Chapter 1
Introduction: Anticipation in Natural and Artificial Cognition

Giovanni Pezzulo, Martin V. Butz, Cristiano Castelfranchi, and Rino Falcone

The purpose of brains is to produce future. *Paul Valery*

1.1 Introduction

What will artificial cognitive systems of the future look like? If we are asked to imagine robots, or intelligent software agents, several features come to our mind such as the capability to adapt to their environments and to satisfy their goals with only limited human intervention, to plan sequences of actions for realizing long-term objectives, to act collectively in view of complex objectives, to interact and cooperate with us, with and without natural language, to take decisions (also in our place), etc.

Currently these capabilities are far beyond the possibilities of robots and other artificial systems. In the next years a huge effort will be required for scaling up the potentialities of the artificial systems that we are able to build nowadays. One way to overcome these limitations is to take inspiration form the functioning of living organisms. A large body of evidence, which we review in this chapter, indicates that natural cognitive systems are not reactive but essentially anticipatory systems. We do not think that this is a mere coincidence. On the contrary, we claim that anticipation is a crucial—and foundational—phenomenon in natural cognition. Individual behavior is guided by anticipatory mechanisms that are used for behavioral control, perceptual processing, goal-directed behavior, and learning. And also effective social behavior relies on the anticipation of the behavior of other agents. We argue that anticipation is a key ingredient for the design of autonomous, artificial cognitive agents of the future: *Only cognitive systems with anticipation mechanisms can be credible, adaptive, and successful in interaction with both the environment and other autonomous systems and humans*. This is the challenge that we anticipate for the future of cognitive systems research: the passage from reactive to anticipatory cognitive embodied systems.

G. Pezzulo et al. (Eds.): The Challenge of Anticipation, LNAI 5225, pp. 3–22, 2008.

1.1.0.1 From Reactive to Anticipatory Cognitive Embodied Systems

Overall, we propose an integrated approach to the study of anticipation that encompass empirical, theoretical, and computational approaches. The study of anticipation has a long history in the empirical literature, that we will review extensively in the next section. In Chapter 2, we focus on the investigation of anticipatory functionalities from a conceptual point of view as well as from a computational one. Despite some studies available on this topic, we believe that a unitary approach to the study of anticipation is still missing. This book intends to address exactly this challenge.

1.1.0.2 Overview of This Chapter

The first part of this chapter is conceptual in nature. It consists in finding 'conceptual keys' to understand the phenomena of cognition and behavior, and to use them to inspire our design methodology. For this reason, in Section 1.2 we propose theoretical arguments for assessing the relevance of anticipation and anticipatory behavior in cognition. We start from a discussion of the role anticipation has played in cognitive science and artificial systems research. Particularly, we highlight why living organisms endowed with anticipatory capabilities are able to develop cognitive capabilities—from simpler to more complex ones—including those based on representations. In conclusion, we propose that *cognitive minds inevitably have to be anticipatory devices*. We also argue that studying anticipatory mechanisms in the brain in isolation does not suffice. Rather, it is also necessary to understand the nature of future-oriented behavior, and of future-oriented representations: which are their specific advantages, why do cognitive agents need to anticipate, etc.

Since this book aims to offer design principles for endowing artificial systems with anticipatory capabilities, it is essential to analyze in detail the specific roles of anticipation in several cognitive functions, including sensorimotor control, attention, internal preparation to action, emotional regulation, learning, exploration, curiosity, and decision making. For this purpose, in the second part of this chapter we focus on how anticipatory mechanisms actually work in living organisms: Section 1.3 gives a review of psychological and neuroscientific theories and models of anticipation. This systematic exploration of natural anticipatory systems is meant to be a source of inspiration for the sake of designing artificial anticipatory systems that have the same levels of adaptivity, flexibility, and autonomy.

1.2 The Path to Anticipatory Cognitive Systems

Before we delve into the different aspects of predictions and anticipatory capabilities, we first sketch out the research path that lead to explicit studies of anticipatory cognitive mechanisms such as the formulation of the theory of *anticipatory behavioral control* (Hoffmann, 1993; Hoffmann et al., 2004), the study of *anticipatory behavior* (Butz et al., 2003b, 2007b), the proposition of *the mind as an anticipatory device* (Castelfranchi, 2005; Pezzulo and Castelfranchi, 2007), and ultimately the authoring of this book.

1.2.1 Symbolic Behavior, Representation-Less Behavior, and Their Merge to Anticipatory Behavior

Traditionally, artificial intelligence (AI) investigated the functionality of symbol-oriented cognitive mechanisms, such as search, planning, and decision making in well-defined, discrete problems, principles of first-order logic, of learning based on symbolic inputs (Russell and Norvig, 1995). The concept of representation as traditionally defined and used in AI was rather detached from actually available sensory information, yielding impressive performance in well-defined environments, such as the game of chess, but highly unsatisfactory performance in natural environments, such as for adaptive robot control. In the latter case, the traditional AI-based systems suffered from several fundamental problems: (1) The scalability problem restricted the systems to solve only highly simplified and very small toy problems. (2) The symbol grounding problem (Harnad, 1990) prevented them from identifying effective sensory discretizations, so that effective symbolic representations, which may be suitable for planning or decision making, did not emerge. (3) The frame problem (McCarthy and Hayes, 1969) prevented systems from effectively representing action-affected and -unaffected parts of the environment with logic-based representations.

As a consequence, the situated approach to cognition gained popularity (Brooks, 1991; Chiel and Beer, 1997; Pfeifer and Scheier, 1999). The situated approach challenges several weak points of traditional AI methodology such as that all 'cognitive' functions (including perception, categorization, etc.) are assumed to be based on (the manipulation of) internal representations and that reasoning is overemphasized in comparison with situated motor activity. The subsequent situated AI approach had a great impact on our current understanding of cognition, learning, and adaptive behavior. For example, it was shown that the effective coupling of brain-body-environment dynamics, without representational processes, can yield very efficient, representation-less behavioral patterns (Braitenberg, 1984; Brooks, 1991). Many of the consequentially realized cognitive functions were previously considered representational in nature. The consequence of these behavioral successes lead to a critical reconsideration of the role of representations and internal processes.

A fundamental side effect of this approach is the emphasis on *reactive mechanisms*, which mainly originates from the necessity to avoid the grounding problem and at the same time to de-emphasize the role of internal representations. This fact has produced a lack of interest for representational processes, which are however widespread in natural cognition, and resulted in skepticism with respect to the need of internal models and representations.

Although reactive and non-representational systems have shown a range of capabilities that were unsuspected, it became also increasingly clear that they will never reach the levels of complexity observable in natural cognitive systems. On the contrary, the important role for anticipations and anticipatory mechanisms in natural cognition is highlighted in several empirical studies. Nature seems to have found a suitable way to overcome the shortcomings of reactive systems by endowing living organisms with anticipatory mechanisms.

1.2.2 The Power of Anticipation: From Reactivity to Proactivity

Reactive systems are those that produce behavior as a response or reaction to (sensed) environmental conditions and internal needs. They do not need to have a complex representation of their environment since it is sufficient for them to sense it. Take as an example a reactive driving rule: *If you see the car in front of you stopping (e.g., you see the red lights indicating the stop), then press the brake*. In normal traffic conditions, a reactive system endowed with this rule is able to avoid accidents most of the time. In artificial systems research, reactive rules (independently of how they are implemented) lead to efficient systems, since the computation they have to carry out is simple and cheap. However, these systems are not versatile, since they tend to have stereotyped responses, and they are not able to prepare for future conditions, but rather they have to wait for the conditions to occur first.

On the contrary, a system endowed with predictive capabilities can use the following rule: *If the car in front of you is close to a crossing, then it is likely to stop, so stop in advance or at least get ready to stop when a crossing is ahead*. Thus, a system endowed with predictive capabilities can take into account (possible) future events to decide on and prepare current behavior.

Predictive capabilities permit even much more subtle behavior. Anticipatory systems can, for example, select an action whose anticipated effect is judged to be positive, prevent dangers before experiencing them, actively search for information that is expected to be relevant, etc. All these capabilities, that are based on processing information relative to the future, are the keys for passing from mere *reactivity* to *proactivity* and *goal-oriented behavior*.

1.2.3 The Anticipatory Approach to Cognitive Systems

By proposing the *anticipatory approach* to cognitive systems, we argue that—now that the criticism of traditional AI from the situated approach is quite well accepted—it is time to reconsider representations. However, these representations need to be integrated into the situated approach to AI. That is, representations may emerge out of representation-less systems and may suitably alter the capabilities of the situated systems. Thus, representations may no longer be detached symbols, but they will need to be grounded in the body's sensory and motor systems and situated in the perceived environment. Our strong belief is that such modern, behaviorally-suitable representations can emerge from the anticipatory approach to cognition. We therefore propose to focus on how anticipations are realized in living organisms, and to investigate how anticipatory representations permit the realization of cognitive functions.

The Mind Is an Anticipatory Device One central tenet of our anticipatory approach to cognition is that a true cognitive mind serves for (and has evolved for) anticipation: *The mind is an anticipatory device* (Castelfranchi, 2005; Pezzulo and Castelfranchi, 2007). Anticipation is not only required for several cognitive functions, but it is an 'ordering principle' of cognition and its development. For this reason, the study of anticipatory phenomena can shed light onto natural cognition—a view that is currently also gaining consensus in the neuroscientific community (Bar, 2007; Frith, 2007; Hawkins and Blakeslee, 2004). For example, after reviewing a number of memory studies and theoretical analysis, Schacter et al. (2007, pg. 660) concludes that:

> Given the adaptive priority of future planning, we find it helpful to think of the brain as a fundamentally prospective organ that is designed to use information from the past and the present to generate predictions about the future.

There are two main reasons for conceiving cognitive minds as essentially anticipatory and future-oriented. First, cognition should be described as an active and productive activity rather than a passive stimuli-processing system. Second, representational and symbolic capabilities were only able to develop due to adaptive advantages of anticipating and dealing with the future. For this reason, the capability to form grounded representations and symbols depends on the capability to anticipate. We illustrate these points in further detail in the two following sections.

1.2.3.1 The Productive View of Cognition

The *productive view of cognition* that we put forward in our anticipatory approach originates from Kant's (1998) idea that, although our knowledge begins with experience, it does not purely arise from experience, since our productive, generative apparatus determines what we know. We do not passively process environmental stimuli, but actively produce representations by means of our categorical apparatus.

> That all our knowledge begins with experience there can be no doubt. [...] But, though all our knowledge begins with experience, it by no means follows that all arises out of experience. For, on the contrary, it is quite possible that our empirical knowledge is a combination of that which we receive through impressions, and [additional knowledge] altogether independent of experience [...] which the faculty of cognition supplies from itself, sensory impressions giving merely the occasion. (Kant, 1998, Introduction)

This idea has been very important in the theory of Piaget (1954), that introduces as an important element of novelty an emphasis on the situated and action-based origin and nature of representations.

> Any piece of knowledge is connected with an action . . . [T]o know an object or a happening is to make use of it by assimilation into an action schema . . . [namely] whatever there is in common between various repetitions or superpositions of the same action. (Piaget, 1971, pg. 6-7)

Central in the Piagetian theory is the concept of schemas (or, better, sensorimotor schemas), which is a highly recognized concept in cognitive science (Arbib, 1992, 2003; Bartlett, 1932; Neisser, 1976). He describes cognitive development in humans as a process of *assimilation* and *accommodation*, in which schemas of increasing complexity are formed to make sense of the world and to operate on it. A central posit, which is generally adopted in schema-based computational modeling (Arbib, 1992; Drescher, 1991; Pezzulo and Calvi, 2007b; Roy, 2005), is that acting on the basis of an action schema also entails the expectation of action effects. Expectations are crucial in the control of action and categorization—two sides of the same coin—since the compliance of action implicitly verifies the expectations and permits the categorization of an object or an event. Expectations are also used for assimilation, that is, learning of a novel schema when current expectations are not met (cf. Pezzulo and Calvi, 2007b for a further discussion on the relations between the pragmatic and epistemic sides of action schemas).

The productive aspect of cognition is particularly important in cognitive, goal-oriented agents. Contrary to the view that cognitive agents can be represented as input-output devices that passively receive inputs for reacting appropriately, we argue that expectations precede stimuli, both factually and conceptually. Factually, anticipatory representations are already there before inputs are received. Conceptually, a true goal-oriented behavior begins with a goal and not with a stimulus.

Recently a great deal of evidence has been accumulated that strengthens this view. Bar (2007) proposes that the mind, thanks to associative mechanisms, is proactive and continuously generates predictions approximating the relevant future—a position that is consistent with the idea of the mind as an anticipatory device. Also in accordance with the view put forward in this chapter, he suggests therefore that anticipation is one (of the few) unifying principle(s) of brain functioning.

With a slightly different emphasis (on memory studies rather than proactivity and goal orientedness), several authors have recently proposed that the essential function of memory is not ‘storage’ but enabling dealing with the future. Some examples are ‘mental time travel’ (Tulving, 1983), ‘memory of the future’ (Ingvar, 1985), ‘memory for the future’ (Glenberg, 1997), and the ‘prospective brain’ (Schacter et al., 2007). All these studies highlight complementary aspects of the productive view of cognition and indicate anticipation as a basic, unitary capability of cognition that produces cognitive and behavioral effects.

1.2.3.2 The Mind Originates from the Need to Deal with the Future

We have reviewed evidence indicating that the organization of the perceptual and motor apparatus is biased toward the future. However, here we do one more step and suggest that the cognitive mind’s main function is to anticipate and to deal with the future. In describing animals as machines evolved for the survival and propagation of their genes, Dawkins (1989, pg. 59) argues that

> Survival machines that can simulate the future are one jump ahead of survival machines who can only learn on the basis of overt trial and error.

We have already discussed how anticipation permits the channelization of epistemic and pragmatic activity. But there is another, perhaps more fundamental reason for considering the role of anticipation as essential in the development of cognitive minds: internal representations might have emerged from anticipatory mechanisms such as internal models for the sake of dealing with the future.

The Emergence of Representations from Anticipation What distinguishes a cognitive from a merely adapted system is the capability to form internal representations (and in particular expectations) and to work on them internally before, or instead of, operating directly on the environment. Endogenously producing representations, and working on them internally instead of immediately acting (Piaget's *substitution*) is a hallmark of cognition.

However, building an internal model before acting is costly. The knowledge-based approach in AI has received criticisms by the situated approach, exemplified by Brooks's (1991) idea that 'the world is the best representation of itself', and in theory, due to the frame problem that mines the idea that we can formulate a symbolic representation of the world and possible actions before acting –and in fact Dennett's (1984) analogy with the *Buridan's ass* illustrates the problems of model based approaches.

One crucial challenge is then to understand how representations can arise in situated systems and how the cost of representation is balanced by appropriate gains. It is extremely advantageous for a situated system to be able to deal with the future and not only the present. An anticipation of the future implies gains in specific cognitive functions such as attention and motor control as well as the development of completely new capabilities such as planning (we review extensively the benefits of anticipation in Chapter 3).

Thus, representations might have originated thanks to the need for dealing with the future, for which the direct way is to predict and represent the future. In turn, representing the future and detaching one's own representations from the current sensorimotor interaction might have provided several other advantages. Representing the future, including future events, our own actions in the future as well as other's actions, is essential for coordinating one's own acts for a long time span, for coordinating with others, for realizing future states that are desirable (this includes controlling others), or for avoiding dangerous futures.

Another related (and not concurrent) hypothesis is put forward by König and Krüger (2006) who argue that discrete entities (symbols) emerge in the brain in the process of feature extraction as a byproduct of data compression for the sake of permitting better predictions:

> In the process of mutual optimization of features and predictions, symbols emerge as condensed entities on which predictions are performed. (König and Krüger, 2006, pg. 14)

In a similar vein, several researchers stress the role of anticipation in the development of cognitive capabilities, from the simplest sensorimotor ones to more complex and symbolic ones such as language (Clark and Grush, 1999; Gardenfors and Orvath, 2005; Grush, 2004).

Several studies in cognitive robotics are now beginning to investigate anticipation from a situated perspective. Although they mainly focus on the control of action and basilar social abilities so far, this direction seems very promising to scale up to higher level cognitive and social functions. Understanding the nature and functioning of anticipatory behavior in unitary perspective will be, however, a big theoretical challenge since anticipatory representations have a double-sided nature, being both *grounded* and *detached*. According to the situated approach to cognition, representations have to be *grounded* (Harnad, 1990): An agent can only act adaptively if it stays intimately coupled with its environment. At the same time, representations and especially expectations and distal goals are by definition about future states of affairs (sometimes even impossible ones), and they are therefore *detached* from the current sensorimotor engagement. Thus, groundedness and detachment seem to be at odds: how can both be obtained?

The answer can come from an investigation, in a developmental perspective, of the *detachment process* that permits to develop anticipatory representations starting only from sensorimotor engagement: representations are not *born* detached, but *become* detached. The reviewed literature offers some indications about possible stages of the detachment process. Accordingly, several authors (Clark and Grush, 1999; Grush, 2004; Pezzulo and Castelfranchi, 2007) propose that internal models permitting to *emulate* the external reality are firstly developed for the sake of action control. Once established, they are exapted (i.e., used for a function other than that for which it was developed) for bootstrapping increasingly complex functionalities such as simulative planning and pursuing distal goals. The diversity of anticipatory capabilities existing in nature thus depends on a progressive process of disengagement from the current sensorimotor cycle, enabled by progressively detached anticipatory representations. Since this process exploits an ‘inner simulation’ mechanism, however, anticipatory representations remain intimately related to situated action and maintain its nature (Barsalou, 1999; Damasio, 1994; Grush, 2004; Hesslow, 2002). The same process can also be in place for social cognition, since anticipatory mechanisms for engaging in future-oriented and social-oriented activity share the same neural substrate.

Consistently, Clark and Grush (1999) put forward the challenge of understanding the anticipatory aspects of cognition in a naturalistic framework, and as originating from situated action. They argue that a crucial step toward truly *cognitive* robotics is reframing the concept of representation in a situated and embodied perspective. They propose that anticipation is the key element, and anticipatory mechanisms (in particular simulative mechanisms) are responsible for bootstrapping grounded representations:

> agents [that genuinely cognize their worlds] are able to substitute inner dynamics for ongoing environmental stimulation, and command adaptively valuable inner spaces that they use to sculpt and modulate their more direct engagements with the world. It is these ‘Cartesian Agents’ we believe, that must form the proper subject matter of any truly cognitive robotics. Clark and Grush (1999, p. 13)

Grush (2004) has put forward this idea and proposed the *emulation theory of representation*: the most comprehensive account nowadays on how representation originates from anticipatory mechanisms that can be used online for action control and re-enacted off-line for enabling a number of sophisticated cognitive capabilities such as visual imagery, reasoning, theory of mind phenomena, and language. A conceptual analysis of the passage from sensorimotor skills to higher level cognitive capabilities based on anticipation can also be found elsewhere (Gardenfors, 2003; Hurley, 2005; Pezzulo and Castelfranchi, 2007; Pezzulo, 2008a).

Representations Are Grounded Because They Remain Related to Prediction If we assume that representations arise for prediction, and continue to depend functionally on prediction, we can understand them in a novel perspective, that is intimately related to situated action. One definition that is compliant with this view is provided by Bickhard and Terveen's (1995) *interactivism*: representations are ways for setting up indications of further interactive potentialities, and thus serve for future interactions. A related view is Smith's (1996) *intentional dance*.

An example of robotic implementation of the Piagetian, constructivist approach to the formation of object representations can help illustrate this point. In the study by Drescher (1991) concepts for objects are developed autonomously on the basis of (actual or expected) interaction effects. Objects (called *synthetic items*) are not provided to the agent but 'discovered', or better postulated, as a common cause of the expected success of a number of actions; objects are then explanations of sensorimotor patterns. If the agent moves its hand to the left and touches (or expects to touch) a surface, or moves its eyes to the left and sees a circular shape, etc., it can postulate that there is a common cause in all these behavioral effects and then 'create' a synthetic item. Later on, it can predicate based upon this item for forming more complex representations, including action representations.

In a similar vein, Roy (2005) has proposed that concepts for objects, which are, for example, reachable or graspable, are grounded by object schemas (similar in spirit to synthetic objects), which regulate actual behavior and at the same time encode predictions on the consequences of an expected interaction. One advantage of this framework is that schemas for actions, objects, and linguistic symbols share the same representational basis and have demonstrated to be successful in complex cognitive tasks such as robot control and linguistic communication.

This approach has implications for the symbol grounding problem as well. Roy (2005) has proposed that it depends on two mechanisms relating agent and environment: causation (from environment to agent) and anticipation (from agent to environment). Representations, including goal representations, are thus grounded thanks to the circular causality of ACTION + EXPECTATION and OBSERVATION + CAUSATION. This circular mechanism could also explain how we attribute *causality*: it is our productive apparatus that permits the reading of events in the world as causally (instead of simply statistically) related, which is essentially the Kantian solution to the problem of causality.

From a philosophical point of view, this approach can solve the problem of how to justify representations without falling in the grounding problem. On the one side we want to highlight the role of representational, anticipatory processes in cognitive agents, but on the other side we need a naturalistic account for representations that is entirely compatible with situated and embodied approaches to cognition. This point is illustrated nicely in Clark's (1998) *minimal representationalism*: "Minds may be essentially embodied and embedded and still depend crucially on brains which compute and represent."

1.2.4 The Unitary Nature of Anticipation

Overall, we have illustrated how several conceptual frameworks have been developed that indicate anticipation as a key element for cognition and for the development from simpler to more complex cognitive capabilities. As a conclusion to this section, we want to stress one of the innovative aspects of the anticipatory approach to cognition. Notwithstanding the fact that anticipation has multiple facets and has multiple realizations in brains and behaviors, we believe that it has to be considered a *unitary phenomenon*, a hallmark for natural and artificial cognition. Consistently with the idea of a cognitive mind as an anticipatory device, we argue that anticipation is inherently involved in—and in many cases a necessary condition for—several cognitive functions. Once a cognitive mind has evolved the power to deal with the future, this opens the possibility of an entirely new set of capabilities and opportunities, and it can realize proactive and goal-oriented behavior.

For this reason, we believe that a real understanding of the phenomenon of anticipation will come from a study of its unitary aspects rather than (or, better, together with) its different realizations, behavioral and neural. Our objective is then to provide a unitary perspective on the study of cognition and its development by focusing on anticipation and, more in general, on the capability and need to deal with the future. In the rest of the book we pursue this objective in several ways: We provide adequate definitions and taxonomies that help highlight the unitary aspects of prediction and anticipation, and we analyze the powers and limitations of anticipation in natural and artificial cognition, also providing a number of examples.

Since our analysis is grounded in biological and psychological evidence, we now proceed with reviewing evidence for anticipatory phenomena in cognition, both in simple and complex organisms, and we illustrate a unifying view of natural cognition based on anticipation.

1.3 Anticipation in Living Organisms

Besides the conceptual perspective on AI research progress, that is, from symbols, to reactivity and situatedness, back to combinations of these mediated by anticipatory mechanisms, there is also a biological, psychological perspective that strongly points toward the ubiquity of anticipatory mechanisms in animals and humans.

Several converging directions of empirical research indicate a crucial role for anticipatory mechanisms in cognitive functions. These mechanisms range from simple, such as sensorimotor coordination, to highly complex, such as decision making in social domains, social imitation and learning, or communication. In this section, we review biological and psychological evidence for anticipatory mechanisms in the brain and the consequent behavioral capabilities of animals and humans.

1.3.1 Anticipatory Natural Cognition

Several animal and human capabilities require an estimation of future states of affairs for compensating the dynamicity of the environment: for example, the motor preparation of the prey-catching behavior of the jumping spider (Schomaker, 2004) or balancing a pole with one hand (Mehta and Schaal, 2002). It has even been proposed that all motor control is mediated by anticipatory information, which is generated by internal predictive models that permit the emulation of the environment (Doya, 1999; Kawato, 1999; Wolpert et al., 1995).

Visual attention is also greatly influenced by expectations, as testified by classic experiments. Yarbus (1967), for example, showed that a visual scene is scanned differently depending on the observer's intentions. This influence of expected stimuli for orienting attention has been reported not only in humans, but also in pigeons (Roitblat, 1980) and monkeys (Colombo and Graziano, 1994). The constructive and active aspects of perception, and in particular the top-down influences, are discussed in detail in Engel et al. (2001). On this basis, models of the visual apparatus including (hierarchical) predictions have been proposed such as *predictive coding* (Rao and Ballard, 1999) and *prospective coding* (Rainer et al., 1999).

More complex anticipatory capabilities, which are referred to as 'simulative', permit the prediction and processing of expected stimuli in advance. For example, Hesslow (2002) describes how rats are able to 'plan in simulation' and compare alternative paths in a T-maze *before acting in practice*. Simulation can also be used for the prediction of danger. Damasio (1994) argues that during decision making humans engage in 'what-if' simulated loops of interaction with the environment in order to evaluate in advance, via *somatic markers*, possible negative consequences of their actions.

Constructivists such as Piaget (1954) have argued that *sensorimotor schemas*, which enable the prediction of action effects, are progressively developed by means of an active exploration and interaction with the environment, leading to understanding and categorization. The view of situated activity as the basis of cognition, including conceptualization, has been recently revitalized and 'motor' approaches are gaining popularity. One piece of evidence that understanding comes from activity and exploration comes from an experiment performed by Held and Hein (1963): Kittens that were unable to move autonomously in the environment (i.e., those being only passively moved) failed to categorize it correctly, being, for example, unable to avoid cliffs, which shows that they did not develop appropriate depth perception. On the other hand, kittens that were raised similarly (they essentially had nearly the same perceptual input) but that had the possibility to move showed successful categorization and depth perception.

1.3.1.1 Anticipatory Human Cognition

We humans are able to perform a plethora of anticipatory mechanisms that seem to go far beyond the capabilities of other species. We are the "symbolic species" (Deacon, 1997), which was able to develop language and rather complex social structures and cultures. Some interesting examples of these capabilities include:

- We can formulate novel goals and plan in view of *future* needs (this includes abstract and distal ones such as having fun or becoming famous). The possibility to anticipate oneself could have lead to the capability to coordinate one's own actions in the present and in the future, and to have a sense of 'persisting self'.
- We can formulate expectations at an increasingly high level of abstraction and can use these to regulate our actions. For example, we can decide whether or not to apply for a job depending on our expectations about the satisfaction it will provide us, the salary, the free time, the success, etc. Not only can we formulate such abstract expectations, but we also can 'match' them with imaginary futures and select among them (albeit often only with a certain degree of success).
- We are capable of *substitution* (Piaget, 1954), that is, to manipulate mentally our representations before or instead of acting in practice. Probably several animal species are able to use their *internal models* of phenomena for making mental manipulations, but we humans are able to use that ability systematically. A mechanic can assemble and dismantle a motor in his mind before doing it in practice. An architect can propose different plans for restructuring a house. Thanks to anticipation it is possible to deal with entities also when they are not present as stimuli: an ability that is crucial for defining an agent's *autonomy* (Castelfranchi, 1995).
- We can heavily modify and adapt the environment to ourselves, not only vice-versa. While other species adjust their representations to fit the actual world, we often act in the world in order to make it fit our representations of what we want, that is, our goals. Several animal species have the capability to adapt their environments, such as building a nest, but typically they do that in a very stereotypic way. We humans do not have this limitation and have heavily modified our environment to fit our present and especially *future goals* (Gardenfors and Orvath, 2005; Pezzulo and Castelfranchi, 2007).
- We can imagine ourself in the future and reason about possible futures. Tulving (2005) has argued that the capability to engage in 'mental time travel' in the past and the future is a uniquely human capability. Although this view has been questioned, and it might be the case that this capability is also available to other animal species to a certain degree (see e.g. Hesslow, 2002), humans can use mental simulation with unchallenged flexibility. Moreover, recent neurobiological studies (see Schacter et al., 2007 for a review) indicate that the process of imagining future events involves the same brain structures that are necessary to form episodic memory traces. This suggests a novel view of memory, whose main adaptive advantage could be providing building blocks for mental simulation and not (only) remembering. This fact could explain the constructive nature of memory: what is needed to imagine the future is the capability to flexibly recombine information from the past rather than simply replaying the past. Although this view is quite novel in psychology and neuroscience, the relevance of mental simulation is highlighted by several research programs, including *prospection* (Buckner and Carroll, 2007), *episodic future thinking* (Atance and ONeill, 2001), *memory for the future* (Ingvar, 1985), and the prospective brain (Bar, 2007).

- Our highly sophisticated social life appears to rely on anticipatory capabilities as well, such as coordination and cooperation, perspective taking, imitation, theory of mind, and language (Knoblich et al., 2005; Frith and Frith, 2006; Gardenfors and Orvath, 2005; Iacoboni, 2003; Rizzolatti and Arbib, 1998).
- We have developed symbols and a symbolic language. Various researchers (Arbib, 2002; Gardenfors, 2003; Gardenfors and Orvath, 2005; Swarup and Gasser, 2007) have recently discussed how anticipation is a precursor to symbolic communication and permits the evolution of symbols, and then the development of humans as *the symbolic species* (Deacon, 1997).
- Gallese's (2001) *Shared Manifold Hypothesis* and Hurley's (2005) *Shared Circuits Hypothesis* describe how anticipation is essential for bootstrapping capabilities in the individual and especially social sphere. Both describe several evolutionary steps necessary for the development of our current cognitive and socio-cognitive capabilities.

This small list could be expanded at will and only intends to point out how ubiquitous anticipatory mechanisms appear to control and guide our behavior and cognition in general. We now proceed with considering concrete neuroscientific and psychological evidence for anticipatory behavior in animals and humans and how such behavior may come about.

1.3.2 Anticipatory Codes in the Brain

What is the neural substrate of these forms of anticipatory behavior? Is there a unique way to predict, or are there many? Are neural substrates for predictions shared among species? Are there specific brain structures that mediate complex forms of anticipation in mammals and in the human brain? Notwithstanding the fact that several aspects are still obscure, the empirical literature is huge, and cognitive psychology and neurobiology continue to unravel mechanisms and processes based on anticipation in humans and other animals. We refer to Fleischer (2007) for a recent, excellent overview of anticipatory mechanisms in the mammalian brain and to Hoffmann (2003) for an extensive survey of the psychological literature.

Here, we instead try to summarize and systematize the empirical findings of anticipations with respect to the neural codes that could be involved in generating expectations—for the sake of discussing the implications of these findings on theoretical models. We can distinguish among two main kinds of anticipatory neural codes that can perform action anticipation and goal prediction. The former focuses on associative links and the latter on generative mechanisms and internal simulation (see Csibra and Gergely, 2007 for a comparison of these mechanisms with teleological reasoning).

1.3.2.1 The Ideomotor Principle and Associative Learning

The first proposal that the brain includes neural codes that relate expectations to action was formulated in the *ideomotor principle* (Herbart, 1825; James, 1890), which has recently received a number of empirical confirmations from psychological studies (Hommel et al., 2001; Kunde et al., 2004, 2007; Prinz, 2005). It has been proposed that an agent can learn to predict the outcomes of its actions and then store ACTION → EXPECTATION (A-E) (a.k.a. ACTION → EFFECT (A-E)) associative links, which may be neurally specified thanks to a *common neural coding* between perception and action (Prinz, 1990, 2003)[1].

What is relevant here is not only that anticipation is deeply integrated with action representation but that expectations can be used for triggering action. ACTION → EXPECTATION (A-E) sensorimotor codes, once learned, can be 'inverted' and become EXPECTATION → ACTION (E-A) links, which permit the activation of an action by its (desired) effects (Hommel, 2004). The relevance of this mechanism consists in its possibility to account for goal-directed actions in a simple and elegant way, since a desired (predicted) effect, and not a stimulus, is responsible for triggering an action.

Another account of prediction based on associative mechanisms is put forward by Bar (2007). Thanks to the similarity between past and novel stimuli, analogies are established that trigger prediction on the basis of associations that capture the most frequent trends in the stimuli. This kind of predictive mechanism is then of the type STIMULUS → STIMULUS and not ACTION → EXPECTATION. These association-based predictions permit the forecasting of what is more likely to happen in the same context, to preventively set up appropriate actions, and to enable priming (perceptual, semantic and contextual, see, for example, Anderson, 1983).

1.3.2.2 Generative Mechanisms: Internal Models

It has been argued that the brain uses *internal models*, which mimic the behavior of external processes, for motor control of action (Doya, 1999; Kawato, 1999; Wolpert et al., 1995). In particular, *forward models* permit the generation of expectations about the next sensed stimuli, given the actual state and motor command. We can further distinguish between *forward sensory models* (STATE + ACTION → SENSORY FEEDBACK) and forward dynamic models (STATE + MOTOR COMMAND → FUTURE STATE).

Inverse models (or controllers) instead take as input actual stimuli and the goal state and provide as output the motor commands necessary to reach the desired state. Taken together, inverse and forward models permit not only the performance of motor plans but also the control of it and in general the regulation of its behavior in noisy and dynamic environments.

Internal models permit an agent to run an 'inner sensorimotor loop' that parallels actual sensorimotor interaction as shown in Fig. 1.1. This inner loop is extremely

---

[1] Since in principle, several effects could be associated with an action, to be efficient, this mechanism needs a guarantee that those effects stored are the only relevant ones, such as those effects originating systematically from the same action, the effects that are rewarding, etc. Since associative mechanisms do not have internal states, this may come only from simple associative forms of learning such as Hebbian learning—see, for example, Butz, 2002a and Drescher, 1991 for principled design approaches to tackle this problem.

useful for regulating motor control. For example, it can compensate for delays in sensory feedback and cancel the self-produced part of the input from sensory stimuli (Blakemore et al., 1998), etc. Empirical evidence is reported for a role of internal models in visuomotor tasks (Mehta and Schaal, 2002), eye movements (Shidara et al., 1993), imagery (Jeannerod, 1994), motor execution (Wolpert and Flanagan, 2001), and sensorimotor learning (Wolpert and Flanagan, 2003). It is worth noting that internal models could provide support for anticipation at different time scales and granularity, for which hierarchical models of action control and recognition over time have been proposed (Haruno et al., 2003; Johnson and Demiris, 2005a).

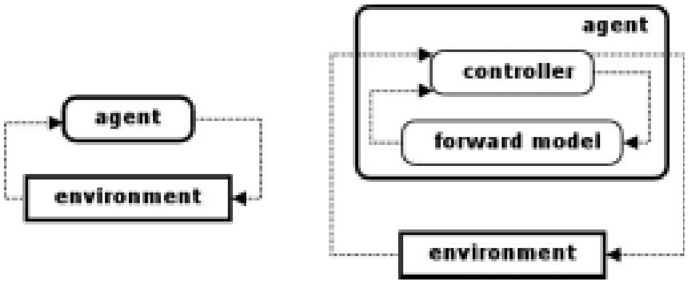

Fig. 1.1 Left: an agent engaged in sensorimotor interaction with its environment. Right: an agent running an ‘inner sensorimotor loop’ which parallels actual interaction.

One advantage of this model from the computational point of view is that it is rooted in standard control theory, and has parallels with the concepts of Kalman filtering (Kalman, 1960). Another advantage is that it provides a unitary view of anticipation in the brain, suggesting that a unique mechanism could mediate action performance, understanding, and imitation, as well as event understanding.

1.3.2.3 Generative Mechanisms: Compact Coding

Predictive coding is an account of the functional architecture of the brain that originates from Helmholtz’s models of perception. It is based on the idea that the sensory brain has a hierarchical structure that has evolved to represent or infer the causes of changes in its sensory inputs (Friston, 2005, 2003; Kilner et al., 2007). It actively does that by means of generative mechanisms that actively predict the input, bias further processing in a top-down manner, and are modulated by bottom-up feedback. This approach has the advantage of being able to integrate cybernetic models of prediction (based on empirical Bayes, Kalman filters, etc.) into a well accepted biological framework.

Besides the further reaching predictive capabilities of such models, though, computational models support this proposition, such as a bidirectional, hierarchical vision architecture, proposed by Rao and Ballard (1999). In this architecture, a higher layer includes ‘concurrent perceptual hypotheses’ that convey priors and modulate the lower layer. In turn, the lower layer sends back prediction error to higher layers as a result of a match or mismatch of perceptual actions.

Several related generative models, such as Bayesian systems and Boltzmann machines, have been used in vision, speech processing, sensorimotor integration, action execution and understanding, and decision making (cf. Dayan et al., 1995; Friston, 2005; Hinton and Dayan, 1996; Kording and Wolpert, 2006; Weber et al., 2006; Wolpert et al., 2003; Yuille and Kersten, 2006).

1.3.2.4 The Case of Mirror Neurons

Besides such strongly sensorimotor models, recent evidence suggests the presence of a *mirror neuron system* in monkeys and humans. This neural system was originally discovered in the ventral premotor cortex (F5) of macaque monkeys, where goal-oriented actions are encoded that are either performed by oneself or only visually observed while being performed by others (Rizzolatti et al., 1996; Rizzolatti and Craighero, 2004). Recently it has been shown that mirror neurons respond to action goals rather than to their surface characteristics, that is, ‘ends’ rather than ‘means’ (Umiltà et al., 2001), also the case with distal goals (Fogassi et al., 2005). This fact suggests a way to understand (and possibly imitate) not only other people’s movements, but also actions and intentions (a model along those lines was proposed in Meltzoff and Moore, 1997).

Mirror neurons show the remarkable capability to encode the prediction of the goal of an action, both performed by self and others, using a single neural circuit. This fact has suggested the possibility of breaking the boundaries between the individual and the social spheres (Gallese et al., 2004), and lead to several suggestions that the mirror system may be involved in a number of socio-cognitive functions such as action understanding, imitation, language (Iacoboni, 2003; Rizzolatti and Arbib, 1998), as well as empathy (Gallese, 2001).

For the sake of our analysis, this fact is particularly relevant since it demonstrates that the same anticipatory mechanisms could be used for goal-oriented actions as well as for action understanding and imitation, so that future-oriented and socially-oriented functions can share the same neural basis (Decety and Chaminade, 2003; Iacoboni, 2003; Jeannerod, 2001).

1.3.3 Simulative Theories of Cognition, and Their Unifying Nature

Recent research on anticipatory and in particular generative mechanisms in the brain has revitalized so-called ‘motor’ or ‘simulative’ views of cognition, which highlight the role of the motor apparatus in all aspects of cognition, ranging from situated actions to high level cognitive capabilities. Simulative theories of cognition indicate that internal mechanisms used for action monitoring and control can be re-enacted for generating long term expectations and ‘covert’ simulation of overt behavior (Cotterill, 1998; Grush, 2004; Hesslow, 2002). At the same time, other cognitive phenomena such as understanding and imitating actions performed by others, reasoning, theory of mind, and language can be accommodated within the same theoretical framework (Blakemore and Decety, 2001; Frith, 2007; Gallese, 2001; Gallese et al., 2004; Iacoboni, 2003; Jeannerod, 2001; Kilner et al., 2007; Rizzolatti et al., 2001; Wolpert et al., 2003). Anticipatory representations produced by anticipatory mechanisms can then be used in action preparation, execution, control, and mental action simulation.

Central to this family of models is the concept of internal simulation, emulation, or re-enactment. This productive aspect, which distinguishes simulative theories from similar views based on associative mechanisms, is now gaining relevance in the literature of mirror neurons (Gallese and Goldman, 1998; Oztop et al., 2006; Rizzolatti et al., 2001) and internal models (Doya, 1999; Wolpert et al., 2003).

According to (Gallese, 2000):

> To observe objects is therefore equivalent to automatically evoking the most suitable motor program required to interact with them. Looking at objects means to unconsciously 'simulate' a potential action. In other words, the object-representation is transiently integrated with the action-simulation (the ongoing simulation of the potential action).

Hesslow's (2002) simulative theory of cognition describes thinking as 'covert' behavior. In this sense, anticipatory capabilities permit the re-enactment of motor programs required for situated interaction. For this reason, there is no gap between the sensorimotor and the cognitive mechanisms that enable behavior. Hesslow (2002) suggests the following three aspects:

> (1) Simulation of actions: we can activate motor structures of the brain in a way that resembles activity during a normal action but does not cause any overt movement. (2) Simulation of perception: imagining perceiving something is essentially the same as actually perceiving it, only the perceptual activity is generated by the brain itself rather than by external stimuli. (3) Anticipation: there exist associative mechanisms that enable both behavioral and perceptual activity to elicit other perceptual activity in the sensory areas of the brain. Most importantly, a simulated action can elicit perceptual activity that resembles the activity that would have occurred if the action had actually been performed.

Related views are put forward by Grush (2004) and Barsalou (1999) under the names of 'emulation' and 'simulation' theories of cognition, respectively. These authors suggest two comprehensive attempts to integrate a plethora of cognitive functions, such as motor control, reasoning, theory of mind phenomena, and language, under the same framework that emphasizes the productive aspects of cognition. These aspects are generated by the capability of the mind to construct models of its environment that can be re-enacted and run either on-line or off-line.

1.3.3.1 Kinds of Internal Simulations and Simulative Theories

While several theories have been proposed as "simulative", the authors are referring usually to different aspects of a "simulation" and to different mechanisms for producing simulations. One important distinction is among mental simulation in the sense of off-line processing, or 'covert' behavior, as highlighted elsewhere (Blakemore and Decety, 2001; Grush, 2004; Hesslow, 2002), and mental simulation as the cognitive basis of social skills such as imitation and mind reading (Gallese and Goldman, 1998).

Similarly, Decety and Grèzes (2006) distinguish among several kinds of simulative approaches, which differ in level of access (automatic vs. conscious) and in scope (motor aspects vs. more complex cognitive states). One view is conscious reactivation of previously executed actions stored in memory (Decety and Ingvar, 1990), which can also be chained for producing long-term expectations (Cotterill, 1998; Hesslow, 2002) and 'simulate' overt behavior. Another view stresses the role of unconscious activation of several aspects of action, including its goal, the means to achieve it, and its consequences (Jeannerod, 1999, 2001): all these representations belong to the covert phase of motor preparation and can be reused for observing actions performed by others. The third view is related to the simulation-theory in philosophy of mind (Goldman, 2005) and explains the capability of understanding other's mental states—including beliefs, desires, and feelings—with the capability to "put ourselves into the other one's shoes" by simulation. It is possible that only some of these mechanisms can be used for generating long-term predictions, which can be used off-line, that is, detached from the current sensorimotor context.

Notwithstanding the differences between the approaches, and assuming that the brain could have alternative ways to simulate and emulate, this book focuses on the understanding of the unitary nature of motor and simulative theories of cognition. Simulative mechanisms, in fact, have been claimed to be involved in a plethora of individual and social cognitive functions such as perception, action performance and understanding, decision making, imitation, intentionality, etc. and have the potential to provide a unitary approach.

1.3.3.2 Action Performance, Understanding, and Imitation with a Unique Mechanism

Simulative theories of cognition, which usually stress the role of anticipatory, generative mechanisms, challenge traditional models of cognition, in which perception and action as well as the individual and social spheres are separated domains. Simulative theories provide means to integrate these domains with perception and action. This fact is extremely relevant because both it suggests a suitable engineering methodology and it has a solid biological basis.

Decety and Grèzes (1999) as well as Jeannerod (1999) show that there is a common neural substrate between action production and imagination—evidence that suggests functional equivalence. For example, mirror neurons, shared neural representations, and simulative processes have been argued to be involved in imitation (Iacoboni, 2002; Meltzoff and Decety, 2003), distinguishing the self from others (Decety and Chaminade, 2003), mind reading (Gallese and Goldman, 1998), and language production and understanding (Rizzolatti and Arbib, 1998). These facts have suggested that simulative mechanisms can explain individual and social capabilities in a unique framework. Three related problems that are inferring which action to perform for achieving a desired goal could be realized by the same mechanism. Several authors argue that generative mechanisms for controlling actions can be reenacted endogenously for perceiving, understanding, and imitating actions performed by other agents in order to understand behavior and to infer intentions from observed actions (Blakemore and Decety, 2001; Gallese, 2001; Gallese et al., 2004; Iacoboni, 2003; Jeannerod, 2001; Kilner et al., 2007; Rizzolatti et al., 2001; Wolpert et al., 2003).

According to Rizzolatti and Arbib (1998, pg. 190): "Individuals recognize actions made by others because the neural patterns elicited in their premotor areas during action observation are similar to that internally generated to produce that action". Imitation can consist in the re-enactment of the internal generative models that better fit the observed agent's goal, providing that it is in the motor repertoire of the imitating agent (Demiris and Khadhouri, 2005; Demiris, 2007; Iacoboni, 2002, 2003).

Anticipatory Mechanisms for Generating Different Kinds of Predictions Are there unitary mechanisms in the brain for producing simulations, or not? Predictions of different kinds (rewards, sensory, different state predictions) may require different mechanisms, and the same can be true for predicting one's own actions and external events. However, there could be anticipatory mechanisms that can flexibly learn to generate different kinds of predictions. One example are forward models, which have been proposed to be involved in the prediction of events, both self-generated and external. Schubotz (2007) distinguishes among prediction of events that we can or can not reproduce. Consider as an example of the first case observing the action of walking, and of performing an highly skilled action such as juggling (assuming that we are not able to juggle). In the former case we can use (re-enact) our own sensorimotor system in order to predict possible effects of other's actions. In the latter case, since our behavior repertoire does not include juggling, our internal models are only able to provide us with partial sensory information. However, as Schubotz (2007, pg. 216) claims,

> Forward models for events are not categorically different from forward models for actions. Forward models for events are just a fraction of forward models for actions, a fraction that misses the full-blown interoceptive and exteroceptive description of action models.

1.3.3.3 Anticipatory Action Control and the Sense of Agency

Anticipation must thus enable not only the simulation of individual and social spheres, but also the distinction of the two. If the same neural states are involved both in action performance and in other's action recognition—consequently being engaged in a 'we-space' (Gallese, 2001)—how do we distinguish ourselves from others?

The development of a sense of agency, which permits the understanding the self-attribution of the effects of (our own) actions, has been discussed by Piaget (1954) and Meltzoff and Moore (1997). They suggest that children learn the 'boundaries of their prediction' and thus develop a *body scheme*.

A comprehensive theoretical framework that relates anticipation, and in particular internal models, to agency has been recently proposed by Frith et al. (2000).

They discuss how the failure to access anticipatory signals that are produced during motor control (e.g., efference copies of motor commands) produces deficits in the sense of agency, and thus conclude that awareness of those anticipatory signals is essential to be able to correctly self-attribute an action or an intention. Along the same lines, Pacherie (2007) discusses the phenomenology of first person agency in terms of simpler experiences: intentional causation, the sense of initiation and the sense of control, all dependent on anticipation.

1.4 Conclusions

This introductory chapter has given an overview of different facets of anticipations and anticipatory behavior from a cognitive science perspective. We have introduced anticipation and anticipatory behavior from the theoretical point of view, we have illustrated how anticipatory mechanisms enable a range of anticipatory capabilities in natural cognition, and we have argued that anticipation is a unitary and foundational phenomenon in cognition, and ultimately that a cognitive mind should be conceived as a future-oriented device.

It has been put forward that goal-oriented systems inevitably need to have anticipatory goal representations to be able to control goals flexibly and adaptively. To learn such goal-based control structures, forward and inverse models need to be learned. Forward models allow the anticipation of a reachable goal and its activation triggering suitable motor commands with parallel inverse models. Moreover, forward models give rise to many more predictive capabilities, such as the capability of goal-inference represented in mirror neurons. The inference options strongly depend on the level of abstraction and modularity with which environmental states and circumstances are represented in the brain. The capability of long-term forward model predictions and goal-oriented behavior needs to reach a symbolic level in order to accomplish the flexibility and determination observable in humans.

Besides these strong planning capabilities, forward simulation also allows the development of self, since motor commands lead to the most reliable sensory effects. Thus, motor commands allow an accurate and reliable representation of motor-dependent forward models. In later developmental stages then, these forward models are used and mirrored to understand the agency of others and in effect, their current intentionality and even emotional state, leading to the capability of language and empathy (Arbib, 2002; Gallese, 2001; Gallese et al., 2004).

In conclusion, we propose again that a crucial challenge for cognitive systems research is to understand the passage from reactive to anticipatory natural cognitive systems, and to do the same thing in the realm of artificial cognitive systems. For the design or artificial cognitive systems, however, it does not seem to be sufficient to simply program a simulative system that uses its predictive capabilities for goal selection, imitation, motor control, reasoning, etc. Rather, the discussed different facets of anticipation need to be pinpointed and then modularly structured, as the brain does. Thus, the next chapter proposes an overarching, modular taxonomy of anticipatory mechanisms and their purpose to yield effective, flexible, and adaptive cognitive agents.