. What is most contrary is the human mind itself where conflicts between preferences are pervasive, and so are the ebb and tide of emotions making necessary a vast repertoire of integrative and regulatory processes in the mind (section 5.2.3). In this, the ultimate enigma to understand and to make good use of is the *self*. The self is as inscrutable as our fellowmen in society and the physical reality around us. Analogous to our theories of the physical and social realities, we all have an implicitly assembled theory of self, based on our *self-image*. However, this theory is the most underdetermined of all and yet the most intransigent against revision because it is guarded by our deepest beliefs and emotions assembled around the axis provided by shame and pride.

We have our consciousness to look at ourselves but once again our grasp of the self is fleeting and fragile — there is no finality in our perceptions and interpretations: not in Nature, not in our fellowmen, and not even within our own selves. Still, we try to regulate the turmoil within us, we fail, and we try again. The worst thing that can happen is when we don't grab at the fleeting chance offered us from time to time to look at ourselves. That chance is the result of a favorable conjunction between two exquisitely complex systems each undergoing an unpredictable course of temporal evolution — the good fortune is that such conjunctions do happen and are not rare.

When the soul, that highest court of settlement of conflicts within us, remains inoperative, our theory of self continues to rule without ever being revised, and everything else fails. As we continue to be in denial, we terribly misjudge our fellowmen, and we develop foundational theories of Nature, only to be proven wrong in our pursuit of a presumed all-embracing coherence ruling reality at large. We reconstruct Man and Nature in accordance with our preferred image. Our pursuits in science assume grotesque proportions, and our social tenets set entire nations aflame — the rot that resides within us comes out and spreads to engulf the world.

Social groups and social formations.

Every individual is at the same time a member of social groups, one embedded within another. Social formations of various sizes, descriptions, and functionalities form a web of relations mediating between an individual and mankind at large. No disposition, no reasoning, no mode of thought of an individual is universally human since all these are conditioned by attitudes an dispositions specific to social formations to which that individual belongs. Generally speaking, the mind is modulated by cultures since experience is always generated within cultures, and the basis on which the mind develops and evolves is that provided by experience.

In a deep-running analogy (section 5.2.8), a social formation has a 'mind' of its own as does an individual, where the social mind is created by means of interactions among individuals, especially among dominant individuals and dominant subgroups. As an extension of this analogy, a social formation generates in its 'mind', 'unconscious' and 'conscious' processes, where the latter are based on the *articulated* and explicitly communicated interactions among these dominant subgroups and the former on ones that remain implicit, muted, and localized. Part of the social mind is adaptive-cognitive, geared towards its continued existence and thriving, and part is affect-based, where the affect of a social group refers to its set of preferences, implicit beliefs, and dispositions that ultimately constitute its 'self' (referred to in this book as 'group-self'). The adaptive-cognitive part looks at the world in more or less clear relational terms, focusing on how the entities of the world are correlated with one another, while the affect-based part appraises the world in terms of its own preferences.

> The group-self is to be distinguished from the group-induced self, where the latter refers to the preferences and dispositions of an individual induced in virtue of his membership in a social formation or group, generated by the preferences and dispositions of the group. Conspicuous instances of group induced self are the set of dispositions generated by a family and those generated by a dominant culture in which an individual remains immersed.

There invariably arises conflicts among the preferences and dispositions of a social formation, and between these preferences and the requirements imposed by the environment in which it operates. Some of these conflicts are resolved by integrative and regulatory processes running within the formation, while some get repressed so as to constitute its 'repressed self'. Finally, the social formation in question possesses a 'soul' of its own, based on its capacity to examine its own self — especially its own repressed dispositions — and to restructure the latter.

It is on the soul of the individual, the souls of social formations, and the soul of mankind as a whole, that our fate hangs.

The continuity underlying binaries in the appraisal of reality.

If there is one single theme running through this entire book, it is the one of binaries being the starting point rather than the end point of an exploration or a discourse, especially one relating to the mind or to reality that the mind confronts. This is a bit of an abstract statement, but becomes concrete when we consider binaries such as the unconscious and the conscious, affect and reason, the mind of an individual and the 'mind' of a group, inductive and deductive, process-one and process-two, self and non-self, and self-linked and shared resources of the mind.

Looking at a binary, things at first appear to be concrete and fine — in any specific problem, one just has to decide as to which pole of the binary 'dominates'. This does work to some extent in real life, only to show small gaps and loopholes that widen later when one finds that the other pole that was ignored earlier makes itself felt with a vengeance.

This is what comes of focusing on a relatively simple part of complex reality since, in focusing on the simple part (a 'model') we throw complexity to the wind, and complexity is not something that can be thrown away or trifled with. The 'truth' that appears to be so convincing in the context of a model loses its appeal as we look beyond the model. The poles of a binary dissolve into a continuum — reason and affect, for instance, cease to be as contrary to each other as one thinks to be the case at first glance. Everything that appears to be simple to start with reveals a substratum spread underneath that makes our theories messy and our life burdensome. But one has to bear the burden that life forces upon us, and has to open one's eyes to what complexity ordains.

The two poles of a binary are spanned by a spectrum that connect them continuously. This does not make the poles superfluous. On the contrary, it is the dichotomy of the two poles that *defines* the spectrum and reveals its scope. At the same time, it is the spectrum that makes more explicit what is implicitly contained in the statement 'unity of opposites'. Affect and reason, for instance, are opposites that cannot be forced apart, because the two are united — a unity that is at once deep, enigmatic, and mind-boggling because it encodes within itself the entire complexity of how the human mind works.

Power relations and the soul.

Of great significance is how *power relations* in society constitute a durable obstacle to the activity of the soul. Power is ubiquitous in human relations, and it is no exaggeration to say that every human relation is, at the same time, an overt or covert power relation in some form or other. Power relations can pervade our existence in innumerable subtle forms, where a person or a social group can be simultaneously a party exerting power over others in a power relation and one at the receiving end in some others.

Power generates affect of an addictive kind, delivering a transcendence from the existential burden resulting from the innumerable bonds that inexorably tie an individual to reality. This inactivates the soul and creates a strong obstacle against the conscious mind confronting and examining the self, in virtue of which preferences and beliefs inimical to social harmony continue to hold the psyche in a death grip, and all our attempts at social cognition and the cognition of Nature at large become abysmally misdirected. Power always generates a narrative that tends to perpetuate its stranglehold over liberating influences in human relations.

The revision of world view.

Adaptability to the world requires that we constantly keep on revising our preferences and beliefs, form new beliefs, and never stop checking all our beliefs against new evidence. But our response to the world around us is not homogeneous — it is built up in stages. We are given a set of capacities of response by the process of evolution, and we build upon those capacities by means of social interactions. Experience generates processes that lead to the realization of our latent capacities. We start with affect and end up with consciousness and reason, where affect and reason make a deep tangle. We make use of affect and reason in trying to understand the physical reality, the social reality, and the self, and to acquire the ability to *predict* by invoking our *theories* — useful things, but fragile too.

As things stand, the highest level of adaptation relates to the *soul*, where consciousness and reason restructure the self. Restructuring the self is not an automatic process, nor is it a one-off process that depends entirely on the exercise of reason. Whether and how it occurs depends on the will, on circumstances, and on a concurrence of the two. Its course is unpredictable too. At times it gets halted as soon as begun, and at other times it transforms individuals and societies almost inside out. It changes the *world view* —

it causes a re-examination of deeply held beliefs, desires, and drives, but never reaches a culmination. As the world gets transformed, new journeys are begun, and new quests are set upon for novel ideas and novel theories, based on a re-constituted metaphysics. It is just possible that we may then shift our focus from 'all-embracing simplicity and harmony' to 'all-embracing complexity, with islands of harmony'.

> One deeply held belief of ours is that the world is explained in terms of binaries. That people, for instance, are either 'good' or 'bad'; or that, if they are good *and* bad, they are either 'on the whole good' or 'on the whole bad'. That an act of a person is either rational or irrational. Or, say, that the world is 'on the whole' governed by simple, beautiful, and harmonious laws. As for these deeply held beliefs, it does not matter much that the idea of 'on the whole' is not a sound one when invoked in the context of a complex system.

Looking at ourselves, our deeply held beliefs and prejudices, is the key to access freedom in the exercise of our will. As we have seen (section 5.1), what is commonly referred to as the paradox of determinism in the context of *free will*, is really a *loss of predictability* — as viewed in terms of shared resources of the mind — owing to the determining role of self-linked ones, while *causality* remains operative all along. It is solely through an act of self-examination that one can, in a real sense, *choose how to act*, though only at a *later* point of time.

We exist in an infinitely complex world and our journey in it meets with infinite hurdles. We build theories to negotiate those hurdles, we succeed, and then we fail. The journey starts afresh, with fresh ideas, fresh theories and — if we are worth our salt — with a fresh world-view. This time around, there will be fresh achievements, fresh anomalies against newly accumulated evidence, fresh conflicts between our preferences, and a fresh scope of application of regulatory processes of the mind. And the highest level of regulation, once again, will be the soul — the soul of the individual, the soul of social groups, and the *soul of humankind*.

Bibliography

[1] Jonathan E. Adler and Lance J. Rips, (editors), *Studies of Human Inference and Its Foundations*, Cambridge University Press, Cambridge, (2008).

[2] Ralph Adolphs and David J. Anderson, *The Neuroscience of Emotion: A New Synthesis*, Princeton University Press, Princeton (2018).

[3] Antonio Alcaro, Stefano Carta, and Jaak Panksepp, 'The Affective Core of the Self: A Neuro-Archetypical Perspective on the Foundations of Human (and Animal) Subjectivity', *Front. Psychol. 8:1424*, (2017).
doi:10.3389/fpsyg.2017.01424.

[4] Jim Baggott, *Farewell to Reality: How Modern Physics Has Betrayed the Search for Scientific Truth*, Pegasus Books, New York (2012).

[5] Jim Baggott, *A Beginner's Guide to Reality*, Penguin Books, London (2005).

[6] Jim Baggott, *Quantum Reality*, Oxford University Press, Oxford (2020).

[7] Marie T. Banich and Stan Floresco, 'Reward systems, cognition, and emotion: Introduction to the special issue', *Cognitive, Affective, & Behavioral Neuroscience, 19*, 409–414 (2019).
(available at)
https://doi.org/10.3758/s13415-019-00725-z

[8] John A. Bargh and Ezequiel Morsella, 'The Unconscious Mind', *Perspectives on Psychological Science, 2008, 3(1)*, 73-79 (2008).

[9] Lisa Feldman Barrett, *How Emotions Are Made: The Secret Life of the Brain*, Houghton Mifflin Harcourt, Boston (2017).

[10] Robert W. Batterman, *The Devil in the Details: Asymptotic Reasoning in Explanation, Reduction, and Emergence*, Oxford University Press, Oxford (2002).

[11] Rommy von Bernhardi, Jaime Eugenín, and Kenneth J. Muller (editors), *The Plastic Brain*, Springer, Switzerland (2017).

[12] Morten L. Kringelbach and Kent C. Berridge, (ed.) *Pleasures of the Brain*, Oxford University Press, Oxford (2010).

[13] , Michael Berry, 'Asymptotics, Singularities and the Reduction of Theories', (in) Dag Prawitz, Brian Skyrms, and Dag Westerståhl (ed.), *LOGIC, METHODOLOGY AND PHILOSOPHY OF SCIENCE IX: Proceedings of the Ninth International Congress of Logic, Methodology and Philosophy of Science, Uppsala, Sweden, August 7-14,1991*, Elsevier, Amsterdam (1994).

[14] Margaret A. Boden, *The Creative Mind: Myths and Mechanisms*, Routledge, London (2004).

[15] Yelena G. Bodien, Zachary D. Threlkeld, and Brian L. Edlow, 'Default mode network dynamics in covert consciousness', *Cortex, 119*, 571-574 (2019).
doi:10.1016/j.cortex.2019.01.014.

[16] C. D. Broad, *Ethics and the History of Philosophy*, Humanities Press, New York (1952).

[17] Gordon M. Burghardt & Matthew A. Cooper, 'Seven Missteps of Desire', *Neuropsychoanalysis, 14 (1)*, 40-43 (2012).

[18] György Buzsáki, *Rhythms of the Brain*, Oxford University Press, Oxford (2006).

[19] György Buzsáki and Rodolfo Llinás, 'Space and time in the brain', *Science, 358(6362)*, 482-485 (2017).
doi:10.1126/science.aan8869.

[20] György Buzsáki and David Tingley, 'Space and Time: The Hippocampus as a Sequence Generator', *Trends in Cognitive Sciences, 22(10)*, 853-869 (2018).

[21] Robert L. Carneiro, 'The transition from quantity to quality: A neglected causal mechanism in accounting for social evolution', *PNAS, 97(23)*, 12926-12931 (2000); https://doi.org/10.1073/pnas.240462397

[22] Nancy Cartwright, *How the Laws of Physics Lie*, Oxford University Press, Oxford (1983).

[23] Alessia Celeghin, Matteo Diano, Arianna Bagnis, Marco Viola, and Marco Tamietto, 'Basic Emotions in Human Neuroscience: Neuroimaging and Beyond', *Front. Psychol., 8*, article 1432 (2017).
doi:10.3389/fpsyg.2017.01432.

[24] Sergio Chibbaro, Lamberto Rondoni, and Angelo Vulpiani, *Reductionism, Emergence and Levels of Reality: The Importance of Being Borderline*, Springer, Heidelberg (2014).

[25] David F. Clayton, Ina Anreiter, Maria Aristizabal, Paul W. Frankland, Elisabeth B. Binder, and Ami Citri, 'The role of the genome in experience-dependent plasticity: Extending the analogy of the genomic action potential', *PNAS, 117(38)*, 23252-23260 (2020).
doi:10.1073/pnas.1820837116.

[26] Harry Collins, *Tacit and Explicit Knowledge*, The Unversity of Chicago Press, Chicago (2010).

[27] Louis Cozolino, *The Neuroscience of Human Relationships: Attachment and the Developing Social Brain*, W.W. Norton & Company, New York (2014).

[28] David Crews, 'Epigenetics, brain, behavior, and the environment', *HORMONES, 9(1)* 41-50 (2010).

[29] Francis Crick, *The Astonishing Hypothesis: The Scientific Search for the Soul*, Charles Scribner's Sons, New York (1994).

[30] Denise D. Cummins, 'The evolution of Reasoning', in: *The Nature of Reasoning*, Jacqueline P. Leighton and Robert J. Sternberg (eds.), Cambridge University Press, Cambridge, chapter 13 (2004).

[31] Luca D'Alessio, Yariv Kafri, Anatoli Polkovnikov, and Marcos Rigol, *Advances in Physics, Vol. 65, No. 3*, 239-362 (2016).
doi:10.1080/00018732.2016.1198134.

[32] Antonio Damasio, *The Self Comes to Mind: Constructing the Conscious Brain*, Pantheon Books, New York (2010).

[33] R. I. Damper, 'Editorial for the Special Issue on 'Emergent Properties of Complex Systems': Emergence and levels of abstraction', *International Journal of Systems Science, 31(7)*, 811-818 (2000).

[34] Christopher G. Davey, Jesus Pujol, Ben J. Harrison, 'Mapping the self in the brain's default mode network', *Neuroimage, May 15, 132* 390-397, (2016).
doi:10.1016/j.neuroimage.2016.02.022.

[35] Christopher G. Davey, Ben J. Harrison, 'The brain's center of gravity: how the default mode network helps us to understand the self', *World Psychiatry 17:3*, 278-279 (2018).

[36] Richard J. Davidson, Klaus R.J. Scherer, H. Hill Goldsmith, (ed.), *Handbook of Affective Sciences* , Oxford University Press, Oxford (2003).

[37] Kenneth L. Davis and Christian Montag, 'Selected Principles of Pankseppian Affective Neuroscience', *Front. Neurosci. 12:1025*, (2019).
doi:10.3389/fnins.2018.01025.

[38] Jeremy J. Day, Daniel Childs, Mikael C. Guzman-Karlsson, Mercy Kibe, Jerome Moulden, Esther Song, Absar Tahir, and J. David Sweatt, 'DNA methylation regulates associative reward learning', *Nat Neurosci., 16(10)*, 1445–1452 (2013).
doi:10.1038/nn.3504.

[39] Jean Decety and William Ickes, (editors), *The Social Neuroscience of Empathy* , MIT Press, Cambridge (2009).

[40] Jean Decety (editor), *The Social Brain: A Developmental Perspective*, The MIT Press, Cambridge (2020).

[41] Stanislas Dehaene, 'Conscious and Nonconscious Processes : Distinct Forms of Evidence Accumulation?', *Séminaire Poincaré XII*, 89-114 (2009).

[42] Paul Dicken, *A critical Introduction to Scientific Realism*, Bloomsbury, London (2016).

[43] Ap Dijksterhuis, 'Think Different: The Merits of Unconscious Thought in Preference Development and Decision Making', *Journal of Personality and Social Psychology, 87:5*, 586-598, (2004).

[44] J. R. Dorfman, *An Introduction to Chaos in Nonequilibrium Statistical Mechanics*, Cambridge University Press, Cambridge (1999).

[45] Jean-Claude Dreher and Léon Tremblay, (editors), *Decision Neuroscience: An Integrative Perspective*, Academic Press, Amsterdam (2017).

[46] Martin J. Erickson, *Introduction to Combinatorics*, John Wiley & Sons, Inc., New York (1996).

[47] Jonathan St. B. T. Evans and David E. Over, *Rationality in Reasoning*, Psychology Press, Sussex (1996).

[48] Aidan Feeney and Evan Heit (editors), Cambridge University Press, Cambridge, (2007).

[49] John Martin Fischer, Robert Kane, Derk Pereboom, and Manuel Vargas, *Four Views on Free Will*, Blackwell Publishing, Oxford (2007).

[50] Erich Fromm, *The Anatomy of Human Destructiveness*, Holt, Rinehart and Winston, New York (1973).

[51] Armin Fuchs, *Nonlinear Dynamics in Complex Systems*, Springer, Berlin (2013).

[52] Shaun Gallagher (ed.), *The Oxford Handbook of the Self*, Oxford University Press, Oxford (2011).

[53] Peter Gärdenfors, 'Reasoning in Conceptual Spaces', chapter 16 in [1].

[54] Peter Gärdenfors, *Conceptual Spaces: The Geometry of Thought*, The MIT Press, Cambridge, (2000).

[55] Dedre Gentner, Keith J. Holyoak, and Boicho N. Kokinov, (ed.) *The Analogical Mind: Perspectives from Cognitive Science*, The MIT Press, Cambridge (2001).

[56] Alexander George and Daniel J. Velleman, *Philosophies of Mathematics*, Blackwell Publishers, Oxford (2002).

[57] Gerd Gigerenzer, *Gut Feelings: The Intelligence of the Unconscious*, Viking, Penguin Group (USA) Inc., (2007).

[58] Gerd Gigerenzer and Reinhard Selten (eds.), *Bounded Rationality: The Adaptive Toolbox*, MIT Press, Massachusetts (2002).

[59] Efrat Ginot, *The Neuropsychology of the Unconscious: Integrating Brain and Mind in Psychotherapy*, W.W. Norton and Company, New York (2015).

[60] Paul W. Glimcher and Ernst Fehr (ed.), *Neuroeconomcs: Decision Making and the Brain* (2nd ed.), Elsevier, Amsterdam (2014).

[61] Simeng Gu, Fushun Wang, Nitesh P. Patel, James A. Bourgeois and Jason H. Huang, 'A Model for Basic Emotions Using Observations of Behavior in Drosophila', *Front. Psychol. 10*, article 781 (2019). doi:10.3389/fpsyg.2019.00781.

[62] John Guckenheimer and Ricardo A. Oliva, 'Chaos in the Hodgkin–Huxley Model', *SIAM J. APPLIED DYNAMICAL SYSTEMS, vol.1, no. 1*, 105-114 (2002).

[63] Jaryd Hiser and Michael Koenigs, 'The multifaceted role of ventromedial prefrontal cortex in emotion, decision-making, social cognition, and psychopathology', *Biol Psychiatry, 83(8)*, 638–647 (2018). doi:10.1016/j.biopsych.2017.10.030.

[64] Nicole Van Hoeck, Patrick D. Watson, and Aron K. Barbey, 'Cognitive neuroscience of human counterfactual reasoning', *Front. Hum. Neurosci. 9:420* 1-18 (2015).

[65] Douglas Hofstadter and Emmanuel Sander, *Surfaces and Essences: Analogy as the Fuel and Fire of Thinking*, Basic Books, New York, (2013).

[66] John H. Holland, *Complexity: A very Short Introduction*, Oxford University Press, Oxford (2014).

[67] Keith J. Holyoak and Paul Thagard, Mental Leaps: Analogy in Creative Thought, The MIT Press, Cambridge (1995).

[68] Struan Jacobs, 'Polanyi's presagement of the incommensurability concept', *Stud. Hist. Phil. Sci. 33*, 105–120 (2002).
DOI:10.1016/s0039-3681(01)00031-0.

[69] Henrik Jeldtoft Jensen, *Self-Organized Criticality: Emergent Complex Behavior in Physical and Biological Systems*, Cambridge University Press, Cambridge (1998).

[70] Philip N. Johnson-Laird 'Freedom and Constraint in Creativity', (in) R. J. Sternberg (ed.), *The nature of creativity: Contemporary psychological perspectives*, Cambridge University Press, Cambridge (1988).

[71] Robert Kane (ed.), *The Oxford Handbook of Free Will: Second Edition*, Oxford University Press, Oxford (2012).

[72] Daniel Kahneman, Paul Slovic, and Amos Tversky, *Judgment under uncertainty: Heuristics and biases*, Cambridge University Press, Cambridge (1982).

[73] Riya R. Kanherkar, Naina Bhatia-Dey and Antonei B. Csoka, 'Epigenetics across the human lifespan', *Frontiers in Cell and developmental Biology, 2(49)* (2014).
doi:10.3389/fcell.2014.00049.

[74] Eric B. Keverne, Donald W. Pfaff , and Inna Tabansky, 'Epigenetic changes in the developing brain: Effects on behavior', *PNAS 112(22)*, 6789-6795 (2015).

[75] Christof Koch, 'What is Consciousness?'. *Nature, 557*, S9-S12 (2018).

[76] Arthur Koestler, *The Act of Creation*, Hutchinson and Co. (Publishers) Ltd, London (1964).

[77] Morten L. Kringelbach and Kent C. Berridge, 'The Affective Core of Emotion: Linking Pleasure, Subjective Well-Being, and Optimal Metastability in the Brain', Emot Rev. 9:3, 191–199, (2017).
doi:10.1177/1754073916684558.

[78] Morten L. Kringelbach and Kent C. Berridge (editors), *Pleasures of the Brain*, Oxford University Press, Oxford (2010).

[79] Kristján Kristjánsson, *The Self and Its Emotions*, Cambridge University Press, Cambridge (2010).

[80] Thomas S. Kuhn, *The Structure of Scientific Revolutions* (Second Edition, Enlarged), University of Chicago Press, Chicago, (1970).

[81] Irene Lacal and Rossella Ventura, 'Epigenetic Inheritance: Concepts, Mechanisms and Perspectives', *Front. Mol. Neurosci. 11:292* (2018).
doi:10.3389/fnmol.2018.00292.

[82] Avijit Lahiri, *Inference Belief and Interpretation in Science*, self-published ebook, (2020)
[available at]
http://dx.doi.org/10.23668/psycharchives.4284

[83] Christopher C. Leary and Lars Kristiansen, *A Friendly Introduction to Mathematical Logic*, Milne Library, SUNY Geneseo, Geneseo, NY (2015).

[84] Mark R. Leary and June Price Tangney (editors), *Handbook of Self and Identity*, The Guilford Press, New York (2012).

[85] Wentian Li, 'Random Texts Exhibit Zipf's-Law-Like Word Frequency Distribution', *SFI WORKING PAPER: 1991-03-016*, (1991).
[available at]
urlhttp://115.28.82.164/sites/default/files/test%20pdf1.pdf

[86] Abninder Litt, Chris Eliasmith, and Paul Thagard, 'Neural affective decision theory: Choices, brains, and emotions', *Cognitive Systems Research 9*, 252-273, (2008).

[87] George Loewenstein and Jennifer S. Lerner, 'The Role of Affect in Decision Making', (in) Richard J. Davidson, Klaus R. Scherer, and H. Hill Goldsmith, (ed.) *Handbook of Affective Sciences*, Oxford University Press, Oxford, (2003).

[88] Steven Lukes, *Power: A Radical View*, The Macmillan Press Ltd., London (1974).

[89] Laura Macchi, Maria Bagassi, and Riccardo Viale (editors) *Cognitive Unconscious and Human Rationality*, MIT Press, Cambridge (2016).

[90] Ken Manktelow, *Reasoning and Thinking*, Psychology Press, Sussex (1999).

[91] R. M. May, 'Simple mathematical models with very complicated dynamics', *Nature 261*, 459–467 (1976).

[92] Emeran A. Mayer, 'Gut feelings: the emerging biology of gut–brain communication', *Nat Rev Neurosci, 12(8)* 453-466, (2011).
(available at)
https://dx.doi.org/10.1038/nrn3071

[93] V. J. McGill and W. T. Parry, 'The Unity of Opposites: A Dialectical Principle', *Science & Society, (published by Guilford Press) Vol. 12, No. 4 (Fall, 1948)*, pp. 418-444 (1948).

[94] Peter B. Medawar, *Induction and Intuition in Scientific Thought*, Methuen & Co Ltd, London (1970).

[95] Jonathan Mercer, 'Emotional Beliefs', *International Organization, 64*, 1-31 (2010). doi:10.1017/S0020818309990221.

[96] Rami Mohieddine, 'Chaos in the Hodgkin-Huxley Equations: The Takens-Bodganov Cusp Bifurcation', *Cornell University Mathematics Department Senior Thesis*, (2008).
(available at)
https://pi.math.cornell.edu/m/sites/default/files/imported/Research/SeniorTheses/2008/mohieddineThesis.pdf

[97] Andrea Nani, Jordi Manuello, Lorenzo Mancuso, Donato Liloia, Tommaso Costa, and Franco Cauda, 'The Neural Correlates of Consciousness and Attention: Two Sister Processes of the Brain', *Front. Neurosci., 13:1169* (2019). doi:10.3389/fnins.2019.01169.

[98] Donald L. Nathanson, *Shame and Pride: Affect, Sex, and the Birth of the Self*, W.W. Norton & Company, New York (1994).

[99] Ben R. Newell and David R. Shanks, 'Unconscious influences on decision making: A critical review', *BEHAVIORAL AND BRAIN SCIENCES, 37:1*, 1-61, (2014). doi:10.1017/S0140525X12003214.

[100] Ikka Niiniluoto, 'Scientific Progress', (in) Juha Saatsi (ed.), *The Routledge Handbook of Scientific Realism*, Routledge, London (2018).

[101] Timothy O'Connor, "Emergent Properties", The Stanford Encyclopedia of Philosophy (Winter 2021 Edition), Edward N. Zalta (ed.), (2021).
(available at)
https://plato.stanford.edu/archives/win2021/entries/properties-emergent/.

[102] Hadas Okon-Singer, Talma Hendler, Luiz Pessoa, and Alexander J. Shackman, 'The neurobiology of emotion–cognition interactions: fundamental questions and strategies for future research', *Frontiers in Human Neuroscience, 9(58)*, (2015). doi:10.3389/fnhum.2015.00058.

[103] Jaak Panksepp, *Affective Neuroscience: The Foundations of Human and Animal Emotions*, Oxford University Press, Oxford (2004).

[104] Elaine Perry, Daniel Collerton, Fiona LeBeau, and Heather Ashton (editors), *New Horizons in the Neuroscience of Consciousness*, John Benjamins Publishing Company, Amsterdam (2010).

[105] Elaine Perry, Heather Ashton, and Allan Young (editors), *Neurochemistry of Consciousness: Neurotransmitters in mind*, John Benjamins Publishing Company, Amsterdam (2002)

[106] Thomas Pink, *Free Will: A Very Short Introduction*, Oxford University Press, Oxford (2004).

[107] Michael Polanyi, *The Tacit Dimension*, The University of Chicago Press, Chicago (1966).

[108] Marcus E. Raichle, 'Creativity and the Brain's Default Mode Network', (in) Suzanne Nalbantian and Paul M. Matthews (ed.), Secrets of Creativity: What Neu-

roscience, the Arts, and Our Minds Reveal, Oxford Scholarship Online, September (2019).
doi:10.1093/oso/9780190462321.001.0001.

[109] Arthur Reber, *Implicit Learning and Tacit Knowledge*, Oxford University Press, New York, Oxford (1993).

[110] Thomas P. Reber, Roger Luechinger, Peter Boesiger, and Katharina Henke, 'Detecting analogies unconsciously', *Frontiers in Behavioral Neuroscience, 8*, 1-13, (2014).

[111] Mathieu Roy, Daphna Shohamy, and Tor D. Wager, 'Ventromedial prefrontal-subcortical systems and the generation of affective meaning', *Trends Cogn Sci., 16(3)*, 147–156 (2012).
doi:10.1016/j.tics.2012.01.005.

[112] David Ruelle, *Chance and Chaos*, Princeton University Press, Princeton (1991).

[113] Jutta Schickore and Friedrich Steinle, *Revisiting Discovery and Justification: Historical and philosophical perspectives on the context distinction*, Springer, Dordrecht (2006).

[114] Allan N. Schore, *The Development of the Unconscious Mind*, W.W. Norton and Company, New York (2019).

[115] Gerhard Schurz, 'Thruthlikeness and Approximate Truth', (in) Juha Saatsi (ed.), *The Routledge Handbook of Scientific Realism*, Routledge, London (2018).

[116] John R. Searle, *The Mystery of Consciousness*, New York Review, New York (1997).

[117] Constantine Sedikides and Steven J. Spencer (editors), The Self, Psychology Press, New York (2007).

[118] Daniel.J. Siegel, *The Developing Mind: How Relationships and the Brain Interact to Shape Who We Are*,

[119] Goran Simic, Mladenka Tkalcic, Vana Vukic, Damir Mulc, Ena Spanic, Marina Sagud, Francisco E. Olucha-Bordonau, Mario Vuksic, and Patrick R. Hof, 'Understanding Emotions: Origins and Roles of the Amygdala', *Biomolecules, 11*, 823

(2021).
https://doi.org/10.3390/biom11060823.

[120] Dean Keith Simonton, *Origins of Genius: Darwinian Perspectives on Creativity*, Oxford University Press, New York (1999).

[121] Steven H. Shmurak, 'Demystifying Emotion: Introducing the Affect Theory of Silvan Tomkins to Objectivists', *The Journal of Ayn Rand Studies (Penn State University Press), 8:1*, 1-18 (2006).

[122] Herbert A. Simon, 'The Scientist as Problem Solver', in: Complex Information Processing: The Impact of Herbert A. Simon, David Klahr and Kenneth Kotovsky (eds.), Lawrence Erlbaum Associates, Inc., Hillsdale, NJ, chapter 14 (1989).

[123] Aaron C.T. Smith, *Cognitive Mechanisms of Belief Change*, Palgrave MacMillan, London, (2016).

[124] Léna Soler, Howard Sankey, and Paul Hoyningen-Huene (ed.), *Rethinking Scientific Change and Theory Comparison: Stabilities, Ruptures, and Incommensurabilities?*, Springer, Dordrecht (2008).

[125] Chun Siong Soon, Anna Hanxi He, Stefan Bode, and John-Dylan Haynes, 'Predicting free choices for abstract intentions', *PNAS April 9, 110(15)*, 6217-6222 (2013).
https://doi.org/10.1073/pnas.1212218110

[126] Olaf Sporns, *Networks of the Brain*, The MIT Press, Massachusetts (2011).

[127] Larry R. Squire and Adam J.O. Dede, 'Conscious and Unconscious Memory Systems', *Cold Spring Harb Perspect Biol., 7(3)*, a021667 (2015).

[128] Nicholas F Stang, 'Kant's Transcendental Idealism', The Stanford Encyclopedia of Philosophy (Spring 2021 Edition), Edward N. Zalta (ed.) (2021).
(available at)
https://plato.stanford.edu/archives/spr2021/entries/kant-transcendental-idealism/

[129] Keith E. Stanovich,*Decision Making and Rationality in the Modern World*, Oxford University press, Oxford (2003).

[130] Daniel L. Stein and Charles M. Newman, *Spin glasses and Complexity*, Princeton University Press, Princeton, (2013).

[131] Hans Steiner, (with) Rebecca Hall, *Your Secret Mind: Getting to know and Living with Your Unconscious*, Karnac Books Limited, London (2017).

[132] Joan Stiles and Terry L. Jernigan, 'The Basics of Brain Development', *Neuropsychol Rev., 20(4)*, 327–348 (2010).
doi:10.1007/s11065-010-9148-4.

[133] Stefan Thurner, Rudolf Hanel, and Peter Klimek, *Introduction to the Theory of Complex Systems*, Oxford University Press, Oxford (2018).

[134] , Paola Tognini, Debora Napoli and Tommaso Pizzorusso, 'Dynamic DNA methylation in the brain: a new epigenetic mark for experience-dependent plasticity', *Front. Cell. Neurosci. 9:331* (2015).
doi:10.3389/fncel.2015.00331.

[135] Michael Tomasello, *A Natural History of Human Thinking*, Harvard University Press, Cambridge, MA (2014).

[136] Giulio Tononi and Christof Koch, 'The Neural Correlates of Consciousness: An Update', *Ann. N.Y. Acad. Sci. 1124*, 239–261 (2008).
doi:10.1196/annals.1440.004.

[137] Albert Tsao, Jørgen Sugar, Li Lu, Cheng Wang, James J. Knierim, May-Britt Moser & Edvard I. Moser, 'Integrating time from experience in the lateral entorhinal cortex', *Nature, 561*, 57-62 (2018).
doi:10.1038/s41586-018-0459-6.

[138] Donald L. Turcotte and John B. Rundle, 'Self-organized complexity in the physical, biological, and social sciences', Proc Natl Acad Sci USA, 99(Suppl 1), 2463-2465 (2002).
doi:10.1073/pnas.012579399.

[139] Bas van Fraassen, *The Scientific Image*, Oxford University Press, Oxford (1980).

[140] Oshin Vartanian and David R. Mandel (ed.), *Neuroscience of Decision Making*, Psychology Press, New York (2011).

[141] Vlatko Vedral, *Decoding Reality: The Universe as Quantum Information*, Oxford University Press, Oxford (2010).

[142] Stella Vosniadou and Andrew Ortony, (ed.), *Similarity and analogical reasoning*, Cambridge University Press, Cambridge (1989).

[143] Jeanne C. Watson and Leslie S. Greenberg, 'Empathic Resonance: A Neuroscience Perspective', chapter 10 in [39] (2009).

[144] Daniel M.Wegner, 'Who is the Controller of Controlled processes?', in: *The New Unconscious*, Ran R. Hassin, James S. Uleman, John A. Bargh, (eds.), Oxford University Press, Oxford, chapter 1 (2005).

[145] Eugene P. Wigner, 'The Unreasonable Effectiveness of Mathematics in the Natural Sciences', *Communications on Pure and Applied Mathematics, VOL. XIII*, 1-14 (1960).

[146] Wikiversity contributors, 'Motivation and emotion/Book/2017/Neurotransmitters and emotion', Wikiversity, 24 August 2021, 12:25 UTC, https://en.wikiversity.org/w/index.php?title=Motivation_and_emotion/Book/2017/Neurotransmitters_and_emotion&oldid=2309265 [accessed 24 August 2021]

[147] Timothy D. Wilson, *Strangers to Ourselves: Discovering the Adaptive Unconscious*, Harvard University Press, Cambridge (2002).

[148] Piotr Winkielman and Kent C. Berridge, 'Unconscious Emotion', *Current Directions in Psychological Science, 13(3)*, 120-123, (2004).

[149] Peter Woit, *Not Even Wrong: The Failure of String Theory and the Continuing Challenge to Unify the Laws of Physics*, Jonathan Cape, London (2006).

[150] Jason S. Wright & Jaak Panksepp, 'An Evolutionary Framework to Understand Foraging, Wanting, and Desire: The Neuropsychology of the SEEKING System', *Neuropsychoanalysis, 14(1)*, 5-39 (2012).

[151] Wayne Wu, 'The Neuroscience of Consciousness', The Stanford Encyclopedia of Philosophy (Winter 2018 Edition), Edward N. Zalta (ed.).
https://plato.stanford.edu/archives/win2018/entries/consciousness-neuroscience/

[152] Yaara Yeshurun, Mai Nguyen, Uri Hasson, 'The default mode network: where the idiosyncratic self meets the shared social world', *Nat Rev Neurosci., 22(3)*, 181–192 (2021).
doi:10.1038/s41583-020-00420-w.

[153] Philip David Zelazo, Morris Moscovitch and Evan Thompson (editors), *The Cambridge Handbook of Consciousness*, Cambridge University Press, Cambridge (2007).

Index