

# 12

## Representational Semiotics

---

After instructional semiotics, with this chapter we shall begin the examination of the second essential aspect of semiotic activity: The representational one [Sec. 8.1]. As in the previous chapter, I shall confine the present examination to multicellular eukaryotes, paying particular attention to animals. What we shall discover is that the best way of dealing with behavior and cognitive processes is to identify this subject as being rooted in the previous treatment of biological processes and systems. Reciprocally, it is also true that biology can no longer ignore behavior and cognitive functions due to the growing importance of field of studies dealing with the environmental impact of biological species [Sec. 10.3].

After a short introduction on the meaning of representation, we shall deal with the crucial problem of the origin of representation. Then, I shall consider the main forms of representational process. An important issue is that of connectionism and distributed networks. The classification of different types of representations (schemata and categories) as well as the issue of animal communication are also crucial. Then, we shall consider this problem from a slightly different point of view and deal with the three main systems of representation.

### 12.1 What Representation Is

Let us again consider the organism under the ontogenetic point of view. As I have said [Secs. 8.1–8.3], the true mystery of life is that living beings are able to exercise information control on the environment, and at least some of them are even able to build representations of their environment and act accordingly. I recall that *signs* are any form of being or activity in which some pattern (icon) is connected with a referent through an indexical relation. Since *representations* are a particular kind of signs, I understand by them any pattern that can be associated to an event or an object as a referent in a proper relation, such that this pattern can be said to *stand for* the referred event or object.<sup>1</sup> I recall that a *pattern* may be defined as a relationship among elements in a set such that, when an arrangement of a subset of these elements is specified, the probability of guessing the arrangement of the remainder generally increases with the size of the previous subset [Subsec. 6.3.1].

<sup>1</sup>[PEIRCE *CP*, 1.541]. It is a historical merit of the French school of Enlightenment to have understood the deep connection between representational activity and use of signs [CONDILLAC 1746, pp. 55–64] [CABANIS 1802, pp. 87–9]. In this context, I would like to stress that in attributing a semiotic activity to any biological system, I am trying to find a middle way between this school, which attributed a semiotic activity only to representational processes, and Peirce's approach [Sec. 8.1], which ontologized semiotics and applied it to the whole universe [PEIRCE 1904, pp. 303–4] [PEIRCE 1906, p. 394].

Representations, considered in their iconic aspect, can be external (like a picture of a person) or internal (like an excitation pattern of the brain). However, as far as their basic properties are considered, there is no substantial difference between these two varieties. In this part of the book, I deal mainly with internal representations that are common to all living beings with a certain complexity and showing a certain kind of behavior. Therefore, when I speak of representations without specifications, it should be understood that I am dealing with *internal* representation.

Representation's icons, consisting of specific spatial and temporal structures, are *tokens* and not types, and for this reason they are not codified information (which as such must be independent of a specific code and therefore also of any specific instantiation)<sup>2</sup>: I have indeed remarked that the brain is hardware without software [Subsec. 3.3.1]. This means, for instance, that the excitation patterns *are* the contents of our brain.<sup>3</sup> Here, we need to substitute a true biological, i.e. dynamic and population-like approach to these problems<sup>4</sup> [Chs. 8–11] to an insufficient information-processing and information acquiring understanding of cognitive processes [Sec. 6.1]. Representations acquire a schematic status of quasitypes when they are *fixed by natural selection* and become species-specific.

However, patterns (icons) are merely one side of the coin. Up till now we have not clarified what the mechanism is that enables representations to refer to external events and objects. We know that there are some who have tried to prove that representations are not necessary for living beings and that only referential-dynamic aspects need to be considered<sup>5</sup> [Subsec. 5.3.4]. This seems very implausible. However, Brook's criticism has contributed in overcoming a passive and purely associationist view of representation [Subsec. 3.8.1 and Sec. 6.1].

## 12.2 The Origin of Representation

As I have said, (internal) representational activity consists of a certain structural pattern (icon) that can be associated with an external event or object according to a certain regularity.<sup>6</sup> These patterns are not necessarily neural for elementary forms of quasirepresentation [Subsec. 11.4.5], but become neural when true representations guide an active search. Now, I shall try to explain on the most general grounds and at a rather abstract level, how quasirepresentations or representations (the difference shall be explained in due course) can be produced. I am not interested here in dealing with the evolutionary origin of representation. I am rather interested in showing how, by having very elementary components and a minimal complexity of iconic patterns at one's disposal, representations can be produced. I remark that this explanation is strongly related to the above interpretation of epigeny [Ch. 11]. It should also be noted that when I am dealing with representations, the environment is explicitly considered as a part of the whole representational system. As I shall show, representations are indeed connected with the issue of monitoring the consequences of the organism's action on the environment. Let us now collect in few points and generalize the results that have been found so far.

- (1) From the first moment, the organism can be assumed to produce a certain pattern in an autonomous manner. The way these patterns are produced is in accordance with the laws of self-organization of complex systems, and in particular with reaction diffusion mechanisms, oscillatory mechanisms, etc. [Subsec. 6.5.2, Sec. 11.3, and Appendices to Chs. 6 and 11].

<sup>2</sup>[BATESON 1968].

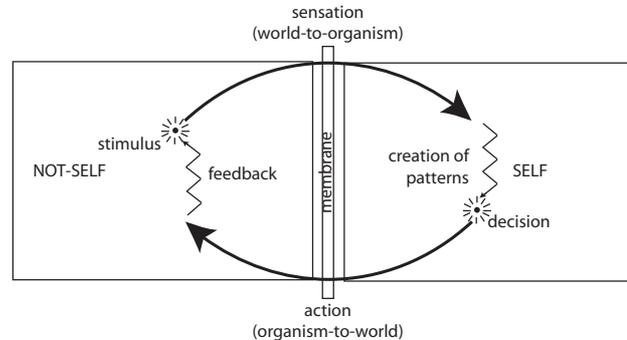
<sup>3</sup>[HOBSON 2005].

<sup>4</sup>[EDELMAN 1987, pp. 8–10 and 23–6].

<sup>5</sup>[BROOKS 1999a]. See also [VAN GELDER 1995] for discussion.

<sup>6</sup>For this reason I accept, at least to a certain extent, a fundamental tenet of associationism, like that of Hume [HUME 1739–40] or, more recently, of Fodor [FODOR 1983, FODOR 1987, FODOR 1990].

## 380 Representational Semiotics



**Fig. 12.1** Elementary semiotic organism–environment relation. In this context, see [VON UEXKÜLL 1926, pp. 125–36 and 155–77]. The pattern emerging in a wave-like manner in the organism determines a point-like decision that allows a certain action on the environment. This gives rise again to a complex process that finally produces a point-like stimulus that starts the cycle again [see also Figs. 8.4 and 8.9]. It is a true bottom-up and top-down process [YUILLE/KERSTEN 2006].

- (2) This pattern will have a bigger or smaller influence on the organism’s decider–selection system, inducing it to undertake a certain action on the external environment (we can now skip the details of this procedure [Subsec. 7.6.2 and Sec. 8.3]). This action will produce a certain effect on the environment and will therefore give rise to a feedback on the organism, which is negative to some extent [Subsecs. 4.4.5 and 8.2.1, Fig. 12.1]. In other words, there is never a complete fit between self and not-self [Subsec. 8.2.7 and Sec. 8.4]. In this sense, representations are necessarily “general” (or at least generic) and never exactly tuned to a specific situation,<sup>7</sup> while stimuli, as well as sensations, are necessarily unrepeatable selection events [Subsec. 2.3.2 and Sec. 4.1].<sup>8</sup> As we shall see, representations of “individuals” are indeed produced at a later evolutionary stage, only by high organisms. I recall that, from the “point of view” of the environment, this feedback is a pure mechanical (disruptive) action on the organism [Subsec. 2.4.1, Secs. 7.1 and 10.2]. Instead, from the point of view of the organism, the stimulus is acquired in terms of a digital information codification—something like either correct or wrong [Subsec. 3.3.1 and Sec. 8.1]. This also explains the specific way in which organisms, in their representational or perceptual activity, combine *digital* peripheral information coding and *analogical* “central” processing through complex structures and networks.
- (3) If this feedback does not have disruptive consequences on the organism and does not represent a clear inadequacy of a previous action, through teleonomic and teleologic processes mutual information between the organism and the environment will be established (or reinforced) as a consequence [Sec. 8.2 and Subsec. 10.3.3]. In other words, if the stimulus fits into a certain expected or tolerance window, it is taken at the very least as a partial confirmation of the correctness of the previous behavior. However, as any environmental stimulus in itself only

<sup>7</sup>Though with a different approach, it is a point also stressed in [KINTSCH 1998, pp. 20–1].

<sup>8</sup>Stimuli are more individual and vivid when they strongly conflict with our previous representations. The idea that stimuli are vivid individual experiences is very common in Anglo-Saxon philosophy [HUME 1777, pp. 17–18] [SPENCER 1860–2, pp. 123–32], but also [VON HELMHOLTZ 1867, pp. 610–11]. Hume calls the effect of stimuli *impressions* and calls representations *ideas*, while Spencer calls the former *vivid manifestations* and the latter *faint manifestations*. According to Spencer, the main difference is in the fact that vivid manifestations occur without previous presentation of their antecedents (unconditionally) and therefore constitute a break in the flux of representations. See also [CABANIS 1802, p. 93].

represents negative feedback, such an agreement is a *consequence* of the organism's capability to carve out the stimulus. In this way, the cycle can be closed. Any established association between referent and representational (iconic) pattern constitutes the semiotic value that this referent has from the point of view of that icon [Sec. 8.1].

In general, the organism has a certain repertoire of fixed values (which in higher organisms are represented by affordances [Subsec. 4.4.3]) and possible responses, but the partial discrepancy between expectation and result *must* nevertheless induce at least a small change in the organism [Subsec. 5.3.2] even if only in the form of a memorization of this experience. Obviously, the extent of this change can be bigger if there is a true mismatch between the organism's action according to its pattern and the environment. Now, although a change must occur as a result of this initial experience [Sec. 8.4], it is in no way a snapshot of the original event.<sup>9</sup> It is rather a modification of the organism such that when a similar event recurs, the *consequent* processing operations are interpreted both in terms of the *current event* and in terms of the *changes caused* by the first or the previous experience. This is the basis of both memory and learning, which are dynamical processes, as we shall see. Here, we have arrived at a crucial point. Even in the most elementary organism perception is already tightly connected with expectation and cannot be conceived without expectation (i.e. without the endogenous and autonomous activity of the organism through which it determines strategies for dealing with the environment).

- (4) Suppose instead that the environmental feedback is not disruptive, but nevertheless represents a rejection of the organism's current action, and therefore—indirectly—of the representational pattern at the source of that action. Now, the organism is obliged to begin a new cycle. In particular, it must begin a new search in the space of representational patterns in order to produce a fit response, in a way that is somehow reminiscent of the random search in the genetic or epigenetic landscape [Subsec. 11.7.2]. This demonstrates a very important point that I have already mentioned [Sec. 12.1]: Representations are not static images, and their purpose is not to depict something. They are essentially *dynamic responses*.
- (5) The most difficult problem is when the organism must change its representation of the *same* item in order to survive. This very often takes place. Most predators are able to camouflage themselves. If the prey survives, it must change its false representation of the predator as soon as possible. This is a true error correction. How is it possible? The weakest aspect of connectionism is the explanation of this point—recall the Humean problem [Subsec. 6.1.5]. I explained that representations are aimed at actions. Now, we have learned that animals have in general at least two independent mechanisms of perception [Secs. 4.3 and 5.1]: One for individuation and the other for identification (this is true for vision and sound reception, but can probably be generalized to any perception and to any form of multicellular organism). Individuation is strictly connected with motor activities and motion perception (in higher animals it takes milliseconds).<sup>10</sup> When individuating, the organism creates no representation of the item. On the contrary, representation is provided by the second mechanism (taking seconds). Individuating is the result of top-down causation, while identification is connected with a bottom-up cascade process.

In the same instant in which the organism receives a discontinuous stimulus (in primates it is associated with the beta rhythm [Subsec. 3.3.3]), and much more so when this is a sudden,

<sup>9</sup>[CRAIK 2002] [AULETTA 2002].

<sup>10</sup>As a matter of fact, human children are attracted by novelty and especially by moving *objects* [KAGAN 2002, pp. 51–5]. For this reason, two-year-old children's vocabulary is filled with words referring to motion and the sudden appearance and disappearance of objects.

## 382 Representational Semiotics

unforeseen event (a surprise), this stimulus then becomes the sign of something that has been *individuated* but must still be *identified*; thus, it must be a sign of a referent that will become subsequently interpreted with an associated representational icon (pattern). This is why, according to Mountcastle, sensory systems can be considered *hypothes; generators*,<sup>11</sup> or, following von Helmholtz,<sup>12</sup> I have remarked that perception in humans is connected with some form of inference and hypothesis [Subsecs. 4.4.5, 6.1.1, and 7.6.2]. This is true however for *any* organism and gives rise to forms of protoinference, as we shall see below. Such a mechanisms explains that the organism, when in motor interaction with the item, can individuate it *even with a false representation* that can later be corrected in a new representational cycle.<sup>13</sup> Therefore, under the impact of environmental feedback, the organism is in a process of continuous representation- and response-changing.<sup>14</sup> This definitively clarifies why a simple theory of the isomorphism between representation and represented object, as supported by many scientists,<sup>15</sup> cannot suffice. In the case of humans, the consequence is that a clear distinction is not made between a lower representational level and an upper knowledge level,<sup>16</sup> where symbolic structures provide new forms of access to knowledge.

Summing up, global frames (icons) in which items are embedded may function as semiotic pattern detectors, so that the perceptual knowledge in them can be used for relatively automatic pattern recognition and comprehension.<sup>17</sup> Expected objects may be identified by making use of these *global* patterns (an automatic procedure), whereas unexpected ones require more *local* analysis of the details [Subsec. 2.2.5]. As a consequence, fixation duration of the unexpected items is longer than that of expected objects. On recognition tests, human subjects noticed changes that had been made only to the unexpected objects. They did not notice when expected objects were deleted or replaced. This means that objects instantiated in the same frame are indistinguishable as long as they are represented as arguments in the frame. This also shows that the source of similarity between perceived items is indistinguishability [Sec. 4.1].

- (6) In this way, through successive accommodations, the organism eventually arrives at sharing some information with the environment. But, as explained, the immediate effect of these representational cycles is to modify the representational structure of the organism to a greater or lesser extent. Now, the crucial point is the following: If the net effect of each stimulus were only to partially modify the computational path and therefore the final response of the organism, we would again make use, though in a more sophisticated way, of the old idea that the output (the reaction) is a consequence of the external input. However, this is not the only reaction of the organism. Indeed, there is no accommodation here without assimilation, and no channelization without canalization [Sec. 8.4 and Subsec. 10.3.3]. The organism feels any new situation (the informational surprise it detects) as more or less being outside its homeorhetic path and therefore as a perturbation [Subsec. 7.6.2 and Sec. 10.2]. Therefore, according to the dynamicity principle [Subsec. 8.2.7] it also tries to *efface* the effects of a shocking stimulus and to come back to its default situation.<sup>18</sup> Here, the major role is played by the comparative and regulative aspects of the organism [Sec. 8.3]. In this way, the latter tries to make the effects of

<sup>11</sup>Quoted in [KANDEL 2006, p. 302]. Peirce spoke of a “shadowing of inferences into perceptual judgements and even into perceptions” [PEIRCE 1903c, PEIRCE 1903d]. See also [GREGORY 1998, p. 10].

<sup>12</sup>[VON HELMHOLTZ 1867, pp. 586–93 and 601–2].

<sup>13</sup>[AULETTA 2002, AULETTA 2003a, AULETTA 2003b].

<sup>14</sup>I am therefore assuming the opposite of [EDELMAN 1987, p. 60], namely that the dynamic aspect of reference is continuously updated.

<sup>15</sup>For instance, see [NEWELL 1990, p. 59]. For critical analysis see [AULETTA 2002].

<sup>16</sup>[NEWELL 1982].

<sup>17</sup>[FRIEDMAN 1979].

<sup>18</sup>[DE BIRAN 1803, p. 77].

the stimulus reversible.<sup>19</sup> This is accomplished in two ways: (a) By modifying the environment, i.e. through some external action that eventually has an indirect effect on the organism itself, as in niche construction [Sec. 10.3], or (b) by incorporating the new representation in the net of representations that have already turned out to be useful or tested [Subsecs. 5.3.2–5.3.3]. In this way, the organism explores properties of objects and of the environment, i.e. it builds finer representations that would otherwise be impossible without such an active antifeedback.<sup>20</sup> Behavior (b) is the internal counterpart of the environment’s modification (a form of internal assimilation), and its aim is to reduce the “novelty” of the new representation. Evidence of this reversible antifeedback [Subsec. 8.2.1] can be found in higher organisms, especially in their dreaming activity. Atlan<sup>21</sup> understood very well that, when dreaming, one recreates a state where any initial association that had become progressively forbidden when awake is once again allowed. In other words, both strong associations (those that have been selected during experience) and weak ones (those that have not passed the test) are here on the same footing.<sup>22</sup> As a matter of fact, dreaming is similar to madness for its disconnection with reality. It is not by chance that this activity is dependent on deactivation of the dorsolateral prefrontal cortex, which is essential for working memory, attention, and goal-directed activity,<sup>23</sup> i.e. all the cognitive aspects that are needed in animals for reference to objects and events. There has been evidence of this antifeedback during dreaming and resting in recent studies,<sup>24</sup> that show that rats’ hippocampus plays in reverse the behavioral sequence during memory consolidation. Obviously, this process can never be total, since (i) by effacing all mutual information with the environment one would never learn, and this with high nonadaptive effects, and (ii) any shock always leaves some trace: This is the price for the partial openness of the organism.

- (7) In its effort to eliminate the effects of an environmental feedback and therefore to restore its initial state, the organism integrates with more or less success the correction induced by the new stimulus into its previous representational net. In other words, we have a dynamical process of integration of two opposite forces<sup>25</sup> (accommodation and assimilation, as mentioned), whose result is the reduction of the distance not only between representation and stimulus but also between old representations and new responses. For this reason, Hebb<sup>26</sup> said that all perceiving is schema with correction. Also, according to Walter Freeman,<sup>27</sup> in higher brain activity, a change constitutes a trajectory in cortical state space, which never returns exactly to a prior state, but returns sufficiently close to the prior state so that the cortical output places a target into the same basin of attraction as did the prior output, following in this way an itinerant path [Subsecs. 8.2.7 and 10.2.3]. This dynamical understanding of representation can integrate the fundamental tenet expressed in Sec. 5.2 and Subsec. 5.3.4 without renouncing to representations. This dynamical, smoothing integration process is the biological basis of any schematization process but even of categorization as well as of intelligent behavior. Indeed, *schemata* are stored patterns which can stand for a *whole class* of stimuli.

A more complex organism, capable of responding to a wider range of environmental circumstances, will be exposed to many more errors. How will the organism react in the long (evolutionary) run?

<sup>19</sup>This is exactly the way in which Herbart understood representation, using this insight as the cornerstone of his psychology: As a resistance against external perturbations [HERBART 1816]. See also [DE BIRAN 1803, pp. 97–9]. M. Jeannerod is a follower of this tradition, whose biological foundations can be traced to [BICHAT 1800], stressing the dynamics aspect of cognition [JEANNEROD 2006]. In other words, the present book is an effort at combining the Anglo-Saxon associationist tradition and the Continental active one.

<sup>20</sup>[DE BIRAN 1803, pp. 25–34].

<sup>21</sup>[ATLAN 1972, ATLAN 1974].

<sup>22</sup>[LUR’IA 1972, p. 113].

<sup>23</sup>[HOBSON 2004].

<sup>24</sup>[FOSTER/WILSON 2006].

<sup>25</sup>[HERBART 1816] [HERBART 1824–5, Parr. 41–3].

<sup>26</sup>[HEBB 1949, p. 111].

<sup>27</sup>[FREEMAN 1995, p. 100].

**384** Representational Semiotics

Since Shannon's tenth theorem establishes that the ability of a control system to correct deviations from an optimal mode of operation cannot exceed the amount of information of the system's input (source)<sup>28</sup>, one needs to increase the mutual information between the environment and sensory channels.<sup>29</sup> Since mutual information is equal to input information minus equivocation [Subsec. 2.3.3, especially comments regarding Eq. (2.16)], then one has to increase either the reception (input) or the reliability [see also Sec. 9.9]:

- The first solution can be performed by making receptors interactive: When two channels with independent capacities less than unity, i.e. with a certain amount of equivocation, are combined by making their input sets interactive, the capacity of the resulting integrated channel is greater than the sum of the capacities of the two channels acting separately. This is constantly done in evolution by integrating both elements into a sensory system and information coming from different sensory systems. We shall find splendid examples of this, especially in mammals.
- The second possibility can be realized by tightening up the connections with the environment and also along the pathway from peripheral stimulus to CNS. This is again a matter of (internal) information control. This solution cannot be implemented too strongly since it could run against the opposite exigency of plasticity.

Therefore, the nonlinearity of representations (being complex patterns, Subsec. 6.5.1) can be considered as a consequence of the fact that at a certain evolutionary point, to have a single item standing for many stimuli, which somehow refer to or are related with a certain situation, becomes necessary. There is also an additional reason: When an organism reacts to many different stimuli with many different patterns, we have a combinatorial explosion. At an intermediate evolutionary stage, different marks may have led to the same situation, so that parallel perceptions of the same situation could be more and more integrated in a whole. It would be interesting to verify such a hypothesis by applying it to very elementary forms of animals. The final lesson is that the organism comes to share information with the environment even if there is no instruction by the latter, and the icon does not reproduce the external world in itself [Subsecs. 2.2.3 and 3.7.2].

## 12.3 Basic Issues of Representation

### 12.3.1 Main Strategies

Reaction (which is not a true representation but a quasirepresentation, even if both representations and quasirepresentations can be placed in the common category of representational activity [Sec. 8.1]) is a basic way of dealing with an unknown environment; here the initiative comes from the environment and the control is exercised only on the possible energetic sources [Subsec. 8.3.1]. Another possibility is that the organism takes the initiative and tries to control much more environmental parameters [Subsec. 8.3.2]. As a further development of this basic form of activity, there are essentially two main strategies [Sec. 10.3]:

- To *integrate* the environment in an extended and open organism or clusters of organisms. This is the strategy followed by green plants, which have transformed the physical conditions of the Earth's surface and atmosphere in terms of oxygen production (a high reactive element), in increasing the capacity of the soil to absorb water, in the ability for water retention and in

<sup>28</sup>[SHANNON 1948].

<sup>29</sup>[SAYRE 1986].

temperature smoothing. In this way, a single biosphere (a single biological system, Sec. 7.2) also integrating bacteria and animals is constituted<sup>30</sup> [Appendix to Ch. 10], and the external environment becomes a part of the metabolic network of plants.<sup>31</sup> To understand this, it is sufficient to compare the Earth with the Moon (where there is a pure physical environment without life) and Mars (a planet where a form of bacterial life probably developed but was unable to reach the level of an ecosystem).

- *Action on the environment from the “outside” by individual organisms, whose motion is the spatial expression of this activity, which requires representations in order to control the results of their own activity on the environment.*<sup>32</sup> This is the strategy followed by animals. In other words, representation is not born for dealing with external stimuli.<sup>33</sup> Already the reactive behavior and the rudimentary active behavior of unicellular eukaryotes do this (even if at a lower level). Representations deal in a controlled way with the *consequences of actions whose source is the organism itself* [Sec. 5.1]. This means that representation is tightly connected with movement and motor aspects from the beginning. Here and in the following I am mainly concerned with this second strategy.

As I have said, representations are necessary for acting on the environment and controlling the consequences of action. This explains the strict boundary that has formed between a nervous system and representations, since the nervous system was born as a control and command center for movement and action [Ch. 5]. This helps us to fix the distinction between representation and quasirepresentation as depending on whether a nervous system is involved or not. There are indeed three general conditions for true representation<sup>34</sup>:

- (1) Organisms can discriminate among several states (a minimal complexity of sensory areas),
- (2) There is a distinction between sensory neurons, motor neurons, and interneurons,
- (3) The motor system can perform more than one type of behavior.

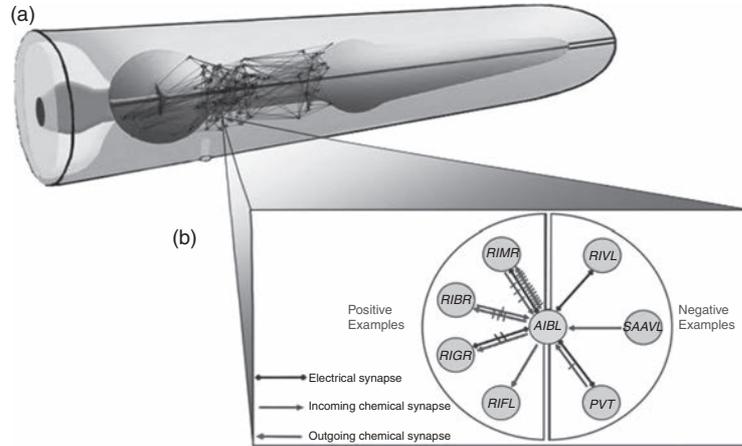
We may assume, following an insight of G. H. Parker,<sup>35</sup> that the first nervous system emerged in elementary sponge-like animals that were able to perform some action through contractions evoked by external chemical stimuli. The fact that plants display a plastic and adaptive behavior shows also that neural tissue is not needed to have an informational interplay with the environment.<sup>36</sup> An extreme example is represented by the dodder, a parasitic plant, which assesses the exploitability of a new host within an hour or two of its initial touch contact.<sup>37</sup> If this is deemed to be insufficient, the plant continues the search; but if the answer is positive, then the dodder coils about the host with a particular number of coils depending on the assessed future return. A. Trewavas hypothesizes that this plastic behavior without a brain is allowed by cellular calcium, which mediates most plant signals. Calcium waves inside cells offer computational possibilities and can, to a certain extent, replace neural nets by constituting complex patterns that somehow mimic neural ones. At the threshold between rudimentary forms of activity and a true representational activity there is the primitive creature *Ascidacea* (sea squirts), a sac-like sessile marine being that develops a nervous system (of 300 cells) for only one day during its larval stage to search for a new place.<sup>38</sup> First examples of permanent fixation of elementary neural circuits can be found in animals like the

<sup>30</sup>[VERNADSKY 1998, pp. 111–12] [LINDEMAN 1942].      <sup>31</sup>[SMIL 2002].      <sup>32</sup>[VON HELMHOLTZ 1867, p. 587].

<sup>33</sup>[MATURANA 1970, p. 13].      <sup>34</sup>[BICKERTON 1990, p. 82].      <sup>35</sup>Quoted in [SWANSON 1999].

<sup>36</sup>[SILVERTOWN/GORDON 1989].      <sup>37</sup>[TREWAVAS 2002].      <sup>38</sup>[LLINÁS 2001, pp. 15–17].

## 386 Representational Semiotics



**Fig. 12.2** (A) A standard schematic representation of a worm's head with a network depiction of a part of *C. elegans*'s neural network on the right side of the nerve ring. Neurons are in their real relative location. (B) An example of a neighborhood of one neuron. The neuron AIBL (amphid interneuron) introduces all types of combinations of synaptic relations with other neurons. For each such combination one neuron has been chosen to demonstrate it. For example the neuron RIVL (ring interneuron) is representative of the group of neurons that form only electrical synapses with AIBL. Each cross on a synapse represents one more observed additional identical synapse. The neighborhood of a neuron is defined as the group of neurons that forms a synapse with it (a chemical or electrical synapse in either direction). Neurons that are in the same neighborhood must be in spatial proximity in the worm's body. A positive example is created when a neuron "chooses" to be presynaptic to another neuron in its neighborhood and a negative example is created when a neuron "chooses" not to be presynaptic to another neuron in its neighborhood. Adapted from <http://www.wormatlas.org/>, the richest web resource on *C. elegans*.

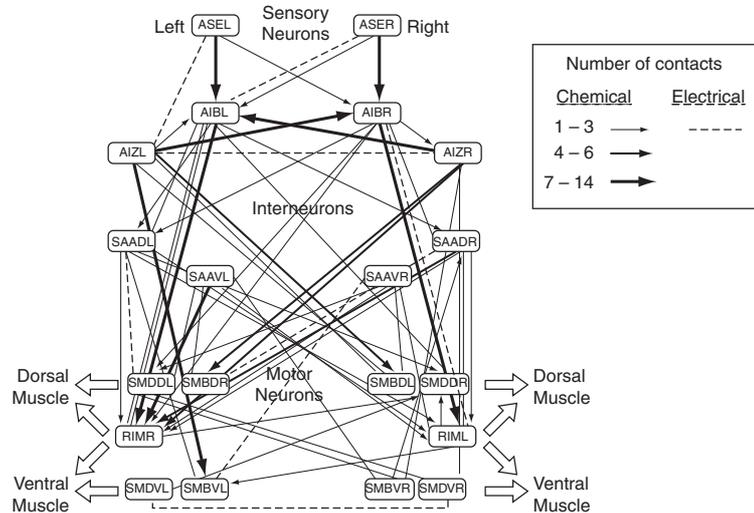
worm *Caenorhabditis elegans*, which has only 302 neurons and is sensible to several changes in its sensory surroundings<sup>39</sup> [Figs. 12.2–12.3].

The brain is the consequence of modularization when complexity grows. Recall that differentiation and integration always go together [Subsec. 2.4.4]. Indeed, the brain is a specialized organ for movement and representation but at the same time a sort of general controlling instance of the organism. The reason is that these functions are the most relevant ones for the animal's information control on the environment. This thesis is also consistent with the fact that in each integrated network there is always a hub [Subsecs. 6.3.2, 8.2.4 and 9.5.3]. We shall see that, also inside the brain, a modular differentiation process [Sec. 3.6] goes on across evolution with the emergence of a sort of brain inside the brain.

### 12.3.2 The Scope of Representation

As mentioned, true representations of external features enable the organism to control more actively environmental signals and energetic sources and to display a wide-ranging activity. For this reason,

<sup>39</sup>[WHITE *et al.* 1986].



**Fig. 12.3** Summary of the whole neural network of *C. elegans* (<http://www.csi.uoregon.edu/projects/cel-egans/talks/nips1996poster.html>). It is interesting to note that essentially we already have here the main neural divisions that we found in higher animals [Fig. 5.10]: Sensory neurons, interneurons playing the role of a regulatory network [Sec. 5.3], and motor neurons. Here, the connections go from sensory neurons through the interneurons (constituting a subnetwork) to the motor neurons. It is understood that sensory neurons are also able to catch the effects on the environment of the actions of the organism.

they require a higher structural complexity (which can be found only in pluricellular organisms with a brain or at least a developed nervous system) [Secs. 6.4-6.5]. The general law is therefore:

All representations are either nervous patterns or brain patterns of higher complexity, even if semirepresentational patterns already require some structural complexity.

When dealing with information, we also need a certain variety at the start [Subsec. 2.3.2]. According to Ashby,<sup>40</sup> information control can be obtained if the variety of the controller is at least as great as the variety of the situation to be controlled [Subsec. 11.3.1]. This is again an application of the tenth theorem of Shannon's information theory, which establishes that the ability of a control system to correct deviations from an optimal mode of operation cannot exceed the amount of potential information at the system's input<sup>41</sup> [Sec. 12.2]. This explains why representations require complex patterns in order to exercise a certain control on an environment. Obviously, the complexity of the environment can be very high, but it can still be controlled (according to a result of Bastolla and Parisi<sup>42</sup>) if complex systems like organisms organize themselves in a modular way and thus are able to partition the environment into different areas or aspects [Subsec. 6.3.2], which again demand a certain level of complexity.

This shows that not all elements of a complex network are always simultaneously active.<sup>43</sup> This is a necessary requirement for obtaining representations: If all elements were always active, we would have a single invariant representation, that is, no representation at all. In other words, we

<sup>40</sup>[ASHBY 1956].

<sup>41</sup>[SHANNON 1948].

<sup>42</sup>[BASTOLLA/PARISI 1998]. See also [RICHARDSON 2004b].

<sup>43</sup>[RICHARDSON 2004a].

**388** Representational Semiotics

need here not only a variety of patterns but also a variety of *alternative* patterns and a (nonlinear) combinatorics. It is important to understand that, since representations do not represent codified information and, at least in their iconic aspect, are therefore tokens, there is always a certain opacity in “translating” from one to another. This is also a general problem of analogue representation relative to digital codification.

We have distinguished between instructional and representational semiotics and said that both find their source in information codification, the first case in the DNA, the second case at the membrane or in the peripheral nervous system [Sec. 8.1]. Any true representation is different in this respect since it relies not only on peripheral codification but it *integrates information codification in its neural network* by the fact that each neuron is a digital device [Subsec. 3.3.2]. Instead, any epigenetic process as well as any quasirepresentational process only has codification at the source (in the genome in the former case, in the peripheral nervous system in the latter case), while the constitution of the triggered patterns obeys pure chemical and complex-like laws [Sec. 11.3]. I think that this epigenetic model also applies to the quasirepresentational activity of plants.<sup>44</sup>

**12.3.3 Representation of Time**

The most basic aspect of representations is their function of informing an organism about external space and time, which are necessary for movement. Here, let me consider in particular how a representation of time may in principle arise from pure epigenetic patterns. There are many sorts of biological clocks and biophysical oscillations [Subsec. 6.5.3] that are quasirepresentations of time<sup>45</sup>: It is a universal feature that can be found in virtually all living forms. In zebrafish, the expression of the clock gene *Period3* (*Per3*) oscillates throughout embryogenesis in the central nervous system and retina.<sup>46</sup> *Per3* rhythmic expression is free-running and is set at rest by light but not by delay caused by low temperature. Also the fertilization time has no effect on *Per3*. The circadian oscillator of some cyanobacteria, like those in eukaryotes, is connected to environmental cues.<sup>47</sup>

When speaking of a true representation of time, apart from the involvement of the nervous system, two additional issues are relevant<sup>48</sup>:

- In each living being there are several oscillations (patterns) with widely different periods, which range from fractions of seconds to years: Without this variety, time could not be represented at all.
- These forms of time representation in general are *not driven* by external rhythmic stimuli, though they respond to certain periodically recurring extrinsic events (the referents of this representation), in a way that is able to maintain a fixed phase relationship between internal and external cycles, i.e. any internal cycle is phase dependent on—but does not necessarily have the same period as—an external cycle, otherwise internal oscillations could not be representations at all [Subsec. 1.2.1].

Without such an endogenous character, the oscillations set in motion by the occurrence of one thing should be kept distinct from the oscillations set in motion by the occurrence of another thing, with the consequence that there would be an extraordinary number of event-labeled oscillations when the number of relevant items increases [Sec. 12.2]. Indeed, as I have stressed, representations are *not* reactions to a *single* environmental parameter. It is not by chance that C. Gallistel distinguishes

<sup>44</sup>And for this reason I do not think it is necessary here to speak of a neurobiology either [BRENNER *et al.* 2006], [http://en.wikipedia.org/wiki/Plant\\_neurobiology](http://en.wikipedia.org/wiki/Plant_neurobiology).

<sup>45</sup>[GLASS/MACKEY 1988] [WINFREE 1980]. <sup>46</sup>[DELAUNAY *et al.* 2000].

<sup>47</sup>[SCHMITZ *et al.* 2000]. <sup>48</sup>[GALLISTEL 1990, pp. 221–41].

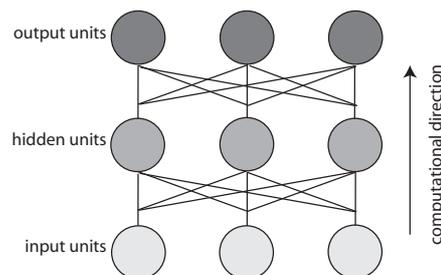
between phase-sense, which refers to the ability to anticipate events (the referents) that recur at a fixed time, for instance a particular time of day (this is common for bees), and the internal sense, which refers to the ability to respond to something that happens in a fixed amount of time, after an event that has occurred at varying points in the day–night cycle (this is common among high vertebrates).

## 12.4 Connectionism and Distributed Networks

Let us now come back to the subject of Sec. 3.8 in order to see on a rather formal plane how representational patterns can be constituted (which, being developed through the nervous system and especially the brain, have some distinctive features relative to other previously considered biological patterns).

### 12.4.1 Multilayered Networks

Early nets, like Rosenblatt’s perceptron [Fig. 3.21], were one-layered: They only had a single level of computational units. This lowered their computational efficacy a lot, as was discovered by Minsky and Papert.<sup>49</sup> Instead, neural networks that were developed later are multilayered. In this case we distinguish at the very least, input units, hidden units (which mediate the processing), and output units [Fig. 12.4]. Obviously, in the case of classical computers or early neural networks, an intermediate device is also needed to transform the input into the output [Subsec. 2.3.1]. However, as far as this happens according to a linear or Boolean transformation (following the rules of classical logic), the output is clearly a function of the input; whereas when an intermediate layer is explicitly present in the network, this is not necessarily (and in general is not) linear. We see again that in order to obtain a useful model of representation, a simple input–output model or a pure linear information transformation does not suffice, but there must be an additional level as well (here represented by the hidden units).<sup>50</sup> Indeed, Hebb<sup>51</sup> criticized the behaviorist theory



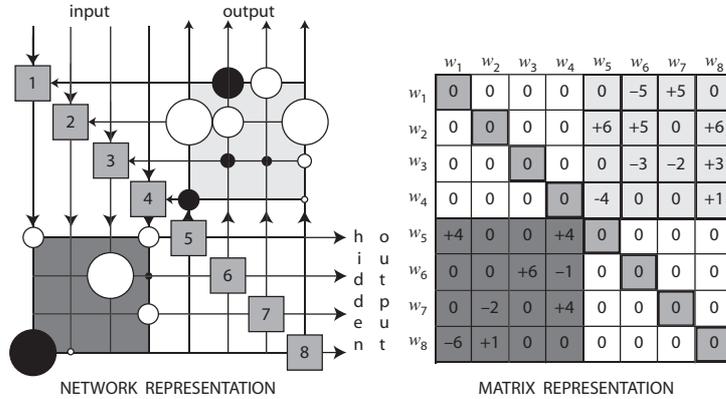
**Fig. 12.4** Here, a *three-level* feedforward network—with an input level, a middle layer (hidden units), and an output level—is shown. For the sake of simplicity each layer has three units. A purely feedforward system represents a reactive system. Nets with feedback, instead, are recurrent nets for dealing with representations.

<sup>49</sup>[MINSKY/PAPERT 1969] [RUMELHART/MCCLELLAND 1986].

<sup>50</sup>A point that has also been stressed by ANN theoreticians [AMIT 1989, p. 37].

<sup>51</sup>[HEBB 1949, pp. 171–81].

390 Representational Semiotics



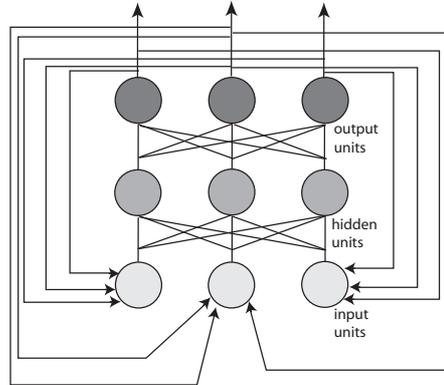
**Fig. 12.5** Units 1–4 are input units, units 5–8 are hidden ones. The computational units are constituted by open (excitatory) and filled (inhibitory) disks. The size of the disk is proportional to the strength of the operation. The matrix on the right shows explicitly the weights for the first two steps of computation. For instance, the value  $-6$  in the first column and 8th row ( $w_{81}$ , where I have dropped the superscript for the sake of simplicity) represents the strong inhibitory action of unit 1 on unit 8 and corresponds to the big black disk on the lower left corner of the net diagram, which in weight overcomes the open disk above (represented by the value  $+4$  in the 5th row and 1st column ( $w_{51}$ ) of the matrix) and the small one on its right (represented by the value  $+1$  in the 8th row and 2nd column ( $w_{82}$ ) of the matrix). The units themselves are represented by diagonal weights ( $w_{11}, w_{22}, \dots, w_{88}$ ). The region of  $\hat{W}^{(1)}$  is represented in dark gray while the region of  $\hat{W}^{(2)}$  is represented in light gray. Adapted from [RUMELHART *et al.* 1986b, p. 50].

based on the stimulus–response coupling: Between the two, there is always the filter/selection of the cellular assembly (this is perhaps even the root of the hidden unities idea). As mentioned, any *single unit* (like any neuron) is a binary information codifier [Subsecs. 3.3.1–3.3.2 and 3.8.3]—but the *whole net*, due to this nonlinear effect, does not instantiate any information codification [Subsec. 7.4.5]. This enables neural networks to be good simulations of the brain and something distinct from other patterns such as as those occurring during epigeny—as we have seen in the previous chapter—which present information codification only in the genetic *source* of the process, although epigenetic processes support cognitive processes [Subsecs. 12.3.2–12.3.3].

The input to the net, at an initial time  $t_1$ , can be represented by an activation vector

$$|a(t_1)\rangle = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ \dots \\ a_n \end{pmatrix}. \tag{12.1}$$

All the units are connected by weights  $w_{jk}$ , which are positive numbers if unit  $u_j$  excites unit  $u_k$ , and negative ones if unit  $u_j$  inhibits unit  $u_k$ . Therefore, the first-step transformation through the



**Fig. 12.6** A recurrent net. Here, the input layer is that which is controlled also by the vectors coming from the output level; these vectors contain information about the prior states of the computation.

net (from the inputs to the hidden units) can be represented by the matrix

$$\hat{W}^{(1)} = \begin{bmatrix} w_{11}^1 & w_{12}^1 & w_{13}^1 & \dots \\ w_{21}^1 & w_{22}^1 & w_{23}^1 & \dots \\ w_{31}^1 & w_{32}^1 & w_{33}^1 & \dots \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & w_{nn}^1 \end{bmatrix}. \tag{12.2}$$

This means that, at time  $t_2$ , we have the resulting hidden-unit vector [Fig. 12.5]

$$|h(t_2)\rangle = \hat{W}^{(1)} |a(t_1)\rangle, \tag{12.3}$$

or

$$\begin{pmatrix} h_1 \\ h_2 \\ h_3 \\ \dots \\ h_n \end{pmatrix} = \begin{bmatrix} w_{11}^1 & w_{12}^1 & w_{13}^1 & \dots \\ w_{21}^1 & w_{22}^1 & w_{23}^1 & \dots \\ w_{31}^1 & w_{32}^1 & w_{33}^1 & \dots \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & w_{nn}^1 \end{bmatrix} \begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ \dots \\ a_n \end{pmatrix}. \tag{12.4}$$

Then, in a second step (from hidden units to outputs), we shall have

$$\hat{W}^{(2)} |h(t_2)\rangle = |o(t_3)\rangle, \tag{12.5}$$

where  $|o(t_3)\rangle$  is the output vector, and for the sake of simplicity, we have considered an equal number of input, hidden, and output unities.

Until now, we have dealt with a feedforward net. If there is feedback (and therefore nonlinear effects with pattern formation [Subsec. 6.5.2]), we also need a rule which takes the output vector of the units and combines it with the connectivity matrix to produce a net input for each type of input, such that, in the easiest case, we have the new input given by the original input (repeated

## 392 Representational Semiotics

over and over) plus the previous output [Fig. 12.6]:

$$|a(t_4)\rangle = |o(t_3)\rangle + |a(t_3)\rangle, \quad (12.6)$$

where  $|a(t_3)\rangle = |a(t_1)\rangle$ . This feedback has already been inserted into Fig. 12.5 alongside the two pure feedforward outputs (those on the right side of the network, which are also simple-layered as in Eqs. (3.5)–(3.7), and those on the top). Indeed, the input unit 4 is excitatory on the output unit 5 ( $w_{54} = +4$ ) but the latter unit has an inhibitory influx on the former ( $w_{45} = -4$ ). This means that here we are assuming that unit 1 =  $a_1$ , 2 =  $a_2$ , 3 =  $a_3$ , 4 =  $a_4$ , 5 =  $h_1$ , 6 =  $h_2$ , 7 =  $h_3$ , 8 =  $h_4$ , so that, if we take, for the sake of simplicity, all the initial input values to be +1, then in the subsequent steps we have

$$\hat{W}^{(1)} |a(t_1)\rangle = \begin{bmatrix} +4 & 0 & 0 & +4 \\ 0 & 0 & +6 & -1 \\ 0 & -2 & 0 & +4 \\ -6 & +1 & 0 & 0 \end{bmatrix} \begin{pmatrix} +1 \\ +1 \\ +1 \\ +1 \end{pmatrix} = \begin{pmatrix} +8 \\ +5 \\ +2 \\ -5 \end{pmatrix} = |h(t_2)\rangle, \quad (12.7a)$$

$$\hat{W}^{(2)} |h(t_2)\rangle = \begin{bmatrix} 0 & -5 & +5 & 0 \\ +6 & +5 & 0 & +6 \\ 0 & -3 & -2 & +3 \\ -4 & 0 & 0 & +1 \end{bmatrix} \begin{pmatrix} +8 \\ +5 \\ +2 \\ -5 \end{pmatrix} = \begin{pmatrix} -15 \\ +43 \\ -34 \\ -37 \end{pmatrix} = |o(t_3)\rangle, \quad (12.7b)$$

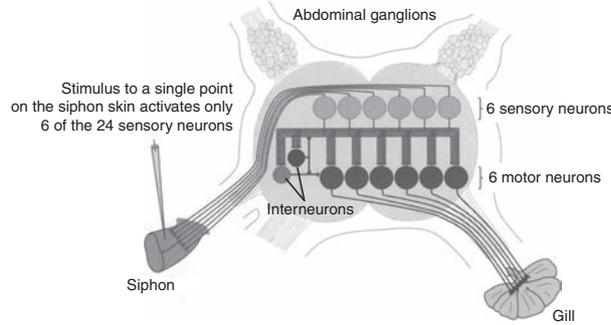
$$\hat{W}^{(1)} |a(t_4)\rangle = \begin{bmatrix} +4 & 0 & 0 & +4 \\ 0 & 0 & +6 & -1 \\ 0 & -2 & 0 & +4 \\ -6 & +1 & 0 & 0 \end{bmatrix} \begin{pmatrix} -14 \\ +44 \\ -33 \\ -36 \end{pmatrix} = \begin{pmatrix} -200 \\ -162 \\ -232 \\ +128 \end{pmatrix} = |h(t_5)\rangle, \quad (12.7c)$$

where, for the sake of simplicity, I have again considered  $a_1 = a_2 = a_3 = a_4 = +1$  for  $|a(t_3)\rangle$ . Consequently, the *whole* transformation of Fig. 12.5 for a single cycle can be written as

$$\hat{W} (|a(t_1)\rangle \otimes |h(t_2)\rangle) = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & -5 & +5 & 0 \\ 0 & 0 & 0 & 0 & +6 & +5 & 0 & +6 \\ 0 & 0 & 0 & 0 & 0 & -3 & -2 & +3 \\ 0 & 0 & 0 & 0 & -4 & 0 & 0 & +1 \\ +4 & 0 & 0 & +4 & 0 & 0 & 0 & 0 \\ 0 & 0 & +6 & -1 & 0 & 0 & 0 & 0 \\ 0 & -2 & 0 & +4 & 0 & 0 & 0 & 0 \\ -6 & +1 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{pmatrix} +1 \\ +1 \\ +1 \\ +1 \\ +8 \\ +5 \\ +2 \\ -5 \end{pmatrix} = \begin{pmatrix} -15 \\ +43 \\ -34 \\ -37 \\ +8 \\ +5 \\ +2 \\ -5 \end{pmatrix}.$$

Obviously, we can also decide to change the weights connecting the single units, in which case we need to change also the transformation matrices for different cycles of computation. A further possibility is when inputs vary, as with true environmental signals. An interesting example of this type of computation through a true neural circuitry is represented by the *Aplysia* [Fig. 12.7].

In general, the matrix  $\hat{W}$  acts on some function of both  $|a\rangle$  and  $|o\rangle$ , where I have omitted the time dependency for the sake of simplicity (it is understood that they simultaneously participate in the new input). If the vector  $|a\rangle$  represents the input activation vector and the vector  $|o\rangle$  represents the output, their scalar product yields a measure of the overlap between them and



**Fig. 12.7** The *Aplysia*'s siphon system has 24 sensory neurons, but a stimulus applied to any point of the skin activates only 6 of them. It is an example of multilayered neural computation with the activation of a subset of all units [Subsecs. 12.3.2–12.3.3]. The interneurons—the hidden units bridging between sensory and motor neurons [Subsec. 3.3.2, Sec. 5.3, Fig. 12.3]—are responsible for very slow synaptic potentials, even lasting for some minutes at a time, in the sensory neurons. Interneurons enhance the strength of the connections between the sensory and motor neurons. Adapted from [KANDEL 2006, p. 196].

therefore of their similarity [Subsec. 1.2.3]:

$$\langle a | o \rangle = (a_1 \ a_2 \ a_3 \ \dots \ a_n) \begin{pmatrix} o_1 \\ o_2 \\ o_3 \\ \dots \\ o_n \end{pmatrix}. \tag{12.8}$$

Edelman<sup>52</sup> saw in this phenomenon the basis of reentry, which is especially interesting when there is feedback into primary receptive areas that has already been elaborated information [Secs. 4.3–4.4 and Subsec. 11.5.3].

The Hebbian learning rule [Subsec. 3.8.2] tells us that, if a unit in the state  $u_j$  receives an input from another unit in the state  $u_i$ , and if both are highly active, then the weight  $w_{ji}$  should be strengthened; that is,

$$\Delta w_{ij} = f(a_j(t), \tau_j(t))g(o_i(t), w_{ji}) > 0, \tag{12.9}$$

where  $\tau_j(t)$  can be understood as a kind of teaching input to unit  $u_j$  and the change in the connection from  $u_i$  to  $u_j$  is given by the product of

- A function  $f$  of the activation  $a_j(t)$  at time  $t$  of the unit  $u_j$  and its teaching input  $\tau_j$ , and
- The function  $g$  of the output value of the unit in state  $u_i$  and the connection strength  $w_{ij}$ .

The mean error—that is, the difference between the state  $a_{xj}$  of the unit  $j$  given the input  $x$  and the desired target output  $T_{xj}$ —is mathematically expressed by the Mean Standard Error<sup>53</sup> (MSE)

$$E_x = \frac{1}{2} \sum_j (T_{xj} - a_{xj})^2. \tag{12.10}$$

<sup>52</sup>[EDELMAN 1987, p. 60].

<sup>53</sup>[ELLIS/HUMPHREYS 1999, pp. 31–5].

**394** Representational Semiotics

An alternative way of expressing this is by means of the overlap between the activation output vector and the target output vector:

$$\langle a | T \rangle. \quad (12.11)$$

Now, it may be shown that the rate of decrease of the error in the total output pattern with respect to a change in any single connection weight is

$$-\frac{dE_x}{dw_{ij}} = e_{xj} \cdot a_{xi}, \quad (12.12)$$

that is, it is equal to the error term  $e_{xj}$  for the unit  $j$  times the activation state  $a_{xi}$  of the unit  $i$  to which it is connected. In general, the system is initialized with a series of random weights, then the net is trained. For given inputs, each unit will give some (mostly incorrect) outputs (depending on the initial random weight). A supervisory system (often a human or an engine programmed by humans) takes into account the associated correct output, compares it with the actual output and calculates the error and the square of the error. The supervisor then inquires if a slight increase or decrease in the weights would reduce the MSE. Once a network has been trained, its performance is not easily destroyed by modifying or removing individual units of the network, since it shows a certain robustness of its activation patterns against random mutations of its units: The network has hubs [Subsec. 6.3.1]. Obviously, we must also take into account the hidden units, which make the calculation more cumbersome. The method here is to apply the error-correction procedure first to the output layer, then to proceed on the penultimate layer, and then to recursively go back through all the layers. In this way, we back-propagate error correction.

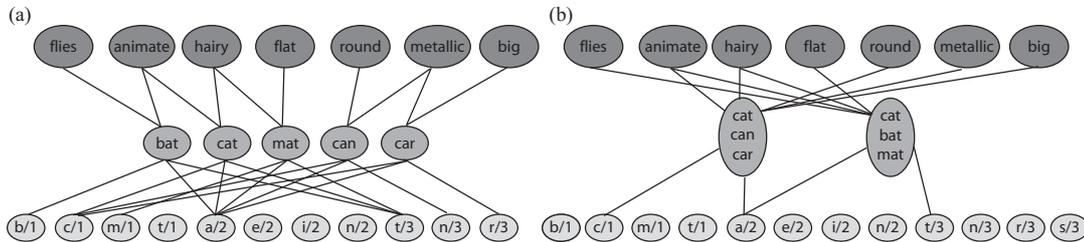
There are some similarities in learning between feedback multilayered networks like NETtalk (for simulating language learning) and humans<sup>54</sup>: (1) Learning follows a power law; (2) the more words the network learns, the better it generalizes and correctly pronounces new words; (3) the performance slowly degrades as some connections are disrupted; (4) relearning after damage is much faster than during the original training; (5) distributed or spaced practices are more effective for long-term retention than massed practices.

Multilayered networks may produce the feeling that associative learning is the only form or at least the canonical form of learning. It is not so. Moreover, we may distinguish between associative learning *strictu sensu* (when an activation pattern is produced on a set of units whenever another particular pattern occurs in another set of units) and at least a form of regularity discovery (when units learn to selectively *respond* to interesting or new patterns present in the input).

A variant of the above scheme is represented by so-called *competitive learning*. The architecture of competitive learning consists in a set of hierarchically layered units in which each layer connects (via excitatory connections) with the layer immediately above it<sup>55</sup>: Each unit of a layer receives an input from each unit in the layer immediately below and projects an output to each unit in the layer above. Within a layer the units are broken into sets of inhibitory clusters in which every element within a cluster inhibits all other elements in the cluster. The more strongly a unit responds to an incoming stimulus, the more it shuts down the other members of its cluster. Competitive learning can serve to learn categorizations of stimuli sets that are not linearly separable [see Fig. 3.22]. In this way, no supervision is necessary. The general applicability of this method alone is doubtful

<sup>54</sup>[SEJNOWSKI/ROSENBERG 1987] [OSHERSON *et al.* 1985].

<sup>55</sup>[RUMELHART/ZIPSER 1986] [ELLIS/HUMPHREYS 1999, pp. 36–7].



**Fig. 12.8** (a) A connectionist net: Here, the output words, represented for simplicity as the middle layer, are a combination of semantic (above) and phonetic (below) aspects. What is relevant here, is that each item is represented by a single icon. (b) A PDP network. Here, we have icons standing for different items to be represented [see also Subsec. 8.2.5].

however, even if it expresses important aspects of the brain’s activity at a formal level, pointed out in particular by Edelman’s neural Darwinism [Subsec. 11.5.3].

### 12.4.2 Dynamic Networks

It is useful to conceptualize a connectionist network as a constraint network in which each unit represents a *hypothesis* of some sort and in which each connection represents *constraints* among the hypotheses [Sec. 12.2]. J. Hopfield showed that it is possible to use the method of gradient descent as a way to move from a state that satisfies a few constraints to a state that satisfies more constraints.<sup>56</sup>

The measure of constraint satisfaction is given by

$$G(t) = \sum_i \sum_j w_{ij} a_i(t) a_j(t) + \sum_i \text{input}_i(t) a_i(t), \tag{12.13}$$

where  $\text{input}_i(t)$  means the sum of all net inputs to the  $i^{\text{th}}$  unit. The measure  $G(t)$  means that the overall goodness-of-fit is given by the sum of the degrees to which each pair of units contribute to the goodness plus the degree to which the units satisfy the input constraints. The contribution of a pair of units is given by the product of their activation values times the weights connecting them. If the weight is positive, each unit wants to be as active as possible. If it is negative, then at least one of the units should be 0 to maximize the pairwise goodness. If the input constraint for a given unit is positive, then its contribution to the total goodness-of-fit is maximized by being the activation of that unit towards its maximal value. If it is negative, the activation value should be decreased toward 0. When the system reaches a state of maximal goodness (a fixed or stable point), it will stay in that state and it can be said to have settled on a solution to the constraint satisfaction problem, i.e. into an appropriate representation of the input. We can consider such a state as a sort of energy minimum, in which case we take the function  $-G(t)$ .<sup>57</sup>

### 12.4.3 PDP

A traditional connectionist net using local representations employs a computing unit for each item to be represented, a sort of pre-representational or quasi-representational stage [Fig. 12.8(a)].

<sup>56</sup>[HOPFIELD 1982, HOPFIELD 1991]. See also [RUMELHART *et al.* 1986c] [ELLIS/HUMPHREYS 1999, pp. 25–31].

<sup>57</sup>I might also mention here that another interesting attempt at combining computation and dynamics comes from the self-organizing map approach [KOHONEN 1995, pp. 80, 87–95, 105–15].

**396** Representational Semiotics

Parallel Distributed Processing (PDP), instead, uses the same unit for representing several items [Fig. 12.8(b)], and only the whole patterns of activity are meaningful; in this way, representations are truly distributed.<sup>58</sup> Note also that decentralization is a common feature of both complex systems<sup>59</sup> and PDP nets. Each active unit not pertaining to the output layer represents a microfeature of an item, and the connection strengths represent plausible inferences between microfeatures. When a new item is stored, one does not need to wipe out existing patterns but only to slightly modify connection weights. This simulates very well the way in which the brain actually works, especially in its representational activity. The fact is that computation patterns do not exist anywhere unless they are actually active, according to our dynamical understanding of representation [Sec. 12.2]: No function without activated operation [Subsec. 8.2.2]. That is, we have here a strict distinction between connections (the set of constraints, the structural aspect), which can be more permanent but only potential, and patterns (the functional aspect), which are actual but transient. In other words, functions cannot be assigned to structures without the latter working in that way. This is why functions exist in organisms only as far as they are alive. So, we have an alternative definition of death [Sec. 7.1]: An organism is dead when there is the cessation of any function. This is also the reason why artifacts display functions only when they have been made [Subsec. 8.2.4] *and* are actually used by humans for that purpose.

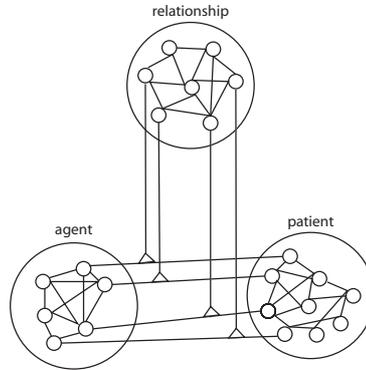
First, let us consider the efficacy of a PDP net more closely from the viewpoint of how memory works<sup>60</sup>—I shall not enter into the details of memorizing here but I shall rather use this issue as a tool for understanding the general principles of PDP computation. People can recall items from *partial* description of their contents. In distributed nets, different items correspond to different patterns of activity over the same group of units. Interactions between units allow the set of active units to influence others, thereby enabling the pattern to be completed. A new item is stored by modifying the relations between the units so as to create a new stable pattern of activity. All patterns will satisfy some inferences and violate others. A stable pattern is one that violates the plausible inferences less than any of the neighboring patterns. Thus, it is clear that there is *no sharp distinction here between genuine memory and plausible reconstruction*.<sup>61</sup> The blurring of the distinction between truthful recall and plausible reconstruction seems to be characteristic of human memory (in fact, as we shall see, memory has a reconstructive character) as well as of other high perceptual activities [Subsec. 4.4.5].

Therefore, when a new item is stored, one should not wipe out or determine conflicts with existing items [Sec. 12.2]. This can be achieved by very slightly modifying a large number of weights (the strength of the connection between units) but risking *interference* between patterns. It is possible to prevent interference by using orthogonal patterns of activity for the various items to be stored,<sup>62</sup> but this blocks the generalization processes, which is an interesting feature of PDP. For instance, by storing the item that “chimpanzees like onions,” it seems useful to represent the concept of onion and the concept of chimpanzee by alternative activity patterns over the *same* set of units. This makes it difficult to represent chimps and onions at the same time. However, we could categorize chimps as agents and onions as patients [Fig. 12.9]. If you learn that other apes and monkeys do not like onions (and therefore it is necessary to correct the previously stored representations), then one should modify the strengths of the connections emanating from all active units, so that a new knowledge about chimps will be partly a property of apes in general

<sup>58</sup>[RUMELHART *et al.* 1986b]. <sup>59</sup>[RESNICK 1994a]. <sup>60</sup>[HINTON *et al.* 1986].

<sup>61</sup>[LOFTUS 1979, LOFTUS 1997] [MILLER/GAZZANIGA 1998].

<sup>62</sup>Therefore, *interference* has a different status here relative to quantum mechanics. There, interference between orthogonal items is also possible due to nonlocal features [Subsecs. 1.2.3 and 1.2.5].



**Fig. 12.9** A network showing the relationship between agent and patient in the example that chimps like onions. Adapted from [HINTON *et al.* 1986, p. 83].

and partly a property of whatever features distinguish chimps from other apes—this strikingly recalls the Aristotelian distinction between genus and specific difference, an issue that will be the object of later investigation here.

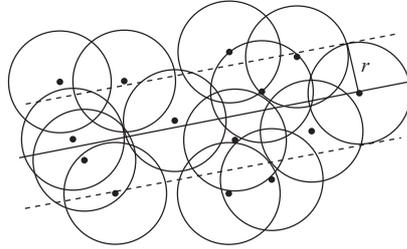
The relation between the schema or the category and one of its instances in a representational activity [Sec. 12.1] can be implemented as a relationship between a set of units and a larger set that includes it (this explains the quasitypicality status of a schema, as pointed out in Point (2) of Sec. 12.2). This kind of coarse coding is indeed a form of distributed representation—we already met coarse coding in the activity of the peripheral sensory system [Subsec. 3.3.1]. We then divide the space into larger, overlapping zones (with the same radius) and assign a unit to each zone. The encoding accuracy  $a$  is proportional to the number  $n$  of zones times their radius  $r$  [Fig. 12.10]. For a  $k$ -dimensional space the accuracy is given by  $a \propto nr^{k-1}$ . Even if each active unit is less specific in its meaning, their combination is far more specific than a pure sum. Coarse coding is only effective when the features to be encoded are relatively sparse. Indeed, if many feature-points are crowded together, each receptive field will contain many features and the activity pattern will not discriminate between many alternative combinations of feature points (we have too much interference). In this case, we would have a loss in the item's identification [Sec. 4.1].

Hopfield<sup>63</sup> successfully applied these ideas to olfactory perception (which I shall come back to below): The variable nature of turbulent air flow makes such a remote sensing problem solvable if the animal can make use of the information conveyed by the fluctuation with time of the mixture of odor sources. Behavioral evidence suggests that such analysis takes place. An adaptive network can solve the essential problem, isolating the quality and intensity of the components within a mixture of several individual unknown odor sources. The network structure then becomes an idealization of olfactory bulb circuitry: The synaptic variables themselves contain information needed by higher processing centers.

Therefore, PDP networks can account for several findings in neurology that have been pointed out by M. Mesulam and J.-P. Changeux<sup>64</sup>: Complex functions are represented by whole circuits; each individual cerebral area represents the neural substrate of several functions and behaviors

<sup>63</sup>[HOPFIELD 1991]. For the sake of simplicity, I am assuming that patterns can be represented by vectors.

<sup>64</sup>[CHANGEUX 1980] [MESULAM 1985]. See also [LIEBERMAN 1991, pp. 28–9].



**Fig. 12.10** The number of zone (set) boundaries is proportional to the number of zone centers (the units) with one-zone radius  $r$  of the line. Adapted from [HINTON *et al.* 1986, p. 92].

(and therefore pertains to several circuits): It is a network showing degeneracy [Subsec. 8.2.5 and Sec. 11.5]. Lesions confined to single regions are likely to result in multiple but partial deficits, and different aspects of the same complex function can be partially impaired through damage to a single cortical area. However, it is also true that in some cases these impairments can be partly healed by recruiting contributions from other areas.

In conclusion, it is clear that a PDP net represents a coarse graining with emerging properties that are not produced by a mere addition of the unities. Moreover, a PDP net allows for simulating interference (this is very common in human recall but also constitutes a computational problem). Additionally, since patterns and schemata are not stored as such, they can be understood as a form of potentiality stored in some arrangement of the net circuitry as well as in reinforced synapses of the brain: If there are suitable environmental conditions, then the pattern is activated and displays a function. Finally, for all these reasons, a PDP net displays a typical wave-like behavior, where there are many global features that cannot be computed by starting from local properties.<sup>65</sup>

#### 12.4.4 The Immune System

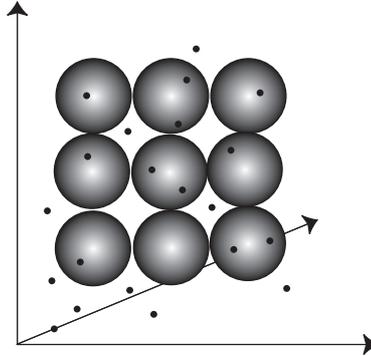
The previous examination is very interesting when dealing with a number of biological problems. I would like to show this by considering the immune system, a somatic selective system that protects the organism against aggression of extraneous bodies. This system, even if it is not a proper representational system, is able to partition the possible space of pathogen agents in subspaces. This is a particularly interesting case as it allows us to see a form of semiotic activity that is somewhere between a pure representational one and an instructional one [Secs. 8.1–8.2]. Moreover, the immune system involves many interesting genetic and epigenetic aspects and therefore also provides an interesting bridge with the biological substrate of cognition.

Antibodies are produced in a random way independently of the external pathogen agents<sup>66</sup>—another instance of generalized Darwinism [Subsec. 2.2.6]. Even if the number of possible pathogen agents is probably huge (potentially infinite), we can represent them in a multidimensional space (where the dimensions represent here some fundamental properties of the agents) in a way that is similar to the arrangement of the sensory space [Fig. 3.2], that is, we can arrange the pathogen agents according to their property-distance from one another<sup>67</sup> [Fig. 3.20]. Now, to fight external but unknown agents with success, an organism needs to produce some random antibodies, each of which will cover a portion (a hypersphere) of the space [Fig. 12.11]. In this way, with a relatively

<sup>65</sup>[MINSKY/PAPERT 1969].

<sup>66</sup>[FORD 1989].

<sup>67</sup>[PERELSON/OSTER 1979].



**Fig. 12.11** How the immune system works. The pathogen agents are represented by points in a pathogen space, whose dimensions are the different relevant properties of the agent themselves. They are arranged according to their property-distances (when spheres superpose, generalization is easier). For the sake of the representation here the space is represented as a 3D space. The antibodies are represented by spheres (hyperspheres when the dimensions are  $> 3$ ). The antibodies are produced randomly. Each one covers a relatively large portion of the space in order to fill the whole space with a relatively small number of them. Spheres may also overlap (not shown in the figure).

small number of spheres (between  $10^8$  and  $10^{15}$  for humans), it is possible to cover the whole space in an efficacious way. It is interesting to note that there is a threshold in the number of spheres, under which the antibody system is ineffective due to the loss of capability of individuating the items. However, there is also an upper limit, beyond which the detection capacity of the systems increases more and more slowly (due to interference effects affecting the discrimination capacity between different items). Obviously, the price to pay is that there is no perfect tuning of the antibody relative to the pathogen agent. However, this is also an advantage, as far as a single antibody may cover more than one agent (an aspect already considered in PDP).

Moreover, such approximation can be optimized through feedback circuits, once the antibody meets the external aggressor. To distinguish between organism and aggressor, the antibody system uses at least two different strategies.

- Antibodies are trained and selected in thymus. Here they live in promiscuity with other proteins of the body in order to become acquainted with them.<sup>68</sup>
- Each of the body's cells, when destroying old proteins, keeps some fragments—made up of exactly 9 amino-acids. Later these fragments are expelled and become attached to the cell's exterior surface.<sup>69</sup> There are around 10,000 protein fragments on every cell, representing almost every single protein that has been made by the cell. In this way, the immune system can easily acknowledge viruses and bacteria, which differ from the genetic material of the body. It is a very useful system of signalling and selecting information.

To understand the feedback mechanism here at play we need to consider that there are an adaptive immune system and a hereditary one.<sup>70</sup> The former is more efficient, and makes use of cells B and T, which are able to identify an aggressor and keep a memory of them (in this way new aggression is thwarted, a fact which is at the basis of vaccines). However, this system cannot function without the latter. In fact, the hereditary immune system produces signaling proteins, called cytokines, which

<sup>68</sup>[HEAT/SCOTT 2002].

<sup>69</sup>[RAMMENSEE 2002].

<sup>70</sup>[O'NEILL 2005]. See also [FORD 1989].

**400** Representational Semiotics

activate B and T cells. The mechanism is this: Particular receptors of the hereditary immune system, called toll-like receptors, are able to catch specific molecules of pathogen agents. Once they have detected the latter through the previously explained mechanism, the hereditary immune system provides a first unspecific attack, simultaneously signaling the aggression to B and T cells. In this way the whole immune system can subsequently adjust and plastically develop.<sup>71</sup>

Another biological example is represented by intercellular signaling [Subsec. 11.3.1], which can also be seen as an instance of PDP. Living cells respond to their environment by means of an interconnected network of receptors, second messengers, protein kinases, and other signaling molecules. Using the response of hepatocytes to glucagon as an example, Bray<sup>72</sup> showed a procedure by which a PDP network could simulate a cell signalling pathway. This procedure involves the following steps: (a) A bounded set of molecules is defined that carries the signals of interest; (b) each of these molecules is represented by a PDP-type of unit, with input and output functions and connection weights corresponding to specific biochemical parameters; (c) a “learning algorithm” is applied in which small random changes are made in the parameters of the cell signaling units and the new network is then tested by a selection procedure in favour of a specific input–output relationship. The analogy with PDP networks shows how living cells can recognize combinations of environmental influences, how cell responses can be stabilized and made resistant to damage, and how novel cell signaling pathways might appear during evolution. There is obviously also a difference between epigenetic networks and true representational ones, as already remarked [Subsec. 12.4.1]: Epigenetic cells are different and their differentiation is *crucial* for the building of the organism [Subsecs. 11.2.2 and 11.3.3]. This means that they are not particularly appropriate for being undifferentiated units distinguished only by their (active or inactive) state, which is crucial for *local* information coding, representing a binary code.

**12.4.5 Difficulties and Results**

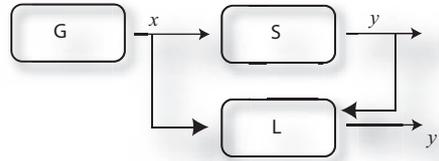
The problem of connectionism and PDP computation is the difficulty of distinguishing between general and specific information<sup>73</sup> and much more between individual token and type, which is more important by far and cannot be reduced to the former. While local classical information-processing is not able to account for emerging properties (as PDP does), it can be very good at storing and handling special items that do not fit very well in general schemata, in particular individual items [Subsec. 4.4.6]. As we shall see, this ability is fundamental for higher cognition.

Indeed, symbolic activity (which we shall deal with in the next part of the book) and distributed representations can be conceived as complementary. In fact, they respond to two completely different exigencies: *Distributed computing* is the best simulation of how assemblies of cells in our brain work for producing representation. Here, indeed, the same network can instantiate a lot of different patterns (it is degenerate). *Symbolic activity*, instead, does not explicitly use representations and is connected with information coding that is absent in a PDP network, apart from the single units themselves. Thus, the most important result of the PDP is that representation can be built without information coding at the global (network) level, but representations, as we shall see, are in turn insufficient to account for many abilities of organisms.

We have already criticized the typical “representationalism” of connectionism<sup>74</sup> [Secs. 6.1 and 12.1], and pointed out that it lacks a connection with the sensorimotor system and to the body.<sup>75</sup> This means that connectionism is not better than classical computation in explaining goal-directed

<sup>71</sup>[DI NOLA/NEUBERGER 2002]. See also [GEARHART 2002]. <sup>72</sup>[BRAY 1990].

<sup>73</sup>[MCCLELLAND/RUMELHART 1986]. <sup>74</sup>[DREYFUS/DREYFUS 1988]. <sup>75</sup>[LAKOFF 1988].



**Fig. 12.12** Statistical learning: The learner  $L$  observes couples of statistical data  $(x, y)$ , the so-called training set, provided by the information generator  $G$  and the supervisor  $S$ , respectively, and tries to guess the input (the output has a value  $y'$ ). Adapted from [VAPNIK 1998, p. 20].

behavior and especially sophisticated forms of purposive behavior like active transmission of signs as it occurs in primates. For this reason, I do not support here Frith's interpretation<sup>76</sup> according to which goal-directed behavior could be explained through the so-called temporal difference algorithm: This procedure allows a machine to find a path toward the desired goal by measuring whether any new step in a search diminishes the distance from the target or not.<sup>77</sup> This is a very interesting and helpful method for understanding how a biological system can *monitor* the chosen procedure for reaching a certain goal, but in general tells us very little about the mechanism through which both the goals and the procedure itself are selected.

Further specific problems have been pointed out by Kohonen<sup>78</sup>: Difficulty in individuating the invariant features in the percepts, difficulties in abstractive processes and in hierarchical organization of information, problems in detecting dynamical processes, difficulties in integrating different functions (motor and perceptive), impossibility of dealing with the problem of consciousness.

Another issue is that PDP networks are in most cases trained while biological systems are self-trained. The necessity of an external trainer is rooted in the very structure of *any* statistical learning. According to this discipline, learning is necessarily a three-system process, including [Fig. 12.12]: (1) A generator of the information (a source of variety), (2) a supervisor that gives additional information, and (3) the learner.<sup>79</sup>

It is also true that we may introduce unsupervised learning. Supervised (SL) and unsupervised (UL) learning accomplish different tasks: While by means of UL only a basic sorting is accomplished, SL is good for extracting a subset of basic combinations as the useful ones. In the case of UL, a good rule is: The more frequently a feature occurs in various input vectors, the more likely it is to be salient in categorizing an input as belonging to different classes.<sup>80</sup> However, this seems to show that the UL of PDP networks does not completely express the complexity of UL in true biological systems.

## 12.5 Kinds of Representation

### 12.5.1 The Indexical Side

As we have seen [Sec. 8.1], any sign expresses or is related to a function connecting an iconic and an indexical part.<sup>81</sup> At a representational level, the two aspects of a sign have very different meanings. The icon embodies the *past* experiences of the organism (or of the sign's user) and is therefore strictly associated with inborn or acquired schemata and memory. The indexical aspect of the sign

<sup>76</sup>[FRITH 2007, pp. 95–7].

<sup>77</sup>[SCHULTZ *et al.* 1997].

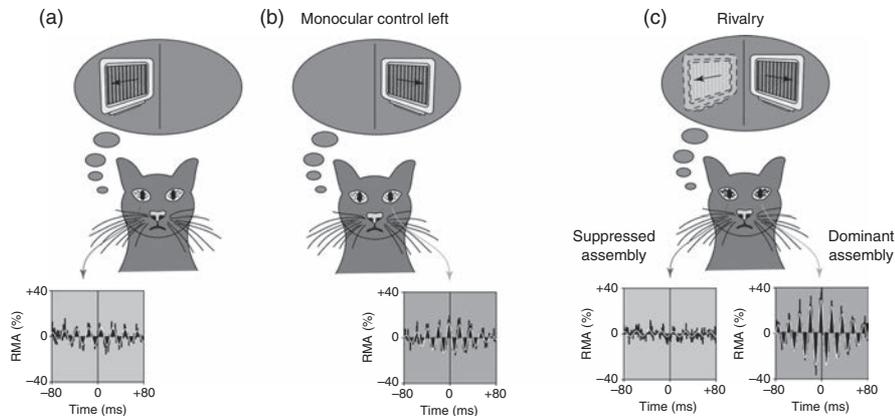
<sup>78</sup>[KOHONEN 1995, pp. 96–9].

<sup>79</sup>[VAPNIK 1998, pp. 19–21].

<sup>80</sup>[CHURCHLAND/SEJNOWSKI 1992, pp. 77–135].

<sup>81</sup>[AULETTA 2002].

## 402 Representational Semiotics



**Fig. 12.13** Synchrony under conditions of binocular rivalry. Strabismic cats were used in these experiments because, in these animals, most cortical cells can be uniquely assigned to either the left or the right eye in terms of their ocular dominance. In front of the awake animal's head, two mirrors were mounted such that each eye viewed a separate computer monitor.

(a, b) Under a monocular control condition, both pairs of cells showed synchronized activity when their preferred eye was stimulated (as shown by clear peaks in the cross-correlograms).

(c) Synchronization changed, compared with the monocular baseline, if both eyes were stimulated concurrently. Correlograms are shown for an epoch where the stimulus presented to the left eye was selected for perception. In this case, the cells driven through the left eye enhanced their correlation (dominant assembly, second and fourth plots), whereas the neurons that represent the suppressed stimulus (suppressed assembly, first and third plots) decreased their temporal coupling. When the stimulus presented to the right eye dominates perception, the strength of the correlations is reversed. The white continuous line superimposed on the correlograms represents a damped cosine function fitted to the data. RMA is the relative modulation amplitude of the center peak in the correlogram, computed as the ratio of peak amplitude over the offset of correlogram modulation. Adapted from [ENGEL/SINGER 2001].

is directed at detecting *novelty* (it is open to future experience). We therefore have a holism of the icon and an atomism of the indexical reference. Icons are built, developed, and activated bottom-up while indexicality is a top-down relation. As we have seen, the indexical relation is connected with motion perception and individuation. When a new event is perceived (generating a surprise, Subsec. 7.6.2), it can violate less essential features of a previous schema or alter essential features. In the first case, we can speak of a *discrepant event*, in the second case we have a true *novelty* (Points (3)–(4) of Sec. 12.2). Novelty is important as such for life, independently of its positive or negative import or evaluation<sup>82</sup> [Subsecs. 8.2.1, 9.5.1, and 9.5.3]: Novelty means change in the environment, an issue that cannot be underestimated by organisms.

Let us consider the neurological aspects of the indexical relation. W. Singer and colleagues found that, in perception, synchrony across recording sites is accompanied by prominent gamma oscillations [Subsec. 3.3.3], which showed the same changes under the rivalry condition [Fig. 12.13]: The power in the gamma band increased for neurons representing the dominant stimulus, although it decreased for cells responding to the suppressed pattern.<sup>83</sup> All the features pertaining to the same

<sup>82</sup>[KAGAN 2002, p. 11].

<sup>83</sup>[MUNK *et al.* 1996] [CASTELO-BRANCO *et al.* 2000]. [ENGEL/SINGER 2001].

object—color, shape, location, and the like—have the same microrate here, namely the gamma rhythm. Then the brain uses the microrate as a signature (mark) for *individuating* objects. It is very important to understand that marks are used both as labels for the *ordinary* indexical relation and as signatures for *novel* and complex structures.

To understand this, examples coming from higher animals can be very useful. Indeed, mammalian brains use two complementary strategies for representing contents: Items that are frequent, of low complexity, or of great behavioral relevance (and therefore also relatively ordinary) are represented by cells with specific response properties with a feedforward mode (therefore in forms that are more reactive and automatized). Items that are infrequent, novel, of high complexity, or too diverse to be represented by individual neurons at a primary level are encoded by dynamically associated assemblies of feature-tuned cells. These assemblies are very specific [Subsecs. 4.4.4 and 4.4.6] and in some cases can even be single neurons, as in the case of face recognition. These specific patterns differ from common (purely associative) cell populations, since they are based on the active association of neurons into functionally coherent representations and cannot be implemented in architectures that possess only feedforward connections.<sup>84</sup>

### 12.5.2 The Iconic Side

Up till now we have considered animals in general, i.e. multicellular (eukaryotic) organisms that are able to display a complex activity with a minimal brain [Subsec. 12.3.1]. In the following I shall mostly consider vertebrates. They are able to acquire at the very least representational schemata, i.e. stored patterns which can stand for a *whole class* of stimuli. Many scholars, following associationism, suppose that schemata should be constructed as a generalization of singular stimuli. This is probably not the correct answer [Sec. 12.2, Point (5)]. A very elementary (reactive) organism is less able than a higher organism to catch the specificity of a stimulus. For this reason, even if the class of stimuli to which an elementary organism can react is very small, they are not very specified but characterized by a certain amount of indetermination. We must be very careful here. Elementary representations and schemata are *not* general. As we shall see soon, *only concepts* can be general, since they can be ordered in class relations and therefore show less or greater levels of generality relative to other concepts. Schemata and elementary responses to stimuli are on the contrary *generic*. In order to understand this point, let us recall that a characterizing feature of life is the ability to partition a whole problematic space in cells [Subsecs. 12.4.3–12.4.4]. A necessity of any organism in order to informationally control the external environment—the nonself—is to subdivide this huge and unknown territory into some broad cells (food and not-food, predator–innocuous, etc.) able to guarantee the minimal activity necessary to survive [Subsec. 12.3.2]. Now, the crucial point is that, when more plastic and differentiated classes of reactions are required as evolution or development goes on, the organisms become sensible to, or interested in, a wider range of stimuli (to increase their information control), and are then able to discriminate among more stimuli and with increasing precision, irrespective of the amount of disturbance caused by the stimulus<sup>85</sup> [Subsecs. 3.3.1 and 7.6.2]. Therefore, the phylogenetic process goes from the less to the more determinate and this is the evolutionary origin of schemata, relative to pure reactions [Subsec. 8.3.1]. It is precisely due to the fact that the range of possible representations has widened that schemata are less generic than less sophisticated responses like pure reactions [Sec. 4.1]. In other words, the rule here (an instance of a principle of economy) is:

<sup>84</sup>[SINGER 2000].

<sup>85</sup>As Romanes already understood [ROMANES 1884, pp. 49 and 62].

**404** Representational Semiotics

Any representation is maintained at a certain level of genericity if the evolutionary (or ontogenetic) pressure does not force for the finding of more specific solutions.

Here, phylogeny and ontogeny are in complete correspondence and ontogenetic cognition processes go from the more general to the more specific according to von Baer's law<sup>86</sup> [Subsec. 9.5.5]. An application of the above rule is the so-called likelihood principle in vision,<sup>87</sup> according to which we assume that the form we are perceiving is the most likely, given certain environmental conditions. Another instance is our perceiving of surfaces as uniformly colored on the assumption that points moving together describe the same object [Subsec. 5.1.3]. All those are examples of a generic representation that is maintained, provided that it satisfies a minimal requirement for surviving.

Any form of dealing with an external stimulus in a representational but durable way could be cast as follows [see also Fig. 4.22, and in general Sec. 4.5]:<sup>88</sup>

(1) *Schemata*. Schemata are different from elementary reactions as far as they (i) apply to different contexts (and therefore, at least indirectly, also to tokens showing significant differences), and (ii) can represent clusters formed by information coming from different sensory channels.<sup>89</sup> However, these different situations are contextually identical (a schema for nesting is *only* a schema for nesting). Schemata are the way in which the organism tries to combine informational evidence about an item. At the beginning they were probably a combination of marks [Sec. 12.2]. Moreover, they are species-specific and therefore are triggered more or less automatically by certain environmental cues. They can be divided into<sup>90</sup>

- *Perceptual schemata*. Perceptual schemata are representations of external events. The environment changes, but perceptual schemata are mostly preserved. Shape is the most relevant feature here although olfactory or auditive components can also be very important. Concerning perceptual schemata, invariant operations are required for animals to identify objects.<sup>91</sup> The ability to schematize objects enables them to learn about their environment economically.
- *Visceral schemata* represent bodily states and are transient. They presuppose an emotional system.
- *Sensorimotor representations* are coordinated motor sequences involving representational features for guiding movement execution [Sec. 5.2]: One-year-old infants, for example, cry when the expected outcome of their action does not occur, but show interest when an expected perceptual event fails to occur (double dissociation). We must distinguish here between (a) *noninteractive* sensorimotor schemata, which are those sensory schemata that are triggered by an external event; and (b) *interactive* sensorimotor schemata, those in which the agent takes into account from the start a possible reaction of another agent as a consequence of its own action. As we shall see, only mammals have this second type of sensorimotor schemata as well as visceral schemata. This kind of interactive sensorimotor schemata are indeed rooted in the typical interactive processes characterizing mammal epigeny [Secs. 11.2–11.3].

<sup>86</sup>[GOULD 1977].

<sup>87</sup>Proposed by Hermann von Helmholtz in [VON HELMHOLTZ 1867]. See also [WILLINGHAM 2001, pp. 73–5].

<sup>88</sup>Some of the following results are due to a sort of refinement of Piaget's important work on child development. In the third part of the book I shall deal with these issues.

<sup>89</sup>[SPENCER 1855, I, pp. 166–7, 177–8]. See also [CHANGEUX 2002, pp. 58–9].

<sup>90</sup>[KAGAN 2002, pp. 26–48]. See also [HERRNSTEIN 1990b]. <sup>91</sup>[JITSUMORI/DELIUS 2001].

- (2) *Categories*. Categories<sup>92</sup> are quite different from schemata. They are not as absolute as schemata since they are connected through a web of perceptual or observable relations,<sup>93</sup> mainly of contiguity and similarity. Moreover, they are not species-specific but generic. It is most likely that only primates have categories. We may assume that categories are formed thanks to sophisticated reentrant mechanisms [Subsec. 11.5.3].
- (3) *Concepts*. The human brain makes distinctions (e.g. by different variants of a syllable) that the mind, at a higher level of abstraction, ignores (they are not linguistically relevant). While schemata are bonded with the physical properties of the objects, concepts are semantic structures incorporating hierarchical relations (it is a rule-based categorization). We will deal with concepts in the third part of the book. I remark that connectionist models are not able to distinguish between schemata, categories, and concepts<sup>94</sup> [Sec. 12.4].

Acquisition of schemata or categories is often misunderstood as acquisition of true concepts.<sup>95</sup> We can group schemata and categories under the common denominator of representations, so that we have (i) prerepresentational reactions and elementary actions, (ii) representations properly (schemata and categories), and (iii) postrepresentational structures (concepts). According to another subdivision, we can focus on the different properties of schemata, categories, and concepts, which can be summarized as follows [Fig. 12.14]:

- Concerning the issue of *specificity* (both species specificity and context specificity), schemata are both species-specific and context-specific while neither categories, nor concepts are specific in any sense.
- Concerning the issue of *relationality*, both categories and concepts are relational while schemata are not.
- Concerning the issue of *hierarchical relations* (of more or less generality), concepts are hierarchically ordered while both schemata and categories are not (they are only generic).

It is very important to realize that categories do not displace schemata (neither do schemata displace reactions), and concepts do not displace categories. Rather, each level is superimposed on the previous ones. Evidence of this is to be found in several impairments in treating information. Probably apperceptive agnosia deals with an impairment in building perceptual schemata, while associative agnosia with an impairment in conceptualization [Subsec. 4.5.2]. Impairment in perceiving faces or functional objects can be considered an impairment in categorization [Subsec. 4.5.3].

Some words about storing are necessary here. To acquire information with the use of a schema is to determine which model best fits the incoming information. Ultimately, *consistent* configurations of schemata are settled, which together offer the best response to the input. Thus schemata are kinds of *procedures*, and not things [Subsec. 12.4.3]. They emerge at the moment they are needed from the interaction of a large number of much simpler elements all working in concert with one another. Therefore, they are pattern-organizing units. Schemata are not explicit entities; they are implicit in our knowledge. They are *potential* entities, like any pattern considered as a pure structure is [Subsecs. 2.4.3 and 4.4.5]. Here, nothing which could be stored corresponds very closely to a schema. What is stored is a set of connection strengths between neurons and brain areas (as connectionism shows) [Secs. 3.8 and 12.4] which, when activated, implicitly have the ability to

<sup>92</sup>I am using this word here and in the following in a very technical sense.

<sup>93</sup>[PREMACK/PREMACK 2003, pp. 177–83]. <sup>94</sup>[KAGAN 2002, p. 64].

<sup>95</sup>[SCHRIER *et al.* 1984]. See also [SCHRIER/BRADY 1987].

## 406 Representational Semiotics

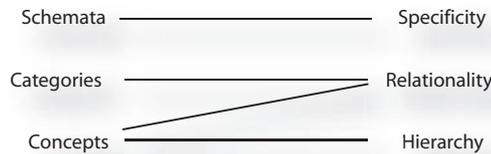


Fig. 12.14 The triad schemata–categories–concepts considered in their basic properties.

generate states that correspond to instantiated schemata. They set *a priori* the connection between units (atoms) and patterns. *A posteriori*, the interaction with a new input eventually reinforces this connection.

### 12.5.3 Affordances

All the potential uses or species-specific (survival) “values” [Sec. 12.2] of items like repair, food, and any other elements relevant for and in the environmental niche [Sec. 10.3] are called *affordances*, and represent the sensory constants of a particular species, which are the measure of the feasibility of carrying out certain actions<sup>96</sup> [Subsec. 4.4.3]. Affordances express not what is necessary but what is preferable to the organism, and therefore are always connected with some form of evaluation. Affordances are the general ways in which schemata are expressed, and at least they are mostly tightly connected with schemata. However, it is well possible, at least in elementary animals like nematodes and planarians, that we have schemata without affordance. In the next subsection, we shall also explore schemata that go beyond species-specific affordances.

When speaking of affordances, perception is considered in highly dynamic terms<sup>97</sup>: Our environment can be partitioned into substances (entities of semisolid or solid matter), media (mainly earth, water, and air), and surfaces. Surfaces are the interface between substances and media and are the place where most of the interactions between substances as well as between substances and media take place. This is very important, since we have already remarked that number internal surfaces and complexity grows during evolution and this is not immediately accessible from the exterior [Subsec. 9.5.2]. In this way, organisms protect their complexity and simultaneously are able to interact with many things without, at a first level of analysis, taking the internal complexity of objects into account. Therefore, the environment (at least from the point of view of the vertebrate organism) is not simply a collection of objects but consists of Earth-universal media like earth, water, and air as well as of all objects’ surfaces thereon or therein. In this way, any object is an ecological unity depending on the dynamical interaction between organism and environment. Indeed, surfaces and thresholds determining the boundaries of objects are dynamic and can shift, since they are continuously updated in the interaction between agent and environment.<sup>98</sup> Such a vision, therefore, helps us to escape the dichotomy between agent and object and to understand how it is possible that tools can be used by humans as extensions of their own body.

Animals’ locomotion is guided by affordances, i.e. the reciprocal relation or fit between physical properties of the actor and the environment that is required to perform a given action. In a recent

<sup>96</sup>[GIBSON 1966, GIBSON 1979]. <sup>97</sup>[GIBSON 1979, pp. 16–32, p. 66].

<sup>98</sup>[GIBSON 1979, p. 41] [MERLEAU-PONTY 1942, MERLEAU-PONTY 1945].

study,<sup>99</sup> locomotion of 14-month-old toddlers and of 8.5-month-old crawling infants was tested, and both groups overestimated their ability to ascend slopes. This is a typical species-specific distortion of perception (at least at a certain stage of postnatal development).

#### 12.5.4 Some Examples

Honeybees trained to recognize complex visual stimuli on the basis of a single feature are able to generalize their choice and apply it to novel stimuli sharing that common feature. The authors of a recent study have shown that honeybees trained with a series of complex patterns sharing a common layout that comprises four edge orientations, remember these orientations simultaneously in their appropriate positions and generalize their response to novel stimuli.<sup>100</sup> They also generalize their response to patterns with fewer correct orientations, depending on their match with the trained layout. It is surprising that the small brain of honeybees is able to perform such tasks.

As mentioned, schemata are much more relevant for vertebrates. In an enlightening study it has been shown how pigeons schematize,<sup>101</sup> for instance, discriminating between plants and trees. Pigeons, in their effort of discrimination, ignored more extended characteristics such as body configuration, and focused on specific blobs and protuberances (specific marks). The experiment also suggested that these animals perceive bits and pieces rather than whole shapes. Pigeons respond to a cluster of features that are neither sufficient nor necessary for recognizing an item [Subsec. 12.4.3]. However, they are able to catch fundamental invariances (schemata showing independence from permutation, for instance).

In the orientation-invariant task pigeons perform even better than humans (perhaps due to their panoramic view when flying).<sup>102</sup> Even humans who do not have previous knowledge of a picture, as well as apes and monkeys, have difficulties in spontaneously recognizing pictures as corresponding to real objects. For instance, monkeys succeed in transferring discrimination to new slides depicting natural objects and other animals, but not to the slides of artificial objects. Pigeons learned to discriminate pictures of trees, bodies of water, or a particular person.<sup>103</sup> Therefore, to some extent pigeons recognize consistency between real objects and their 2D representations (object–picture transfer).<sup>104</sup> Moreover, generalization to new instances of tree pictures was better than to new instances of nontree pictures.<sup>105</sup> The level of discrimination did not depend on whether trees constituted the reinforced or unreinforced category.<sup>106</sup> After being trained on an oak leaf pattern, pigeons responded to other oak leaf patterns but not to other leaf patterns.<sup>107</sup> As mentioned, this is a schematization bound to perception that is critical for surviving. Nevertheless, it is very possible that schematization does not strictly coincide with species-specific affordances.<sup>108</sup> In other words, even if bound up with vital needs, schematization abilities can have further potentialities that are manifested when applied to situations that do not have an immediate vital significance. It is interesting in this respect that pigeons do not seem to possess visual imagery abilities that allow for the performance of mental rotation of objects<sup>109</sup> [Subsec. 3.7.2].

<sup>99</sup>[ADOLPH *et al.* 1993].

<sup>100</sup>[STACH *et al.* 2004]. <sup>101</sup>[CERELLA 1986]. <sup>102</sup>[JITSUMORI/DELIUS 2001].

<sup>103</sup>[HERRNSTEIN *et al.* 1976]. <sup>104</sup>[COLE/HONIG 1994]. <sup>105</sup>[HERRNSTEIN 1979].

<sup>106</sup>As we shall see, this shows that learning (schematization) is not a form of behavior, which is instead subject to reinforcement.

<sup>107</sup>[CERELLA 1979]. <sup>108</sup>[VAUCLAIR 1996, pp. 14–16]. <sup>109</sup>[VAUCLAIR 1996, pp. 22–4].

## 408 Representational Semiotics

Parrots can also give appropriate answers (about 80 percent) to questions like: “What color is X?”, “What shape is X?”, “What object is Y?”, “What object is shape Z?”<sup>110</sup> Pepperberg’s experiments with parrots are related to vocal markers for specific colors and shapes.<sup>111</sup>

Perceptual similarities between items suffice to account for all these findings. As a matter of fact, pigeons can classify new instances of people/nonpeople slide pictures, but the success is determined by the degree of similarity and only occurs at the lower level of abstraction (which is schemata-specific and not generic).<sup>112</sup>

## 12.6 Communication and Sign

Communication consists in exchanging signals among animals.<sup>113</sup> These signals have a semiotic value but do not necessarily imply that there is some information coding or previous agreement between animals that communicate [Secs. 8.1 and 12.1]. Communication only includes phenomena that fulfill the following three conditions which satisfy the classical theory of information exchange and acquisition [Subsec. 2.3.2]:

- (1) *Inconstancy* of signals. Signal production is in fact manifested by an observable change of state of the sender. For this reason, constant states of an animal (such as skin color for many species) are not considered to be forms of communication.
- (2) As a prerequisite of communication, there is some (through teleonomic and teleologic mechanisms) structural or behavioral *coadaptation* [Subsecs. 8.2.1 and 8.2.7], a form of matching or mutual information between sender and receiver. The energy transfer mediated by signal reception must serve as a response trigger rather than as a precipitant in its own right (commanding someone to jump off a bridge is an act of communication, but pushing somebody is not) [Subsecs. 2.2.1 and 3.2.2]. In intraspecific communication, where there is a mutually beneficial flow of information, a potential two-way symmetry is required.
- (3) *Internal processing and patterning*. Animals process internal information and produce patterns. In processing signals, there are some transformation rules. In fact, communicative response to the same signal in the same context may vary at different times [Subsec. 11.3.1]. Responses are defined as any change in the probability of subsequent behavior compared to *expectations* in the absence of signaling. Here, conscious intention is not presupposed while a goal-directed behavior is [Sec. 8.2]. From the point of view of both the sender and the receiver there is a selective pressure for producing and receiving signals in an apt manner. The sender must produce signals (with a goal-directed behavior) such that they stand out from a background that can be considered as noise (environmental information that is not pertinent to the context of communication).

With regard to Point (1), this is the source-variety requirement. Point (2) leads us to an important question: Why is there so great a variety of communication? Let us take a concrete example: The variety of songs in birds. Traditional ideas about communication—as a means of transferring veridical information,<sup>114</sup> a way to manipulate others,<sup>115</sup> a means of altruistically benefiting one’s

<sup>110</sup>[PEPPERBERG 1987a, PEPPERBERG 1987b, PEPPERBERG 1991]. See also [GRIFFIN 2001, pp. 180–86].

<sup>111</sup>[PEPPERBERG 1983].

<sup>112</sup>[HERRNSTEIN *et al.* 1976, HERRNSTEIN/DE VILLIERS 1980] [CERELLA 1979] [WRIGHT *et al.* 1988] [HUBER 2001].

<sup>113</sup>On this subject see [GREEN/MARLER 1979] [SMITH 1977] [HAUSER 1996].

<sup>114</sup>[HAUSER 1996]. <sup>115</sup>[KREBS/DAWKINS 1978].

genetic relatives<sup>116</sup> [Sec. 9.6]—cannot completely account for it. It also depends on sexual selection as well as on other forms of coadaptation [Subsecs. 9.2.1 and 10.3.2], where preferences play an important role [Subsec. 12.5.3]. Variety is greatest when females combine inherited song preferences with a desire to be surprised,<sup>117</sup> an interesting high-level reversal of more basic forms of behavior [Subsec. 7.6.2]. About Point (3), I remark that signals are considered arbitrary (for instance, words) when the mapping function generating them cannot be described as a generalized transformation rule. All signals are somehow arbitrary, and only some of them become biologically constant when fixed through natural selection or ontogenetic habituation. In the presence of an unidentified signal, an animal may compare past knowledge (previous schemata) with the present situation (a new signal is different from an error, the correction of which requires a new indexical act [Subsec. 12.5.1]).

The physical signal is altered in the moment of its liberation from the source as a result of spreading and attenuation due to environmental factors [Subsec. 2.3.1], and reduction due to background noise, but later on it can also be informative about such an environment.<sup>118</sup> Then, the general context in which the signal is received must be considered. For instance, many avians and mammals emit long-distance signals at low frequencies and close-range signals at high frequencies. The perception of a signal depends also on the perceiver's ability and its current attentional status. How different must two stimuli be in order to be detected as different? This is known as the JND (just noticeable difference). Similarity is of no relevance here [Subsec. 12.5.2]: In the real world only the *difference* matters, which on the other hand already supposes a form of choice, as we shall see. I would like to propose that there is also a Just Meaningful Difference (what the system chooses to recognize as a biologically significant difference: Again we can speak here of affordances).

Zahavi<sup>119</sup> stressed that, to be effective, signals need to be reliable, and they are honest if and only if they are costly in producing and maintaining. This leads to the *handicap principle*, stressing that the effective signals appear at a glance to be deleterious for surviving<sup>120</sup>. Instead, they can be helpful even between predator and prey, when there is a common interest; for instance, when the prey wishes to convince the predator that it has apperceived the predator and that it is able to outrun it: In this way both the prey and the predator do not waste their time and energy. This is a common behavior of gazelles toward wolves. In those cases, a sort of coadaptation between different species is evident. The opposite can also happen: A predator like the tiger can show to the prey, e.g. a bull, its intention to attack in order to apperceive the bull's reaction and be able to evaluate its ability to react or fight. A similar behavior also occurs among males of the same species when they are rivals: This allows for many conflicts to be solved without having recourse to physical violence. Moreover, honest signals assume particular relevance when mate choice is at play since they are selected in such a way that they serve as marks of mate quality. Thus, these handicap signals are addressed to the selecting sex as marks of quality as the signals between prey and predator mark the quality of the prey. This obviously does not exclude cheating,<sup>121</sup> as we shall see below.

Such an interpretation may provide an alternative to the other hypotheses previously mentioned. For instance, R. Dawkins and J. Krebs<sup>122</sup> understand signals as a manipulation of behavior. They stress the view that signals do not convey information useful for a reactor (an "ethological" point of view), but they should rather aim to manipulate its behavior (a traditional "game-theory" point of view). However, even if these forms of behavior are important, it is now acknowledged, according to more recent developments of game theory, that there are also cooperative forms of

<sup>116</sup>[ACKLEY/LITTMAN 1994].<sup>117</sup>[HUSBANDS/HARVEY 1997].<sup>118</sup>[HAUSER 1996, pp. 71–109].<sup>119</sup>[ZAHAVI 1975, ZAHAVI 1993].<sup>120</sup>[ZAHAVI/ZAHAVI 1997].<sup>121</sup>[JOHNSTONE/GRAFEN 1993].<sup>122</sup>[KREBS/DAWKINS 1978].

## 410 Representational Semiotics

behavior [Sec. 9.6]. In mating, for instance, manipulation and cooperation can be integrated (as in many affective behaviors).<sup>123</sup>

A wider synthesis comes from the work of T. Guilford and M. Dawkins. Signals have two components, strategic design and efficacy<sup>124</sup>:

- *Strategic design* is concerned with how the signal is constructed to provide the information necessary to make a receiver respond, and in turn it has three main components: Honesty, manipulation, and, at least in the case of humans, mind-reading,<sup>125</sup> a concept which will be discussed in the next part of the book
- Instead, *efficacy* is concerned with how a signal is designed to get that information across to the receiver. Dishonest signals tend to be discarded.

An evolutionary force is therefore represented by the psychology of the receiver, at least in the case of mammals. This means that what is easy to detect, to discriminate, and to recall, plays an important role in the evolution of communication.

## 12.7 Representational Systems

The following examination stems ultimately from Peirce's theory of signs, which is strongly influenced Colin Cherry<sup>126</sup> [Fig. 8.1], who in turn opened the way to the pathbreaking studies of the ethologist Peter Marler in this field. Marler was himself fully aware of Peirce's legacy.<sup>127</sup> In a joint work with S. Green he distinguished between different types of semiosis.<sup>128</sup> I will partly modify and develop this scheme, that has its source in empirical studies.<sup>129</sup> I believe that all representational processes can be assigned to the following categories:

- (1) *Reference* (Marler and Green's indexical reference that is not an active production of a sign), which is basic to all representational semiosis.<sup>130</sup>
- (2) *Addressing* (deictic semiosis in Marler and Green's language, that also covers other contextual forms of communication), which is basic to all forms of communication.
- (3) *Active transmission* of a sign as a sign of something to a partner that is able to understand it as a sign of this something.

The former aspect is probably found in all animals, while the latter two might be evolutionarily derived from the former. We shall see that reference covers perceptual schemata, addressing corresponds to visceral schemata and sensorimotor representations, and active transmission to categories [Subsec. 12.5.2].

### 12.7.1 The First-Order Representational System

The first order of representational activity (reference) can be synthesized as follows: An agent interprets something as a sign of a referent.<sup>131</sup> The triangle<sup>132</sup> shown in Fig. 12.15 goes from the

<sup>123</sup>[HINDE 1981].      <sup>124</sup>[GUILFORD/DAWKINS 1991].

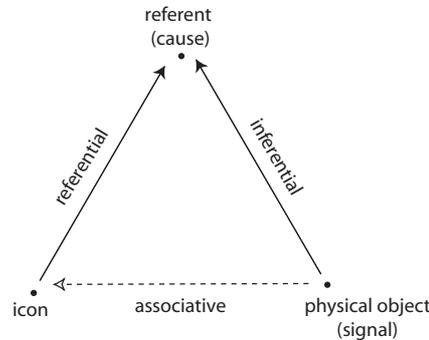
<sup>125</sup>[PREMACK/WOODRUFF 1978a, PREMACK/WOODRUFF 1978b].      <sup>126</sup>[CHERRY 1957].

<sup>127</sup>[MARLER 1961].      <sup>128</sup>[GREEN/MARLER 1979] [MARLER 1992].

<sup>129</sup>[AULETTA 2005c, AULETTA 2007]. In these papers I still supported a relatively incomplete and immature point of view.

<sup>130</sup>Deacon calls this *indexical semiotics* [DEACON 1997, pp. 69–101].      <sup>131</sup>[ALLEN/SAIDEL 1998].

<sup>132</sup>As I have mentioned, for the first time a similar triangle was proposed by Cherry [CHERRY 1957]. See also [MORRIS 1938].



**Fig. 12.15** The three relations of the first-order representational system are shown. An object becomes a sign through association with a representation having an iconic function which is referred to a possible cause of the physical object. The associative relation is a dashed line.

physical object to the icon or the representational schema (the object has then the value of a signal in its property of evoking a certain schema), from the evoked representation to the referent (*qua* referent), and from the physical object to the referent (not as referent but rather as an inferred possible cause of the physical object's presence). It is very important that we are not speaking here of an inference in the proper sense of a logical operation (which is typically symbolic). If there is a proper association and a proper referential relation, the net result is that of a sort of proto-inference. In a paradigmatic case, an animal sees a track (the physical object) and takes it as a sign of possible prey (the referent) of which it has a given representation (the icon). It is very important to stress that the iconic aspect is a schema mostly encompassing some affordances [Subsec. 12.5.3]: The prey is food, the tree can be a sanctuary, and so on.

Let us have a closer look at these relations:

- The relation between physical object and icon is purely associative: The associative relation is justified (at least to a certain extent) by the phylogeny or ontogeny of the animal itself. In the latter case, it has become a habit [Subsec. 12.5.1]. Without at least a habit one could not speak of an association at all. In the case of a phylogenetic species-specific adaptation we speak of an instinct.
- Only indexical relations (as the referential relation shown in the previous figure) can be arbitrary, in the sense that they are “chosen” and conventional: The act through which the animal uses its stored representation (its schema) of a possible (future) prey as an index pointing to the individual cause of the track (as physical object) is purely arbitrary, since there is no justification for this apart from the fact that the animal has actually taken the track as a sign of the possible prey: It is a sort of event-like brute fact [Subsecs. 2.2.3 and 2.2.5].
- The physical object is related to the referent only *through the icon*, and it is only in this way that it acquires the semiotic function of being a sign. Therefore, it is the icon that is directly related to the referent, while the relation between physical object and referent is only indirect and inferred from the perspective of a certain icon and a certain referent. For this reason, the indirect relation that is established between object and referent can be erroneous or misleading. An instance is represented by an animal erroneously following the track of a wolf instead of a lamb. In this case, the animal has used an icon as not standing for the right thing. Such

**412** Representational Semiotics

a relation, being subject to error,<sup>133</sup> is also subject to deceptive tactics. For instance, many animals (and even flowers) can assume colors, postures, or forms (which can be considered as “tracks”) such that they can easily be taken to be something different from what they actually are.

There are reasons to believe that the first representational system can be found in any animal showing a minimal complexity (pattern production), at least in rudimentary form can be found in social insects, and as well-developed schemata in all vertebrates.

There are several examples of deductive proto-inferences made by animals like birds. For instance, pigeon jays use transitive inferences in order to predict social dominance. In some experiments, they were tested by making use of a transparent chamber separated into three compartments.<sup>134</sup> In the central compartment there is an unshelled peanut and in the opposite chambers there are two birds which compete for the food. A third bird is outside the chamber and can observe the behavior of these two animals. Finally the third bird is tested itself in the chamber with one pigeon of the previous pair. This ability should not be overestimated. There have also been limits found on the pigeon’s ability to deduce<sup>135</sup>: In fact, they are not able to understand a series of items (for instance, colors) but rather follow very practical rules about specific tasks (like “when choose yellow, choose blue thereafter”). This is the reason why we cannot speak here of inferences in the true sense of the word but rather use the concept of “proto-inference”, which only supposes an indirect associative link between items.<sup>136</sup>

In another experiment, five ravens “fished” for some food suspended on strings. When a raven is startled while holding a small piece of meat in its bill, it typically flies off without dropping the food; the four ravens which succeeded apparently realized that the string would prevent the meat from being carried away, since they dropped the meat before flying.<sup>137</sup>

We have already considered some time “representation” of elementary organisms [Subsec. 12.3.3] and some further examples of schemata [Subsec. 12.5.4]. In the case of space’s representation [see also Subsec. 5.3.4], Gallistel<sup>138</sup> adduces neurological evidence that animals (obviously, plants are not concerned here) may represent their position in the environment regardless of how they are oriented, and, at the opposite, may also represent a particular direction regardless of where they are. Abstractly speaking, there are three ways to navigate in a given environment:

- By dead reckoning or path integration, i.e. by updating the estimate of one’s own position on the basis of knowledge of how fast one has been moving, in what direction, and for how long; and
- By piloting by means of a reference to a centrally-synthesized map [Subsec. 3.7.1] and the observation of points that are represented on the map (landmarks), which turn out to be the referents of this representation<sup>139</sup>;
- By making use of landmarks only, without maps. This is only, possible in a relatively limited and familiar environment. Another possibility is to use gradients like magnetic fields.<sup>140</sup>

<sup>133</sup>I thus support the idea that representation always presupposes the possibility of misrepresentation [DRETSKE 1988, pp. 64–70].

<sup>134</sup>[PAZ-Y-MIÑO *et al.* 2004].

<sup>135</sup>[TERRACE 1993] [VAUCLAIR 1996, pp. 16–18] [WYNNE 2004, pp. 73–4].

<sup>136</sup>The so-called conditional-discrimination tests (when a discrimination is conditional upon another one) will be very useful for distinguishing between kinds of proto-inferences [REZNIKOVA 2007, pp. 146–51 and 170–1].

<sup>137</sup>[GRIFFIN 2001, pp. 116–18].

<sup>138</sup>[GALLISTEL 1989, GALLISTEL 1990].

<sup>139</sup>[HAFTING *et al.* 2005].

<sup>140</sup>[REZNIKOVA 2007, pp. 109–11].

Let us first consider *piloting*. There are several sorts of maps (a kind of icon). Most animals using piloting take advantage of Euclidean maps (that preserve distance and angle). Moreover, there is evidence that many animals in general use metric maps, i.e. maps that are sensitive to the compression of geometrical shapes. A metric map is therefore a stronger mapping than affine maps, which are the product of a transformation between two vector spaces (strictly speaking, two affine spaces) consisting of a linear transformation followed by a translation. Among the affine maps, there are projective maps (that preserve only the relative position of the points). In fact, bees are sensitive to rectangular compressions of an original square and the heavier the deformation is the more sensitive they are. This does not imply, however, that bees use maps. Since animals are sensible to the geometrical structure of the environment without taking into account other features, the system for establishing position and heading may be understood as a module. Gallistel understands computation as the process that maps one representation or set of representations into another. To a certain extent this is a good definition, but only when computing representations, that is, when the latter are not considered in their function of representing something else, and are considered rather as *endogenous* patterns (in their structural properties), that is, as icons.

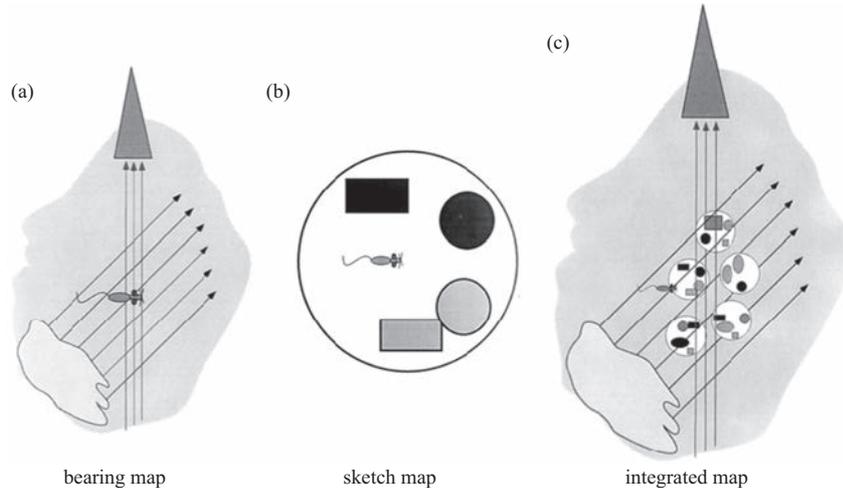
*Dead reckoning* can make use of very different resources<sup>141</sup> and could be performed by using the equivalent of what we consider as either Cartesian or polar coordinates. Dead reckoning by using Cartesian coordinates is performed by considering first the solar heading (i.e. the angle of the lubber line measured) from the actual position, clockwise from the sun, decomposing in the sine and cosine components, and combining this with the calculation of sine and cosine of solar azimuth (ephemeris angle); in this way, the sine and cosine of the angle between the compass heading and the actual direction are obtained. Finally, after some calculations that give the sine and cosine of the compass bearing of the nest, one obtains the distance from the nest (the fixed reference) and the angle representing the bearing of the nest and whose sine and cosine give the position of the nest in a Cartesian coordinates frame (again a form of proto-inference). Polar methods are less satisfactory because they are based on feedback loops in which the bearing of the nest and the distance already enter as inputs. As such, evolution should have preferred the first methodology. It is experimentally proved that ants and bees always maintain a constant angle relative to the sun and that they can extrapolate the changes of the sun's position when it is out of sight.

There is also an important interplay between dead reckoning and piloting: In fact, maps can also be constructed by using dead reckoning. A map is built by combining the metric representation in egocentric coordinates of the relative position with dead reckoning that provides a representation in geocentric coordinates of the vantage points and the angles of view (headings) [Sec. 5.3]. By calculating the actual position (in Cartesian coordinates) relative to the hive (positioned at the origin), by considering a point's position relative to the actual position, and taking into account the heading of the animal, it is possible to obtain the position of the point in Cartesian coordinates. The main task is then to calculate the angle and distance from a given source (reference point). A map is very important in order to find the locations of several points when coming from different positions.

Bees and ants also use *landmarks only* and topographic or sketch maps built using positional information or at most bearing maps [Fig. 5.2], i.e. a navigational system based on goal localization

<sup>141</sup>[REZNIKOVA 2007, pp. 108–9].

## 414 Representational Semiotics



**Fig. 12.16** How cognitive maps are built in mammals. They are constructed from both (a) non-cognitive, pure information-processing maps (actually they are not maps in the true sense of the word), like bearing maps, thanks to crossmodal perception; and (b) a further elaboration of centrally-synthesized topographic maps [Fig. 5.2] resulting in sketch maps [Subsecs. 3.7.1 and 4.4.2, Sec. 4.3]; (c) the final cognitive (centrally synthesized) map is then the integration of a bearing map with many sketch maps. Adapted from [JACOBS/SCHENK 2003].

and some environmental factors,<sup>142</sup> by individuation of landmarks surrounding the goal (when coming back they continuously compare a remembered snapshot of the landmark with their current visual image and try to reduce the discrepancy between the two images). We may further assume that insects (as well as other animals) make use of very different cues for their navigation and are even able to combine them.<sup>143</sup>

Two important points are whether bees make use of cognitive (integrated) maps [Fig. 12.16] and of a language when communicating the results of their explorations to their nestmates. If this were true, they would show levels of representation that are very high, since they should somehow be able to organize many elements in a model of the world and give rise to some active production of signs<sup>144</sup>:

- With regard to the first issue, we have seen<sup>145</sup> that bees also use specific landmarks. Indeed, when these are not visible, bees get confused. As I have stressed, cognitive map-building cannot be supported by dead reckoning and landmark or bearing-map navigation.<sup>146</sup> A general result today widely acknowledged is that only birds and mammals are able to build cognitive maps.<sup>147</sup>
- With regard to the second issue, the problem is that the so-called bees' language is not open-ended, nor allows for the recombining of elements, but is strictly limited to providing information about sources of nectars and possible new homes: Bees cannot adapt their communication system to saying anything else. It is likely that pure olfactory aspects (and not genuine

<sup>142</sup>[WEHNER 1989]. See also [WEHNER 1981, WEHNER 1983] [COUVILLON *et al.* 1991].

<sup>143</sup>[REZNIKOVA 2007, pp. 111–14]. <sup>144</sup>[WYNNE 2004, pp. 13–45]. <sup>145</sup>[DYER 1991].

<sup>146</sup>[MENZEL *et al.* 1990]. <sup>147</sup>[JACOBS/SCHENK 2003]. However, see [REZNIKOVA 2007, pp. 128–31].

communicative ones) also play a role here.<sup>148</sup> Similar considerations should be true for the astonishing pheromone ants' language.<sup>149</sup>

This does not mean that bees are not able to learn and recall things like shapes and scenes related to the two items above. However, as we shall see, learning can be, at a first and elementary stage, purely associative (even if still semiotic in nature). This has been stressed recently by studying the issue of hive bees reproducing the way bumblebees extract nectar from flowers without pollinating them.<sup>150</sup> I recall that Darwin and Romanes still believed in a sort of direct imitation,<sup>151</sup> a tangled issue that will occupy us later. Summing up, we may assume that these insects, from the point of view of the active production of signs, are pure machines that acquire and process information almost mechanically. Instead, from the point of view of making reference to landscape signs and to other bees dancing, they are first-order representational systems.

There are interesting examples of deception already at this level.<sup>152</sup> Ants can be deceived and actually use deception tactics.<sup>153</sup> A queen of the genus *Polyergus* can deceive workers of the genus *Formica formica* and be accepted as their queen by releasing specific chemicals emitted by the proper queen (which has previously been killed). As already mentioned, this is understandable according to the first-order representational system. Indeed, the usurper may simply follow pure Darwinian, genetically hardwired mechanisms, while the workers show an ability to semiotically react to specific pheromones. We see here a very common phenomenon:

Lower forms of semiotic systems can to a certain extent “mimic” higher forms.

This can give the feeling that we are dealing with a much more complex phenomenon. As such it is better to apply a canon of parsimony and explain phenomena with causes that are as low in the psychological scale as possible.<sup>154</sup> However, this circumstance must not be easily dismissed, because, at least in certain contexts [Subsecs. 8.2.7, 9.5.1, and 9.5.3], it allows one to deal with situations where the lower-level semiotic processes only work imperfectly or partially.

The first-order representational system (reference) is so universal that it is not confined to a specific representational process: It corresponds to what I have called above *perceptual schemata* [Subsec. 12.5.2]. Animals as complex as fishes or even reptiles also show noninteractive sensorimotor schemata, which are always automatically activated by the perception of a certain object, for instance prey. I stress that, to speak here of a sensorimotor *schema*, several different motor segments need to be coordinated. This already demands a relatively high-level brain or social organization in order to develop and coordinate sophisticated motor programs.

### 12.7.2 The Second-Order Representational System

The other two forms of representational system are both active: Signs are actively produced *as* signs. In other words, they are produced *purposively*, as part of a strategy (even if not necessarily voluntarily). I would like to stress that it is not simply a goal-directed behavior [Subsec. 8.2.2]: Any of the previously considered semiotic levels in which information control plays a role are goal-

<sup>148</sup>[WENNER/WELLS 1990] [WENNER 1997]. Moreover, bees show a considerable ability to react to workers when they come back, telling the others about some nectar source. This form of behavior, however, is satisfactorily explained by the first representational type of process mode (for instance, to take a dance as a sign of a food source), without having recourse to more sophisticated semiotic forms.

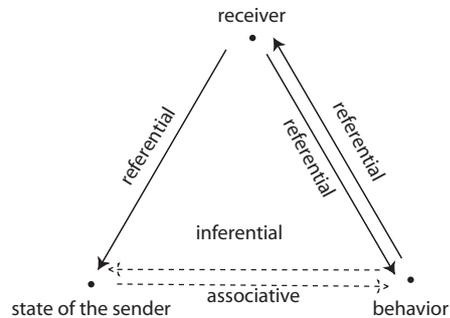
<sup>149</sup>[HÖLDOBLER/WILSON 1990, HÖLDOBLER/WILSON 1994].

<sup>150</sup>[LEADBEATER/CHITTKA 2007]. <sup>151</sup>[ROMANES 1884].

<sup>152</sup>Deception and mimicry is also common in plants [WICKLER 1968]. For a basic overview of some kinds of deception see [GRIFFIN 2001, pp. 212–27]. <sup>153</sup>[TOPOFF/ZIMMERLI 1993].

<sup>154</sup>[LLOYD MORGAN 1894, p. 53][LLOYD MORGAN 1900, pp. 99–100]. See also [REZNIKOVA 2007, pp. 7–8].

## 416 Representational Semiotics



**Fig. 12.17** The second-order representational system. A sender expresses its state through an iconic behavior that is referred to a receiver so that the receiver will understand the behavior as a sign of the state.

directed. Here, however, there is something more: The interaction with the addressed individual (in the second-order representational system) or with the final recipient of information (in the third-order representational system) is *explicitly considered* in this strategy. In other words, purposive behavior is already *interactive at the start* and not only as a consequence of an action.

As to the second order of representation, the relationships may be cast as in Fig. 12.17. It is especially expressed in visceral schemata and sensorimotor interactive representations [Subsec. 12.5.2]. This system can be found in mammals, that is, animals which show an affective system, and to a certain extent in birds, even if rudimentary sensorimotor representation can also be found in lower animals (in this case, it is probably bound with a single sensory channel and without an interactive component). It can also be the case that rudimentary forms of active communication are present in other vertebrates [Sec. 12.6]. Empirical research will tell us more on this point and help us to discriminate more.

Here we have two referential relations, one of which is even doubly referential, and the sign is actively produced as a behavior for a receiver: The sender behaves or, in the most general case, changes some physical feature (not necessarily in a voluntary way), so that another living being (not necessarily of the same species), the receiver, understands this feature as a sign of a certain state of the sender. For instance, a dog is willing to fight (state) and shows its teeth (behavior) in order for a human to be able understand its action as a sign of a disposition to attack.<sup>155</sup> In this case, the sender refers to a possible receiver so that the receiver will infer that the behavior is a sign of a certain state (in other words, so that the receiver will refer in this way to the sender's state):

- The relation between the state of the sender and his behavior is again purely *associative*, whereas, from the receiver's point of view, it is *inferred*.
- The relation between the sender's behavior and the receiver as well as between the receiver and the state are both *referential*. The former is even doubly referential, as far as the behavior is addressed to a receiver but the receiver also refers to the behavior as a sign (through some proto-inference) of the sender's state. This is the ground of the interactive aspect of the schemata used here.

Indeed, this complex relation is subject to error and deception tactics.<sup>156</sup> For instance, an animal can give to a predator a sign that it is wounded when it is not, or give a false alarm for sexual

<sup>155</sup>[MEAD 1934, pp. 42–51].

<sup>156</sup>[MITCHELL 1986].

reasons.<sup>157</sup> The most interesting case in this respect is that of birds trying to appear injured (with a broken wing) with the purpose of preventing attacks on their nest. In a detailed study it has been shown that, in 44 out of 45 cases, birds tried to lead an intruder away from the nest and that they modified their behavior depending on variations in the intruders' behavior.<sup>158</sup> On the other hand, the receiver can also be independently wrong or deceived about the relation between sign and state. It can take the sign as an invitation to mate, when it is not. Or it can pretend to be impressed by a menacing behavior, when it is not (and is rather ready to fight).

With regard to the receiver, experimental evidence has been found for single dynamic patterns of the brain when producing schemata, results that would have been dismissed as noise by employing traditional methods.<sup>159</sup> These patterns were stabilized when animals learned that the behavioral meaning of a stimulus (here, the verbal command "move" or "don't move") was based on the abstract quality of tone direction (rising or falling tone).

This form of representational process comprehends all forms of emotional expression. These have been extensively studied by Darwin.<sup>160</sup> Normally, honest signals [Sec. 12.6] are emotional signals, such as crying. For instance, the pitch of the voice appears to convey information about body weight and motivational state in many nonhuman primates. Human infants emit many types of communicative grunts accompanied by nonvocal gestures, such as pointing to a desired object.<sup>161</sup> However, again, many emotional signals may be deceiving. Many sounds aim to arrest the perceiver, or let him approach. Facial expression can be emotionally expressive but also strategic.<sup>162</sup>

### 12.7.3 The Third-Order Representational System

The third-order representational system (active communication to a partner of some sign having referential import and following quasicodified rules) may be represented as in Fig. 12.18. It is another form of active process and is triply referential. It is likely that such behavior can be found only in highly developed mammals, such as primates and cetaceans.<sup>163</sup> In the following I shall focus on primates.

In this process, a sender presents a possible receiver with a sign in order that the receiver should understand the sign as indicating a third thing (an event, an object, or even a third animal). In this way, the sender refers both to the receiver and to the object (this referential relation is something new insofar as it establishes a direct communicative link with a conspecific, as we shall see). Simultaneously, the sender intends that the receiver refers to the object (a relation that can only be inferred by the sender), doing that by referring at the same time to the sign as a sign of the object (here we again have a proto-inference, but from the perspective of the receiver). In the third-order representational system, the sign is explicitly produced from the start in its semiotic (triadic) function, i.e. as a sign *of* an object or event *for* a receiver who therefore plays the role of an interpretant.<sup>164</sup> By interpretant I mean an animal that is able to understand a sign as used by somebody else *as* a sign standing *for* an object. As we shall see, this is the basis of information interpretation which is the highest and a symbolic form of dealing with information. For this reason, all the relations, apart from the inferences, are referential, and deception can be far richer than in the other two semiotic systems. The sender can deceive the receiver so that he will not be able to

<sup>157</sup>[MUNN 1986a]. <sup>158</sup>[RISTAU 1991b] [*GRIFFIN* 2001, pp. 221–5].

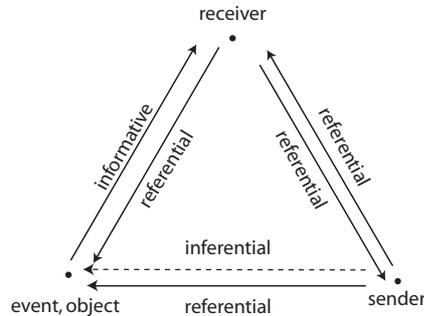
<sup>159</sup>[OHL *et al.* 2001] [KILGARD 2001]. <sup>160</sup>[DARWIN 1872].

<sup>161</sup>[HAUSER 1996, pp. 471–653]. <sup>162</sup>[EKMAN 1984, *EKMAN* 1985].

<sup>163</sup>[CHENEY/SEYFARTH 1990a]. See also [STEKLIS 1985]. I shall deal here with primates. See [TYACK 2000] for cetaceans.

<sup>164</sup>[PEIRCE *CP*, 1.339].

## 418 Representational Semiotics



**Fig. 12.18** Third-order representational process, in which we have several referential relations but no associative one.

understand correctly the event or the object (e.g. the sender can simulate a danger where there is none). He can give the receiver the false information that, e.g., a third individual will attack her or him because she or he fears a possible alliance between the receiver and the third participant. On the other hand, the receiver can fail to understand what the referred event or object is, or can pretend that he or she fails to understand. As we shall see, all these forms of deception are quite common among apes.

In a study by Marler and associates it is shown that, in domestic chicken, a male's food calling communicates information about food quality to a female receiver.<sup>165</sup> This perhaps explains the general mechanism by which the third-order representational system has phylogenetically arisen through the second-order one: Affective involvement may give rise to information communication through a process of growing specialization, leading from interactive sensorimotor schemata to a true active transmission of information.<sup>166</sup> In this context, the concept of active "functional referentiality"<sup>167</sup> can be very helpful: The chicken's alarm may not be dependent on the immediate context but still on a larger or historical context.

I wish to stress that in the third-order representational system a direct representation of objects or events is not communicated (which is impossible, given that representational icons are endogenous patterns in the brain). Rather, the partner is invited to produce a representation that is appropriate to the sign communicated by the sender.<sup>168</sup> Therefore, the content that in the other two representational systems is embodied in an iconic aspect, is delocalized here in the web of the threefold referential relations. For this reason, the third-order representational system is tightly bound with the issue of categories and categorization [Subsec. 12.5.2], as we shall see below. Let me give some specific examples of the third-order representational system:

- First, let us consider call alarms emitted by monkeys. Contrary to previous studies up to the end of the 1970s, it has been shown that nonhuman primate vocalization is under voluntary control. The difficulty with alarm signals is that they are altruistic and costly with regard to survival, but are also very common<sup>169</sup> [Sec. 12.6].

<sup>165</sup>[MARLER *et al.* 1986a]. In a subsequent study [MARLER *et al.* 1986b] it was shown that males' calls for food are lower when there is no presence of conspecifics and lower still when there is a male competitor. In general, an audience effect in avian calls has been acknowledged [GYGER *et al.* 1986, EVANS/MARLER 1994]. See also [MARLER/EVANS 1996].

<sup>166</sup>[MARLER 1992][MARLER *et al.* 1992]. <sup>167</sup>[EVANS *et al.* 1993b]. <sup>168</sup>[AULETTA 2007].

<sup>169</sup>[HAUSER 1996, pp. 363–470].

Call alarms of vervet monkeys denote external referents.<sup>170</sup> In fact, they present a mix of categorical and indexical information. Scholars have also succeeded in classifying different grunts (generally social bound) according to rudimentary proto-syntactic rules. Syntax, in a first approximation, may be defined as any system of rules that will allow us to predict sequences of signals. Marler<sup>171</sup> appropriately distinguished between *phonological* syntax, where the ordering of the elements is important (the same elements in a different order cause different responses) [Subsec. 7.4.1]; and *lexical* syntax, when the meaning of the compound call results from the combination of the meaning of its constituents. The former is only a proto-syntax, while the latter is a true one. There is evidence for the presence of a phonological syntax in nonhuman primates<sup>172</sup>. Gibbons' calls are produced in specific phonological sequences<sup>173</sup>. Young monkeys develop their ability to use signals according to the feedback that they receive from adults (whether they themselves cry confirming the presence of a danger or not).<sup>174</sup> Depending upon the circumstances, vervet monkeys can choose to give an alarm call without an escape response or to flee without giving a call. Solitary vervets do not give alarm calls at all when confronted with a predator. Therefore, vervet monkeys seem to modify their alarm calling rate depending on their audience (the conspecifics they wish to address or not). This is called *audience effect*.<sup>175</sup> They also take into account the signaler's identity: Vervet monkeys ignore signals emitted by unreliable senders. Transfer does not occur if either the sender or the referent changes.<sup>176</sup> Vervets also recognize alarm calls of other species (for example the superb starling). Moreover, monkeys classify their vocalization according to their referents (signals with different acoustic properties are similar for their referent). Vervet monkeys have three different alarm calls (for leopards, eagles, and snakes) which lead to very different responses. They are probably kinds of injunctions rather than statements. In any case, they are messages directed toward conspecifics and referring to events.<sup>177</sup> As we shall see later, "leopard" or "eagle" have here the status of a *category*. Putty-nosed monkeys can even combine the alarm call for eagle and that for leopard, producing a third call that is a trigger of group movements.<sup>178</sup>

The above conclusions have also been confirmed by experiments in the field: Playback experiments on diana monkeys have been performed in which two calls were played in sequence, separated by five minutes of silence, such that they were either (a) similar in acoustic and the so-called semantic features,<sup>179</sup> (b) similar in semantic features only, or (c) different in both acoustic and semantic features.<sup>180</sup> Subjects transferred habituation across acoustic but not semantic features, suggesting that they attended to the call's underlying reference. Summing up, what Cheney and Seyfarth and other scholars of the field have shown is that monkeys respond to acoustic alarms and not directly to the predator itself or to the behavior of the monkeys emitting the alarm call<sup>181</sup>: The alarms be understood as implying something about the threat and not about the evasive action to be taken. Nevertheless, it is crucial to clarify that alarm calls of nonhuman primates are still a semiotic way of communicating and not a symbolic one.<sup>182</sup> Indeed, cognitive aspects involved here are confined to<sup>183</sup>

<sup>170</sup>[MARLER 1977a] [CHENEY/SEYFARTH 1990a, pp. 98–183].

<sup>171</sup>[MARLER 1977b]. <sup>172</sup>[EVANS/MARLER 1995]. <sup>173</sup>[MITANI/MARLER 1989].

<sup>174</sup>[CHANGEUX 2002, p. 137]. <sup>175</sup>[MARLER *et al.* 1991]. <sup>176</sup>[CHENEY/SEYFARTH 1988].

<sup>177</sup>S[SEYFARTH *et al.* 1980a, SEYFARTH *et al.* 1980b]. <sup>178</sup>[ARNOLD/ZUBERBÜHLER 2006].

<sup>179</sup>As we shall see, the word *semantic* is used improperly here. However, for the time being we do not need to worry about this issue.

<sup>180</sup>[ZUBERBÜHLER *et al.* 1999]. <sup>181</sup>[WYNNE 2004, p. 128].

<sup>182</sup>This could be confirmed by the fact that there are striking similarities in vocalization and frequencies between chimpanzees and gorillas [MARLER 1976], while true symbols are not immediately connected with a perceptual or biological substrate (from this point of view they are arbitrary).

<sup>183</sup>[JÜRGENS 1990].

## 420 Representational Semiotics

- (a) Identification of vocal signals, and
- (b) Learning to associate a particular call with a particular context.

The human Wernicke's language area is a multimodal brain area (and this is also the reason why a sign language as well as a vocal one, and all other forms of codification like scripture and so on, are possible), while in nonhuman primates species-specific communicatory signals seem to be decoded in the same brain structures where nonspecies-specific sounds are also processed.

- The second type of example is represented by active transmission of information during close social interplay. Gibbons acknowledge artificially rearranged songs as conspecific territorial signals but respond in a qualitatively different way. Rhesus monkeys (a variety of macaques) from Puerto Rico give five acoustically distinct scream vocalizations during agonistic encounters in order to recruit support from allies against opponents.<sup>184</sup> In primates, alliance formation varies according to the social rank and the matrilinear relatedness of the opponent as well as to the severity of the aggression. Since they are strongly related to the particular class of the opponent and the level of physical aggression, they have referential import. Pigtail macaques of the Yerkes Center also employ acoustically distinct classes of agonistic screams depending on the context, especially on the rank of the opponent and the severity of aggression. Four different screams have been observed. Moreover, contextual usage undergoes developmental modification.<sup>185</sup> Generally, one may distinguish between noisy, arched, tonal, pulsed, and undulating screams.<sup>186</sup> When a monkey hears a grunt, it is immediately informed of many details even though it may be out of sight of the vocalizer and even though the vocalizer itself may not be involved.<sup>187</sup> Another important point is the redirected aggression, depending on the dominance or subordination of both the speaker and the receiver. Indeed, understanding social relations is crucial for managing third-order signs. Vervet monkeys are more likely to threaten a particular individual if they have previously been involved in a fight with that individual's close kin.<sup>188</sup> Java monkeys are very good at discriminating among different mother-child affiliations.<sup>189</sup> It is also interesting to consider that foraging vocalization of rhesus monkeys reared by Japanese monkeys and of Japanese monkeys reared by rhesus monkeys are different relative to their ordinary, conspecifics, and are treated as different by them.<sup>190</sup>

Vervet monkeys also show the ability to recognize the relationships that exists among other individuals.<sup>191</sup> In order to recognize that certain social relationships share similar properties, one must either memorize all relationships to which one has been exposed and then evaluate them according to some criteria, or classify different types of relationships so that they can be compared independently of the particular individuals involved.

- Studies on apes also confirm and extend these results. A study on 115 captive chimpanzees<sup>192</sup> suggests that referential and purposive communicative behaviors, in the form of gestures, exist and are lateralized to the left hemisphere in chimpanzees. A striking example is offered by the chimpanzee Kanzi, raised by Sue Savage-Rumbaugh. Kanzi often hides objects in the woods and remembers the location days later. The hiding of objects can also be a game that is played with pretended objects instead of real ones. Kanzi also hides himself. Hiding oneself not only needs computation of the line of regard, but also imagining what others will do at different points in time, implying the ability to be in a strong communicative interaction. Kanzi also uses signs to

<sup>184</sup>[GOUZOULES *et al.* 1984].<sup>185</sup>[GOUZOULES/GOUZOULES 1989].<sup>186</sup>[GRIFFIN 2001, pp. 170–2].<sup>187</sup>[CHENEY/SEYFARTH 1990a] [GRIFFIN 2001, pp. 166–70].<sup>188</sup>[CHENEY/SEYFARTH 1986].<sup>189</sup>[DASSER 1987].<sup>190</sup>[MASATAKA/FUJITA 1989].<sup>191</sup>[SEYFARTH/CHENEY 1988, CHENEY/SEYFARTH 1990b].<sup>192</sup>[HOPKINS/LEAVENS 1998].

mislead others: This requires the realization that others can react not only as a consequence of what one does, but also by what one communicates (a highly interactive behavior). When Kanzi informs someone that he will take a melon and instead he goes to play with his companions and he knows that the caretakers will let him go, this ability of anticipating what will happen requires the capacity of misrepresenting one's own goals, selecting instead particular goals that are known to be acceptable to others but that can be used indirectly and communicatively to accomplish one's own ends.<sup>193</sup> The chimpanzee is able to use both accurate and misleading information, by taking into account the nature of sender or recipient.<sup>194</sup>

Byrne and Whiten<sup>195</sup> compiled a catalogue of deceiving behaviors of primates, where the common trait is the necessity for the agent to monitor the actual or potential attention of the target relative to the agent:

- Concealment (to conceal something from the target): Hiding from view, acoustic concealment;
- Distraction (to distract the target's attention away from a particular locus, at which point it is directed to a second locus): Distract by looking away, distract by looking away with a linked vocal signal, distract by leading away, distract with intimate behavior;
- Creating an image by means of which one causes the target to misinterpret the behavior's significance for itself in other ways: Present a neutral image, present an affiliative image;
- Manipulation of the target using social tools: Deceptive tool about the agent's involvement with the target, deception of a target about the agent's involvement with the tool;
- Deflection of the target to a fallguy (to divert the target by directing it towards a third party, the fallguy).

Summing up, to deceive in the way that has been reported here, an animal must first be able to predict what another individual would do in another context or as a reaction to a certain event (to deceive is a counterfactual). They are also able to understand that competitors may do the same and to take appropriate action.

## 12.8 Concluding Remarks

Representation is a specific semiotic activity. Neither instructional semiotics nor pure reactive behavior is representational. Representations are typical of animals and are born with the purpose of controlling the consequences of their own actions on the environment:

- Representations have been born and developed under the constraint of combining information of different types and gaining a larger autonomy from the environment. The received negative feedback from the environment stimulates the organism to find a better fit. However, accommodation is not the only tendency. We also have assimilation due to the fact that there is a tendency to restore the homeostasis when there is a perturbation caused by the negative feedback.
- Connectionism is a way to understand the birth of representation in associationist terms. Here, it is very important to acknowledge the three-layered structure that is able to account for the creation of complex (hierarchically organized) patterns.

<sup>193</sup>[SAVAGE-R./MCDONALD 1988]. <sup>194</sup>[WOODRUFF/PREMACK 1979].

<sup>195</sup>[WHITEN/BYRNE 1988c, WHITEN/BYRNE 1988a]. For further impressive examples of deception tactics among primates see [DE WAAL 1982, DE WAAL 1989] [BYRNE/WHITEN 1988] and also [BYRNE 1995].

**422** Representational Semiotics

- In a PDP network single excitation patterns or items can represent different objects or events. In this case, it is possible to cover a very big representational space with a relatively small number of items, each of them covering a portion of the space. This is also the way in which the immune system works.
- PDP shows that there is no representational function that is not instantiated in some current activation.
- Notwithstanding these results, connectionism is not fully capable of explaining higher cognitive performances like those making use of active communication or symbols.
- Representations are organized in perceptual and sensory motor schemata and categories. The general law here is: Any representation is maintained at a certain level of genericity if the evolutionary (or ontogenetic) pressure does not oblige the finding of more specific solutions. Schemata are strictly bound with affordances, the catching of some environmental aspects that possess a specific (survival) value for the organism.
- Following P. Marler, it is necessary to distinguish between three forms of representation: Reference (first-order representational system, centered on perceptual schemata and on noninteractive sensorimotor schemata): It is characteristic of nonmammal vertebrates; addressing or deictic semiosis (second-order representational system, centered on visceral schemata and interactive sensorimotor schemata), which is typical of low mammals; active transmission of signs (third-order representational system, centered on categories), which is typical of primates and cetaceans.

Let me stress that, from a pure relation-theory point of view [Subsec. 2.4.4], the first-order representational system is monadic: A single agent may perform this inference without considering any other agent. Therefore, it does not include communication media. Consequently, there is only a single referential relationship. The second system is dyadic, that is, it is intrinsically interactive with another agent, and therefore there are two referential relations. The third system is triadic, in that there are three referential relations (a triangle sender–object–receiver). Further evidence of this increase in referential relationships may be found in the fact that—as I have stressed—any referential relation somehow presupposes an arbitrary relationship between its two terms and thus can be an object of deception.

After these general considerations on representational semiotics, let us consider once again the machinery of the brain.

# 13

## The Brain as an Information-Control System

---

After a long investigation we come back to the issues raised in the first part of the book [Chs. 3–5]. However, we shall now treat the brain as an information-control system and not only as an information-processing or pure information-acquiring device. Indeed, we have seen the insufficiency of a pure information-processing treatment of the problem as well as of a pure information-acquiring one [Sec. 6.1]: Information-processing is concerned with inputs that are given and eventually mapped according to certain fixed rules; information-acquiring does not have anything to do with the reference or the function of the acquired information. Information control [Sec. 8.2] is concerned with biological functions which are not the result of physical and chemical conditions: The organism must be able to individuate and monitor relevant parameters and systems, in order to perform effective actions for keeping control on its living context as high as possible. Obviously, these parameters and systems must be directly observable in principle.

In this chapter, after some general considerations, we shall discuss phylogenetic brain stratification (the mammal's brain becomes the center of our analysis), and study the brain as a system for integrating information from different sources and for developing coordinated behaviors.

### 13.1 Brain Stratification

#### 13.1.1 Some General Considerations

At an ontogenetic level, we can distinguish between a developed sensory system (vision, some mechanisms for “feeling” sound reception, and at least in vertebrates also a tactile system), an advanced metabolic system, consisting in a regulation that in higher animals happens through hormonal production, and a CNS, whose central organ is the brain [Sec. 3.1]. We have already treated the peripheral sensory system in general [Sec. 3.3] and vision in particular [Ch. 4] as well as the central control of movement [Ch. 5]. In the following, we shall discuss mainly the brain (later on we shall briefly examine the hormonal system). The original function of the brain was probably to control motion [Subsec. 12.3.1], that is, actions on the world and therefore also to select specific reactions, and it was only later extended to cover many other functions. Any organism, being a complex system [Secs. 6.3–6.6], can be considered to a certain extent to embody a fractal, recurrent structure [Subsec. 9.5.4]. Therefore my guess is that a specialized structure like the brain of vertebrates, which originally developed from the selection system of the organism [Secs. 7.6 and 11.5, Subsec. 11.4.4], displays a functional organization that reproduces the basic structures of

**424** The Brain as an Information-Control System

dealing-with-information<sup>1</sup> and especially the general features of information control [Secs. 8.2–8.4], becoming itself articulated in:

- (1) A system for obtaining information from the outside and processing it,
- (2) A system for regulating time and the waking functions of the organism, and
- (3) A system for selecting appropriate actions and goal-directed behavior.

This should not be a surprise, given my assumptions that (i) any dealing-with-information reproduces the main structure shown in Fig. 2.5 and (ii) any biological system is semiotic in its deepest nature [Chs. 8 and 12].<sup>2</sup> Indeed, as Peirce correctly understood,<sup>3</sup> these three aspects are deeply rooted in the nature of signs, and he called the above three forms of semiotic activity the logical, the emotional, and the energetic interpretants (the high semiotic functions that are vitally meaningful for the organism).

**13.1.2 The Brain Does not Code Information**

As we know, the peripheral sensory system codes information [Subsec. 3.3.1]. It is often supposed that the brain also codes information and that there is a universal language of the brain or of thought.<sup>4</sup> A. Clark<sup>5</sup> criticized Fodor’s theory of the language of thought because it assumes the existence of symbolic “atoms” which are differently combined.<sup>6</sup> We have indeed seen that the brain’s representational activity is best expressed in connectionist terms [Sec. 12.4]:

- (1) *Representations can overlap* (partially or totally).
- (2) *The brain shows intrinsic context sensitivity.* One does not need a core representation of a thing but there could be a variety of states again linked by family resemblances. In this case, no generalized computation is necessary.
- (3) The system can increase its representational power, and not suffer big damage if locally disturbed, since it is a *distributed network*.

It is interesting to note that neural patterns are self-organizing nonlinear structures [Subsecs. 6.3.1 and 6.5.2] and therefore show an important difference with respect to atom and molecule combinatorics [Subsec. 6.2.1]: In the brain, there are no elementary units at all apart from single neurons.

Therefore, the main conclusions are

- The brain does not codify information, although single neurons may be excited or not according to a binary code [Subsec. 3.3.2] and information is acquired peripherally through sensory organs in a codified way [Sec. 8.1]. An evidence for this initial and local codification is provided by the fact that the perceptual import shows no nonlinear transformations, which accounts for the Stevens’s law of stimulus intensities<sup>7</sup> [see again Subsec. 3.3.1]. This view makes justice of the distinction between *local* information coding and *global* nonlinear excitation patterns [Secs. 3.1 and 6.5].

<sup>1</sup>As understood perfectly for the first time by A. Lu’ria [LUR’IA 1973, pp. 43–101]. See also [HERBART 1816, Para. 55–6].

<sup>2</sup>The idea that signs are necessary for higher cognitive functions can be found in [CHANGEUX 2002, pp. 113–19], following the French tradition mentioned in Sec. 12.1.

<sup>3</sup>[PEIRCE 1907, p. 409]. <sup>4</sup>[FODOR 1975] [CALVIN 1996]. <sup>5</sup>[CLARK 1993a, pp. 3–112].

<sup>6</sup>For one of the first attempts at interpreting the brain as a symbolic machine see [CRAIK 1947].

<sup>7</sup>[MOUNTCASTLE 1998, pp. 9–11].

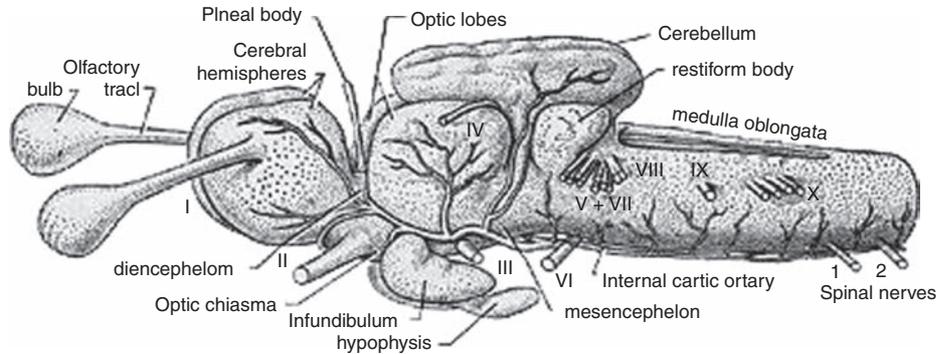


Fig. 13.1 Fish brain. Adapted from <http://www.austmus.gov.au/fishes/faq/pain.htm>.

- When information arrives at the CNS, there is a neuron combinatorics that provides a representation of things, i.e. a given spatial-temporal configuration associated with certain stimuli [Subsec. 12.1]. This representational combinatorics is analogical and not digital or somehow codified [Subsecs. 4.4.1 and 7.4.5]: Bigger objects excite more neurons in the mean, excitation patterns are sensible to rotations, and so on.<sup>8</sup> For this reason, it neither represents, nor uses a symbolic language. I mention here that if the brain were a symbolic machine, then it would be so phylogenetically from the start, but we have no evidence of symbol use by lower vertebrates.

This is fully in accordance with the so-called neural-processing hypothesis of perception, according to which no symbolic mediation is necessary, either in the peripheral nervous system or in the cortex.<sup>9</sup> This is the biggest *qui pro quo* that the AI has to overcome. Summing up, the brain is only a semiotic system, although it is the most general semiotic system on Earth.

### 13.1.3 Three Evolutionary Layers

Preliminarily, recall that the cerebrum has been described as having three phylogenetic parts: The archipallium, the paleopallium, and the neopallium. In fish, the archipallium is the largest part of the cerebrum. Some researchers suggest that the early archipallium gave rise to the human hippocampus, while the paleopallium was further developed into the limbic system. In amphibians, the cerebrum includes the archipallium, paleopallium, and some of the basal nuclei. Reptiles first developed a neopallium, which continued to increase in more recent species to become the neocortex of Old World monkeys and humans. The human pallium envelops the forebrain.

It is therefore convenient to have a short look at the evolution of the brain. For the following considerations the reader may see Subsec. 3.4.1, Secs. 4.2–4.3, and Subsec. 5.3.1. I shall not consider here the great evolutionary gap between the first brain organization and that of developed vertebrates:

- The fish brain [Fig. 13.1] is still a patchwork, where different primary sensory areas (for smell and vision) of the brain are clearly visible, a rudimentary midbrain and a diencephalon are already present as well as the cerebellum. This means that motor information is also acquired and processed and there is probably some kind of feedback loop between sensory and motor information. The metabolic system seems to be relatively independent from these structures.

<sup>8</sup>[BATESON 1966, pp. 372–4].

<sup>9</sup>[MOUNTCASTLE 1998, p. 6].

426 The Brain as an Information-Control System

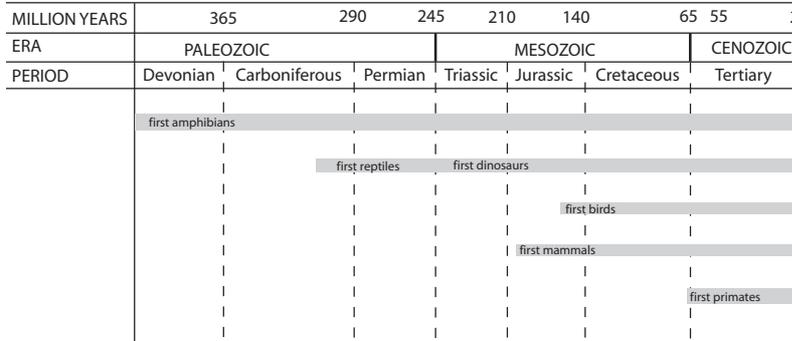


Fig. 13.2 Evolution of out-of-sea animals.

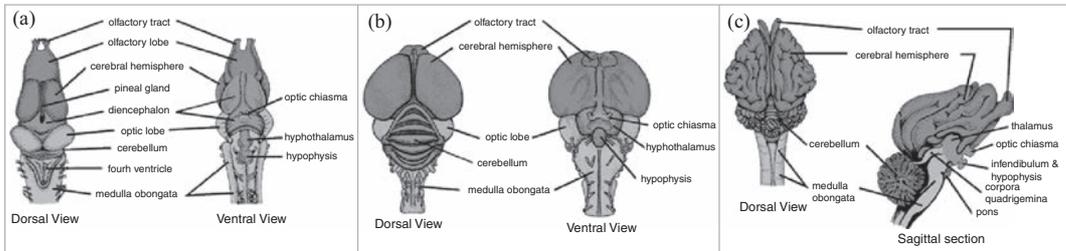


Fig. 13.3 Comparison among (a) amphibian, (b) bird, and (c) mammal (cat) brain. Adapted from <http://www.uoguelph.ca/zoology/devobio/210labs/ecto3.html>.

- Things stand a little bit differently with the first terrestrial animals, namely amphibians [Figs. 13.2 and 13.3]. Common to an amniote’s forebrain organization is the fact that ascending sensory information is relayed through dorsal thalamic nuclei to parts of both the pallium and subpallium.<sup>10</sup> The pallial (cortical) areas tend to be specialized into discrete regions in receipt of afferent information from one particular sensory modality. In nonmammalian amniotes, the dorsal pallium contains the dorsal cortex. In birds, a rostral expansion of the dorsal pallium which receives visual inputs from the lemnothalamic system has been termed the visual Wulst.
- The dorsal ventricular ridge (DVR) is a pallial structure unique to reptiles and birds. Despite differences in DVR organization, common to all reptiles and birds is the basic connectivity of the anterior DVR (ADVR) and the basal DVR (BDVR). P. Ulinski<sup>11</sup> suggested that the basic organization of the DVR serves as a linkage between sensory inputs and motor outputs; an interface between sensory and perceptual processing and mechanisms which modulate behavior. The ADVR receives visual, auditory, and somatosensory information; it sends outputs to the BDVR and basal ganglia. As a matter of fact, it was shown that their brain, especially the midbrain, is able to integrate sensory information in a way that is relatively similar to that of mammals.<sup>12</sup> Unlike amphibians, we clearly see a more integrated information system, in which sensory information acquisition and motor information control tend to constitute a whole

<sup>10</sup>All terrestrial animals and birds that lay eggs covered by a membrane are amniotes.

<sup>11</sup>[ULINSKI 1983]. <sup>12</sup>[STEIN/GAITHER 1981].

plastic circuitry, allowing for a diversification of behavior and for a much better ability to learn. It may be assumed therefore that reptilians are able to integrate sensory information coming from different sources and to have well-developed noninteractive sensorimotor schemata [Subsec. 12.7.1]. Moreover, the reptilian brain consists of a very developed cerebellum and a fully constituted brain stem, creating the appropriate interface between the metabolic system and brain.

- The dorsal and lateral pallia in mammals have undergone a spectacular expansion, developing into the laminar neocortex. The six-layered cortex, with its marked sulci and gyri, is often considered the defining characteristic of mammalian brains, and was once considered a prerequisite for true higher-level sensory and cognitive activities. Subsequent research into the cognitive and sensory capacities of birds and other organisms has moderated this judgment [Subsec. 12.7.2]. The differences, however, between the avian–reptilian dorsal ventricular ridge and mammalian neocortex are instructive, for instance, both in the attempt at reconstructing the visual-system evolution and in understanding visual processing in these very different brains. The old-mammal brain, apart from an initially developed neocortex, presents a transformation of the brain stem into a part of the whole brain circuitry itself, by connecting basic metabolic functions to behavior and memory. In this way, metabolic functions both contribute and are partially controlled by higher cerebral functions: Sensorimotor schemata become interactive [again Subsec. 12.7.2] and tightly bound with expectations.

I would like to stress that the mammal emotional system is a true emergence at an evolutionary scale. However, this does not mean that there was no social interaction between lower organisms and even single cells. In fact, the epigenetic building of the organism shows the opposite [Subsec. 11.2.1]. Recently, it has been shown that even the amoeba *Dictyostelium purpureum* prefers to form groups with its kin in situations where some individuals die in order to assist others.<sup>13</sup>

- Neomammals have realized much more. They have achieved a full integration among the reptilian brain (able to perform a basic learning), the limbic system (capable of a more refined learning through experience), and the neocortex (allowing for learning through thinking). This is essential for allowing large increases in cognitive performance.

### 13.1.4 The Triune Brain

Starting from the previous results, P. Maclean has proposed that the human brain<sup>14</sup> has a three-layer system, the so-called theory of the triune brain [Sec. 12.7]: A reptilian layer (represented especially by basal ganglia) which provides for innate instinctual action tendencies (a first-order representational system), an old mammalian layer which is responsible for the affective system (a second-order representational system), and the neomammalian brain (fully developed neocortex), providing for higher cognitive functions (third-order representational system) [see also Sec. 3.4]. Maclean remarked that, in the evolutionary transition from reptiles to old-mammals, three cardinal developments occurred:

- Nursing, in conjunction with maternal care,
- Audiovocal interactive communication for maintaining mother–offspring contact, and
- Play.

<sup>13</sup>[MEHDIABADI *et al.* 2006].

<sup>14</sup>[MACLEAN 1990]. However, the following idea applies to primates or at least to apes as well.

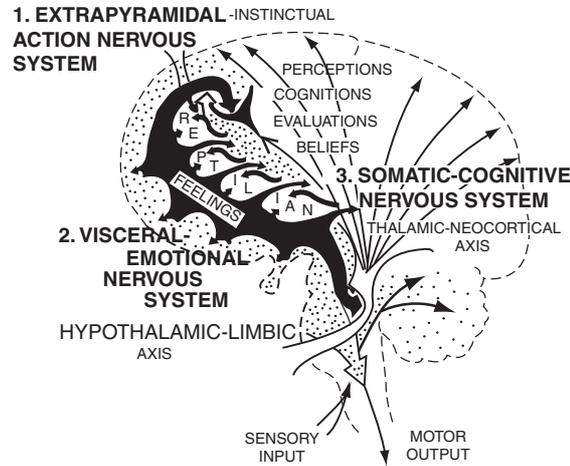


Fig. 13.4 Panksepp's triune brain. Adapted from [PANKSEPP 1998, p. 62].

All three changes deal with sensorimotor interactive schemata [Subsecs. 12.5.2 and 12.7.2]. The new situation represented by mammals determines the new significance of the social environment for the development of the brain.<sup>15</sup>

J. Panksepp has adopted and further developed Maclean's schema, reinterpreting this as it follows<sup>16</sup> [Fig. 13.4]. We have two main paths [see also Sec. 4.3 and Subsec. 5.1.4]: A dorsal stream that goes to the neomammalian brain and is related essentially to external information (through optic, haptic, and auditive signals) and a ventral-visceral path related to chemical or internal signals (taste, smell, temperature, hormones). Both converge in the reptilian brain. Even if I do not follow all the details of this approach (for instance, birds probably already show at least some rudimentary affective behavior: Indeed they take care of their newborns), there is evidence that lower mammals (rabbits, dogs, and so on) make use essentially of chemical signals, while higher mammals (primates especially) rely more on the sense of sight in their dealing-with-information. As we shall see, this introduces a fundamental difference between referring to individuals through sensory marks or acknowledging them in a more integrated fashion. In any case, we can assume that the emotional system is absent in reptiles and amphibians, and arises as a specialization of the regulatory-endocrine system (feelings are actually strictly connected with the homeostatic state of the organism).

Here it is necessary to briefly specify this scheme: It should not be interpreted as saying that the brain of preneomammalian organisms dose not possess a decisional system. Rather, the emergence of neomammals is strictly related to that of a categorical-symbolic system, as we shall see. Moreover, the fact that the emotional and cognitive systems converge with the motor system does not exclude other convergences at all, as we have already seen above and shall further see below. It is very important, in general, to distinguish between evolutionary stratification and the current functions of the organism or of the organ<sup>17</sup> Exaptation plays a decisive role here [Subsec. 9.5.1].

<sup>15</sup>[WEXLER 2006, pp. 85–137].

<sup>16</sup>[PANKSEPP 1998, pp. 59–79].

<sup>17</sup>[LEWONTIN 1998].

Therefore, assuming the previous schema with some provisos, we could perhaps very briefly summarize it in this way:

- Multicellular animals with a CNS, and especially fish, amphibians, and reptilians, show a brain that is essentially characterized at most by perceptual schemata and noninteractive sensorimotor schemata, as far as the cognitive-motor aspect is concerned. They lack a developed emotional system, and present a minimal decision system. They essentially rely on a first-order representational system [Subsec. 12.7.1].
- Probably birds and old-mammals have acquired an emotional system. Moreover, low mammals have also acquired advanced interactive sensorimotor schemata, a primitive ability to protocategorize, and a decisional system far more advanced than that of the previous group. Old-mammals are able to individuate conspecifics and individuals of other species through semiotic marks (especially olfactory ones). This group is characterized by the second-order representational system [Subsec. 12.7.2].
- While humans (the subject of the third part of this book) have developed concepts, and, as we shall see, new forms of affection and decision, nonhuman primates can be considered to be at the threshold of this development. For instance, they have the ability to fully understand individuality and categorize. For this reason, they also make use of the third-order representational system [Subsec. 12.7.3].

To properly understand the complexity of information integration in the brain, a survey of human brain organization (where all the lower and higher aspects are displayed and integrated) can be very useful. However, I stress that some basic characteristics discussed in the next section are also common to all primates.

## 13.2 The Brain as an Integration System

The brain is a system that is structured *both* hierarchically [Sec. 6.3] and horizontally, i.e. distributed<sup>18</sup> [Sec. 12.4]. The primate brain can be considered as structurally organized in three major subsystems (which are also biological systems on their own [Sec. 7.2]), the sensory, the motor, and the emotional. In the first part of the book, we have considered the sensory [Chs. 3–4] and the motor [Ch. 5] systems. The reason for a rudimentary treatment of the sensory system in the first part of the book is obvious: It is a system for acquiring information, and therefore it fits well in that context. The reason for treating the motor system in that context is less obvious since this is not a system for acquiring information. However, the aim of that chapter was to show that a simple model of information acquisition was not adequate to describe basic functions of the organism [Sec. 6.1].

Instead, the proper place for dealing with the emotional system is here (after having presented some basic functions and processes of the organism in the last chapters) because it is strictly connected with the limbic system that also provides for some fundamental metabolic regulations [Sec. 7.3]. Therefore, a preliminary short look at the role of the hypothalamus and the endocrine system would be of benefit. The hypothalamus [Subsec. 3.4.1] controls three systems<sup>19</sup>: (1) The endocrine system, (2) the autonomous nervous system (the visceral and involuntary part of the nervous system, whose effector motor neurons are located in ganglia outside the CNS<sup>20</sup>) [Subsec. 5.3.1], and (3) the neural subsystem involved in motivation for action and (at least for

<sup>18</sup>[CHANGEUX 2002, pp. 28–31] [FRISTON/KIEBEL 2009].

<sup>19</sup>[IVERSEN *et al.* 2000a].

<sup>20</sup>[POWLEY 1999].

**430** The Brain as an Information-Control System

mammals) in emotion. The hypothalamus controls basic physiological needs, like blood pressure, body temperature, feeding, digestion and metabolic rate, reproduction, and emergency responses to stress. In this way, it integrates autonomic and endocrine functions with behavior. Indeed, it takes advantage of three main mechanisms. First, it has access to the sensory information of virtually the entire body. Second, in its regulative specific function, it compares sensory information with biological set points (like the temperature of 37° for humans). Finally, it adjusts an array of autonomic and endocrine responses and contributes to behavior in order to restore homeostasis [Subsec. 8.2.7 and Sec. 12.2]. The hypothalamus controls the endocrine system directly, through the pituitary gland secreting neuroendocrine products, and indirectly, by secreting regulatory hormones into the local portal circulation.<sup>21</sup> The pituitary gland is therefore sometimes called the “master” gland of the endocrine system. It is of the size of a pea and is located at the base of the brain.

Let us now have a closer look at the three subsystems of the brain [Figs. 13.5–13.6]:

- (1) With regard to the *sensory system*, it was observed that, although sensory nerve pathways deliver messages to the CNS that are quite invariant with respect to given sensory stimuli, the manner in which the CNS treats these messages depends on the functional status of each relay station. Thus, rather than being a simple mirror of the external world, the CNS embodies a dialogue between internal representation and the way information is acquired from the senses.<sup>22</sup> Let us consider this mechanism.

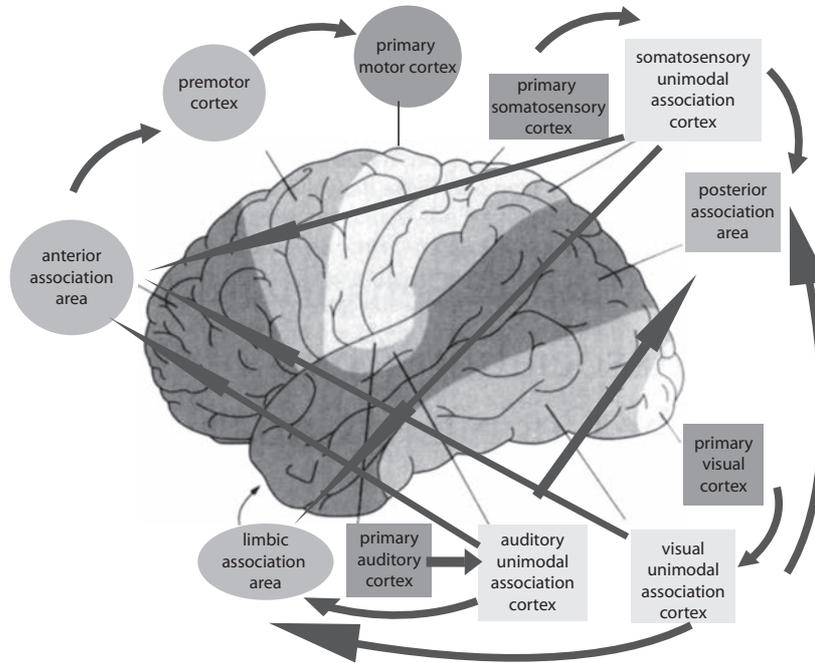
The interface between pure information-acquiring and the production of schemata and categories in vertebrates [Sec. 12.5] is represented by the sensory cortex. Here, excitation patterns that have an iconic function are already produced, but in the first steps are still dependent on the specific sensorial (visual or auditive, and so on) nature of the stimuli. I recall that all information arriving in the brain has essentially the same structure [Sec. 3.3]: Only the frequency (for the intensity of sensation or the speed of movement) and its reciprocal, i.e. the period (for the duration of the sensation or of the movement) of the action potentials matter. In this sense, the peripheral sensory system is a pure information-acquiring system. However, when these different pieces of information are acquired by the CNS, the specific message of an action potential (whether a visual or auditive stimulus, for instance) is determined by the neural pathway that carries it.<sup>23</sup> It is here that it acquires a general semiotic-representational function [Sec. 12.1]. In this way, evolutionarily speaking, the brain is emerged as an integration system able to act on the world and to represent it by sorting and combining information coming from different types of sources [Sec. 12.2]. Then, it has built higher cortical pathways acting as a constraining form of causality that, in a top-down process, can canalize information coming from lower levels [Subsecs. 3.4.3 and 6.3.2].

We have already examined some important structural aspects of the cortex [Sec. 3.4]. Let us now consider the problem a little bit closer,<sup>24</sup> starting with the sensory system.<sup>25</sup> The *primary sensory cortex* is composed of

- The somatosensory cortex (Brodmann’s areas 1, 2, and 3), located in the postcentral gyrus (parietal lobe). The somatosensory system detects experiences such as touch or pressure, temperature (warm or cold), pain, as well as proprioception, visceral (internal) senses, and facial expressions.
- The visual cortex (area 17), located in the banks of calcarine fissure (occipital lobe).

<sup>21</sup>[MURRAY 1989, pp. 166–75].      <sup>22</sup>[LLINÁS 1988].      <sup>23</sup>[MOUNTCASTLE 1978] [KANDEL 2000a, p. 31].

<sup>24</sup>A model that owes a lot to the pioneering work of J. H. Jackson [JACKSON 1915].      <sup>25</sup>[SAPER *et al.* 2000].

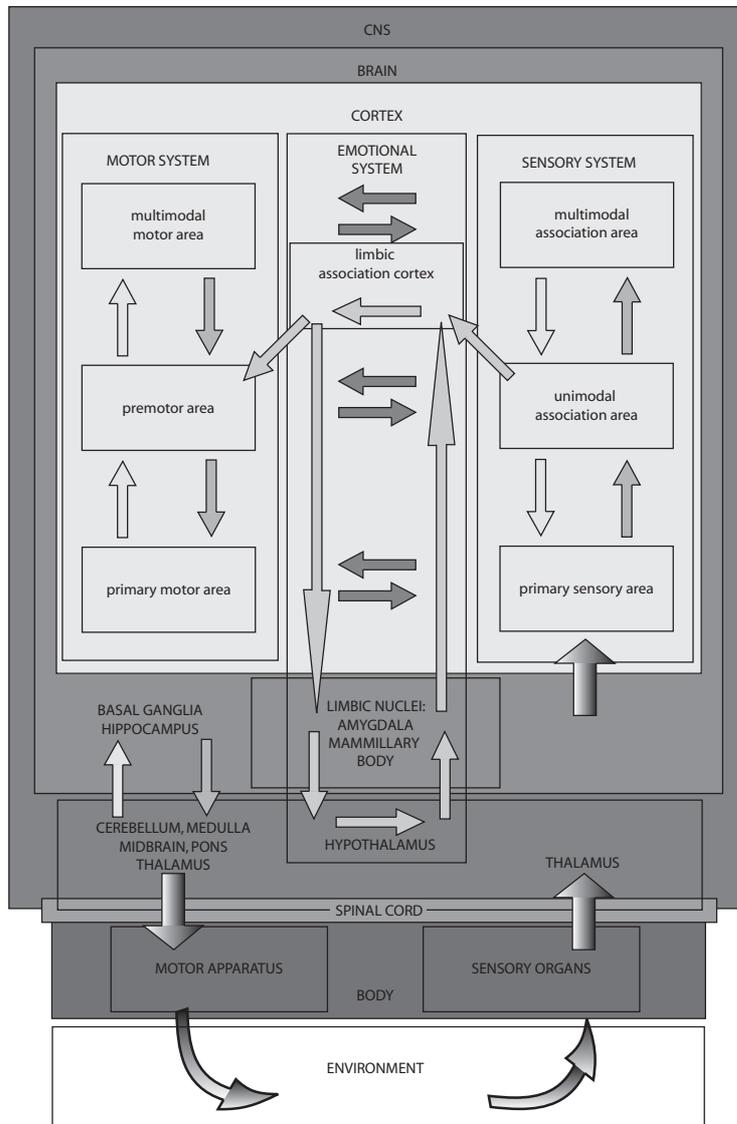


**Fig. 13.5** The lateral surface (left side) of the human cortex shows the regions of the primary sensory and motor cortices (in dark gray), the high-order motor and sensory cortices (in light gray), and the three association cortices (in middle gray). All sensory areas are represented by squares, all motor areas by circles, and the limbic area by an ellipse. Red arrows show only the feedforward connections in information-processing. Thickness of arrows only has a graphical value. Note that arrows go from the primary sensory areas to the respective unimodal sensory areas. From the latter they go to all the three association (posterior sensory, anterior motor, and limbic) areas. Also note that arrows go from the frontal cortex to the premotor motor cortex, and from this to the primary motor cortex. Adapted from [SAPER *et al.* 2000]. (This figure is reproduced in color in the color plate section.)

- The auditory cortex (areas 41 and 42), located in the Heschl's gyrus (temporal lobe). The correspondent *unimodal (secondary) sensory association areas* are<sup>26</sup> [Fig. 13.7]:
- The somatosensory cortex, located in the posterior parietal (parietal lobe).
- The visual cortex, located in the inferolateral surface of occipital (occipitotemporal lobe).
- The auditory cortex, located in the superior temporal gyrus (temporal lobe).

The *multimodal association areas*, where schemata and categories are produced and complex purposeful movements are contributed to, is the posterior multimodal sensory integration (including visuospatial localization, language perception, attention), located in the junction between the lobes (parietotemporal lobe). This is the place for processing at the highest level the sensory information coming from the thalamus, which is the gate between the brain and the rest of the CNS, and going through the primary and secondary sensory cortices.

<sup>26</sup>[LEDOUX 1998, pp. 198–200].



**Fig. 13.6** Sensory-cognitive, emotional, and motor-decisional systems. The green arrows depict the feedforward direction (descending for the motor system and ascending for the sensory system) while yellow arrows represent the feedback direction (ascending for the motor system and descending for the sensory system)—see also [FUSTER 2003, p. 109], showing a significant forgetting of the emotional system. The red arrows stand for the horizontal connections between these two major systems. The emotional system is totally different since it represents a relatively closed circuit (the blue arrows), in the sense that is not directly connected with systems outside of the brain, whose inputs and outputs rely on the other two brain systems as well as on the hormonal system. Moreover, it is a much more heterogeneous system because it is bound to metabolic functions. For evidence of direct connections between primary motor and sensory areas see [TERADA *et al.* 2007]. Although the spinal cord is part of the CNS, it is also the joint between the rest of the CNS and the body.

A comparison with Fig. 12.1 is highly instructive, since it shows that in organisms with a brain the feedback between sensory and motor systems is *internalized*, even if it still plays a major role in the organism's environment. (This figure is reproduced in color in the color plate section.)

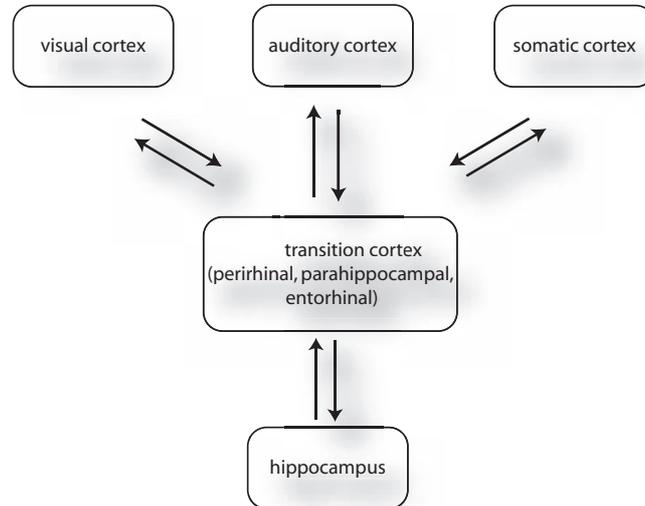


Fig. 13.7 Creation of unimodal percepts.

(2) With regard to the *motor system*, we have the reverse (top-down) path:

- *Anterior multimodal motor integration* (including motor planning, language production, judgment), located in the prefrontal cortex (Brodmann's areas 8, 9, 10, and 46), rostral to premotor areas on dorsal and lateral surfaces (frontal lobe).

The prefrontal cortex (the higher-level part of the motor-decisional system) can in turn be divided into three main parts [Fig. 13.8]: (1) The cortex around the sulcus assures the behavioral continuity for executing a task. It is connected to the working memory, that is, the memory used for the task at the hand. (2) The region ventral to the principal sulcus stores information in the working memory about the what of the object [Sec. 4.3]. (3) The region dorsal to the sulcus holds information about the where of the object. These two pieces of information are also combined in the prefrontal cortex, which is organized hierarchically from the most anterior part going back [Subsec. 5.1.4].<sup>27</sup>

- This multimodal motor association cortex sends information to the *premotor (secondary) system* (for motor preparation and programs), located in the rostral to primary motor cortex (frontal lobe, area 6). Part of the premotor area is the supplementary motor area that is very important for initiation of motion, and whose lesion provokes abulia, i.e. the inability to initiate motion.
- Finally, the *primary motor cortex* (area 4) is located in the precentral gyrus (frontal cortex). Recall that the cerebellum is of fundamental importance for producing patchwork maps [Subsec. 5.3.1] that are then integrated into true sensorimotor schemata in the primary and secondary motor cortex. I stress that, when ascending to higher representational and motor areas, there is an increasing *independence* of external stimuli and therefore a growth of top-down effects.

<sup>27</sup>[JEANNEROD 2009, pp. 131–3].

## 434 The Brain as an Information-Control System

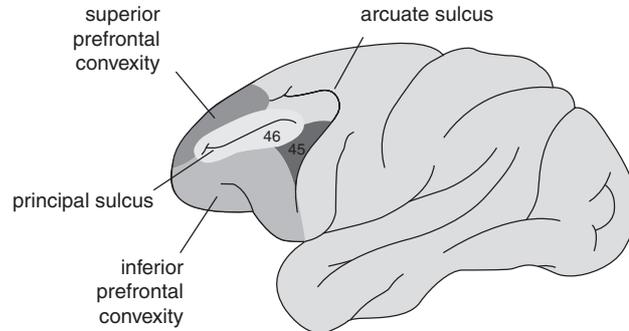


Fig. 13.8 The monkey prefrontal cortex. Adapted from [SAPER *et al.* 2000].

A. Lur'ia, followed by T. Shallice, pointed out that the high motor portion of the frontal lobes is a superstructure above all other parts of the central cortex, performing a far more general function than the other ones, namely as a sort of supervisor.<sup>28</sup> This is not a mystery, since the brain as a whole is an overdevelopment of the selective and decisional system of the organism, so that it is quite normal that the center for planning and steering is inside the brain, a sort of brain inside the brain [Subsec. 12.3.1]. We shall also consider in the third part of the book the relevance of planning and decision-making, and the pragmatic aspects of symbolization in humans. For now, I would only like to reproduce the splendid words of Lur'ia, who vividly describes the purposeless behavior of a dog whose frontal lobes have been destroyed: "It responds to all irrelevant stimuli; then it sees leaves which have fallen on the garden path, it seizes them, chews them, and spits them out; it does not recognize its master, and is distracted by all irrelevant stimuli; it responds to any element of the environment by uninhabitable orienting reflexes, and its distractions by these unimportant elements of the environment disturb the plans and programs of its behavior, making it fragmentary and uncontrolled".<sup>29</sup>

- (3) The *emotional system*, the basic understanding of which we mainly own to Papez and Maclean<sup>30</sup> [Fig. 13.9], is the most difficult brain subsystem to treat. To a certain extent it is a closed circuit integrating the limbic system and taking information from as well as giving information to the sensory and motor systems. Its business is the self-regulation and integration of mammals' psychological activity.<sup>31</sup> It consists itself of three essential layers [see also Fig. 3.12]:

- The brain-stem part represented by the hypothalamus.
- The limbic nuclei: Amygdala, hippocampus,- hippocampal formation, septal nuclei, fornix.
- The association limbic area, consisting of the cingulate gyrus (which also has a cortical part), and other cortical parts like the parahippocampal gyrus and insula.

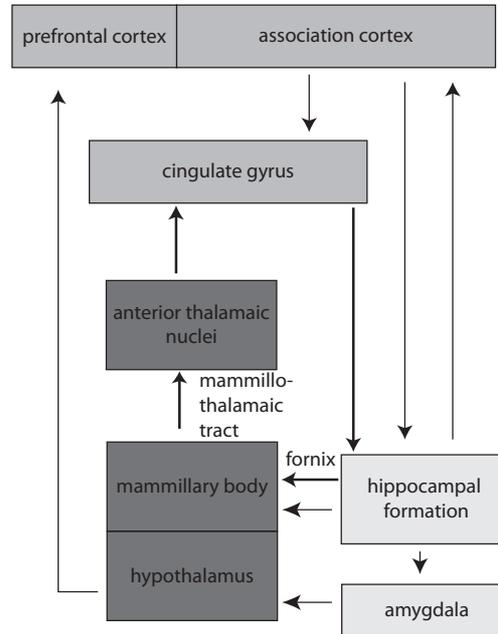
Very schematically, we have a circle between the cortical regions, the limbic nuclei, and the hypothalamus [see also Fig. 13.6]. One must distinguish between the role of the amygdala and that of the hippocampus; this will be clarified later on. For the time being, it is important

<sup>28</sup>[LUR'IA 1973, p. 89] [SHALLICE 1988, pp. 328–52]. See also [JEANNEROD 2009, pp. 136–9].

<sup>29</sup>See also [PASSINGHAM 1993, pp. 233–4].

<sup>30</sup>[PAPEZ 1937] [MACLEAN 1955]. See also [IVERSEN *et al.* 2000b].

<sup>31</sup>[SIEGEL 1999, pp. 239–75].



**Fig. 13.9** The emotional-system circuit as proposed by J. Papez and revised by P. Maclean: The limbic nuclei (dark gray), the thalamic (brain-stem) complex (light gray), and the cortical areas (medium gray). Adapted from [IVERSEN *et al.* 2000b, p. 988].

to stress here that the hippocampus has no direct connection with the primary motor or primary sensory areas, but rather its connections are with the association areas.<sup>32</sup> This has been considered as evidence that the hippocampus is important only for the formation of new associations, while native or old ones do not need its contribution.

Summing up, the frontal cortex is responsible for the attentional and decisional mechanisms, the sensory posterior complex for processing and integrating sensory information, and the limbic system (especially the amygdala) for affective and emotive reactions. Therefore, the brain and the CNS can be divided into two longitudinal halves: A posterior one for reception and an anterior one for action; in the middle is the emotional system. The dividing line is represented by the central fissure and in part by the sylvian fissure.<sup>33</sup>

The whole brain system is highly integrated and interactive [Fig. 13.10]. Here, I shall provide an example of integration: Objects are manipulated by humans for obtaining geometric or force cues. Force cues, in particular, are decisive for understanding the shape of objects.<sup>34</sup> In other words, information (such as shape) that is ordinarily acquired with sight can also be acquired through touch. According to this examination, we may consider a weaker form of modularity rather than the traditional one [Sec. 3.6] and say in accordance with A. Damasio<sup>35</sup> that what may be localized in the brain are *convergence zones*. In other words, I assume that modules are micromodules at the purely functional level, that is, not concerning the systems themselves but the functions (behavior,

<sup>32</sup>[FUSTER 2003, p. 46].

<sup>33</sup>[FUSTER 2003, p. 123].

<sup>34</sup>[ROBLES/HAYWARD 2001].

<sup>35</sup>[DAMASIO 1989].

## 436 The Brain as an Information-Control System

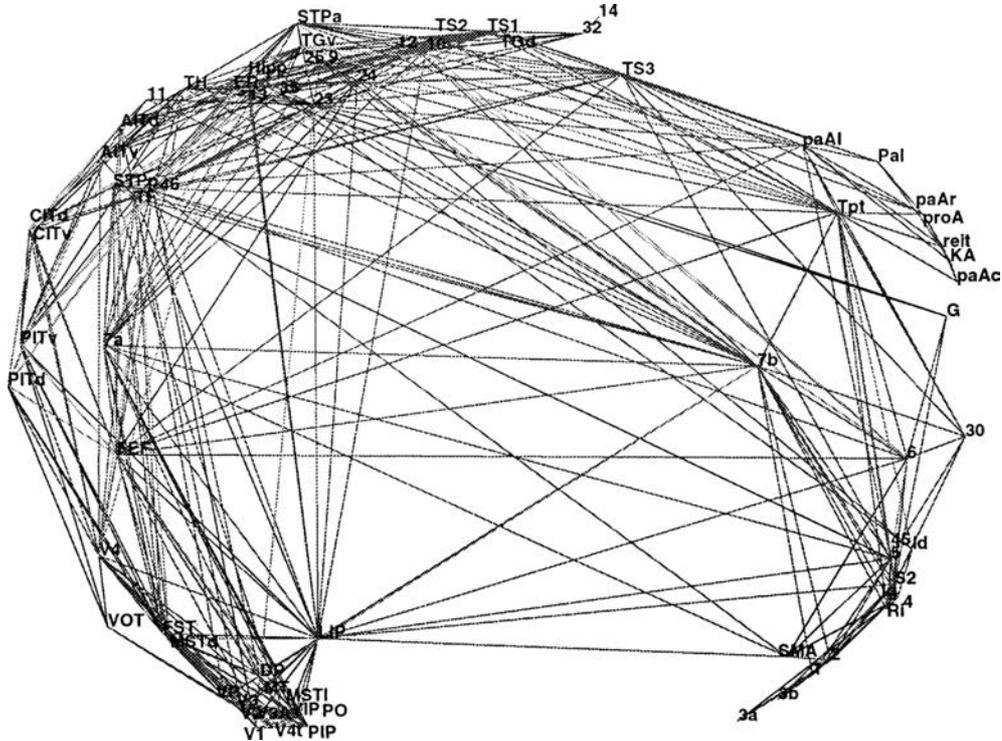


Fig. 13.10 The brain's hierarchical connections. 72 cortical areas of the macaque brain show 758 connections (of which 136 are one-way). The position of each area is specified as that which minimizes the distance from other areas to which it is connected. See in particular at the top left the temporal anterior multimodal area (STPa), and a little bit on the left and in a lower position, the posterior multimodal area (STPp). These are hubs. Another very important center is 7b (with many connections to other sensory systems). All cortical connections are characterized by convergence, divergence, recurrence. Adapted from [MOUNTCASTLE 1998, p. 265].

learning, memory) at hand. They are formed according to pure Hebbian and connectionist rules [Secs. 3.8 and 12.4], and express, at the start, a process of mere habituation: Areas that have worked together several times continue to do so if there are no counter-signals.

Many disturbances of the higher functions of the nervous system (aphasia, apraxia, agnosia) are due to a disconnection of primary receptive and motor areas.<sup>36</sup> In lower mammals these connections are direct, while for higher mammals they are developed through the associative cortex. This means that a disruption of this area can disconnect sensory and motor functions. The connections in the visual association regions have been extensively analyzed and it has been shown that the major outflow from these regions is to the lateral and basal neocortex of the temporal lobe, which in turn connects to limbic structures. Lesions of the lateral and basal temporal lobe lead to a failure in

<sup>36</sup>[GESCHWIND 1965].

activating limbic responses (fight, flight, sexual approach). In other words their affordances become lost. However, it has been pointed out that lesions in the visual area are also relevant.<sup>37</sup>

### 13.3 The Three Information Systems of the Brain

We have seen that like an organism the brain too can be conceived as being organized into three functional systems [Figs. 13.5–13.6]. These systems, constituting the information-control system in mammals, are a further development of the basic ones: A processor, a regulator, and a decider [Sec. 7.2].

The sensory, motor, and affective subsystems are *white boxes*, that is, containers that are empty apart from the pure, generic operation they perform; in other words, boxes that are not opaque but contentless. In the words of J.-P. Changeux, these systems only provide the context and not the content.<sup>38</sup> Behaviorism supposed that the brain was a black box in the sense that it was an inaccessible entity. Nowadays, on the contrary, the brain is studied in its structures and functions<sup>39</sup> even if we are only able to assign very general actions to single subsystems and therefore they are in themselves very undetermined:

- The role of the emotional system is only to provide a general *arousal state*, which as such can be interpreted or give rise either (a) to an ample range of emotional *behaviors*<sup>40</sup> (thanks to the interaction with the decisional system) or (b) to the impression of emotionally charged memories on the brain (this is in collaboration with the cognitive system).<sup>41</sup>
- The decisional system in its isolation can only provide a general *attentional (or alert) state*, which in order to be determined must somehow produce either (a) learning together with the cognitive system or (b) purposive behavior in collaboration with the emotional system.<sup>42</sup>
- Finally, the cognitive system is in itself a pure information processor, not very differently from a computer, without any “understanding” of the information it treats, not to mention its value. It can evaluate this information as a contribution to learning or to memory, but only in collaboration with the (a) decisional and (b) affective systems, respectively.

When the single systems or subsystems are considered in themselves (therefore as white boxes), they seem (and to a certain extent even are) monadic, and sometimes even act as such, especially in the most elementary and automatic processes. When they collaborate with another subsystem, they are interactive and in a dyadic relation. When the whole system represents a single cybernetic circle—which can only happen in the highest form, in symbolic activities—the brain is a single triadic entity [Subsec. 2.4.4].<sup>43</sup>

This proposal seems to catch the fundamental results of Sec. 13.1. As far as I can understand, against it one could make use of the fact that the architecture of the brain seems to be too complex to be reduced to such a tripartition. It seems even to be too difficult to speak of emotions or of cognitive functions in general, given the huge variety of activities displayed by

<sup>37</sup>[LEHRMITTE/BEAUVOIS 1973].      <sup>38</sup>[CHANGEUX 2002, p. 79].      <sup>39</sup>[VAN ESSEN *et al.* 1994].

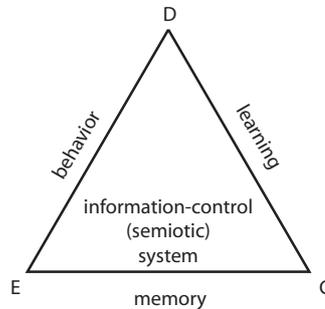
<sup>40</sup>[LEDOUX 1998, pp. 128–37].

<sup>41</sup>Panksepp acknowledges that one can never capture innate emotional dynamics in its pure form [PANKSEPP 1998, p. 26].

<sup>42</sup>[WUNDT 1907, Sec. 15].

<sup>43</sup>Damasio was one of the first scholars to understand the relevance of the emotional system for human cognitive abilities, especially when choice is involved [DAMASIO 1994, DAMASIO 1999].

## 438 The Brain as an Information-Control System



**Fig. 13.11** Relations between the sensory-cognitive (C), decisional (D), and emotional (E) systems. The activity in which the emotional and the decisional-motor systems cooperate is behavior. The activity in which the emotional and the sensory-cognitive systems cooperate is memory. Finally, the activity in which the sensory-cognitive and decisional-motor systems cooperate is the non pure-associative learning. See also Fig. 5.10.

the brain.<sup>44</sup> Thorndike said<sup>45</sup> that the mind and the brain may consist in connections. In this way, Thorndike confounds the systems with the functions displayed by the organism. To establish connections is the office of learning and to fix them is the job of memory, as we shall see: Both of them are brain *functions*. Therefore, in order to avoid misunderstandings, I think that it is necessary to distinguish between system and function:

- The *systems* are the decisional, emotional, and cognitive ones. They have a solid biological and neural basis. These three systems are general-purpose activity poles.
- In my opinion, the diversification of the brain's activities or *functions* is produced in the feedforward and feedback connections *among* the three subsystems considered here.

This proposal is also a way to take into account the concept of function without detaching it from structural considerations [Subsecs. 6.1.4 and 8.2.4–8.2.5], and leans in the direction proposed by Panksepp,<sup>46</sup> who felt the necessity to distinguish between coherently functioning psychobehavioral “organ systems.”

### 13.4 Concluding Remarks

In this chapter, I have stressed that the brain does not code information but only *receives* codified information from the peripheral sensory systems:

- The brain is stratified following a phylogenetic history in: The reptilian brain, centered on fundamental perceptual and nonreactive sensory-motor schemata; the mammal brain, centered on the visceral-emotional system and active sensorimotor schemata; and the neo-mammal brain focused on high-level planning.
- Mammals (and in part birds) are characterized by care of their infants and play.
- The previous three fundamental levels in part superpose with the three main systems of the brain: The sensory system, the emotional system, and the motor system. The sensory system

<sup>44</sup>Such a point of view has been authoritatively supported in [LEDOUX 1998].

<sup>45</sup>[THORNDIKE 1931, pp. 121–2].

receives information from the periphery and sends it to the primary sensory areas, and from here to the unimodal association cortex and then up to the multimodal association area. The motor system gives its inputs starting from the multimodal area down to the premotor area until the primary motor area, where through very complex pathways the signals are traded to the motor apparatus. The emotional system receives and gives contributions to the other two systems.

- The three systems in their isolation are white boxes that are unable to produce specific functions. In order to give rise to behavior, learning, or memory, at least two of them must cooperate.

The synthesis in this chapter highlights an important dialectic between functional segregation and functional integration in the brain. These complementary aspects of functional brain architectures have been a key guide in the use of human brain mapping to understand functional anatomy. Functional segregation refers to the modular or segregated functional specialization of various cortical and subcortical systems, whereas functional integration refers to the distributed integrations among these areas mediated by extrinsic axonal processes. This chapter has highlighted the fundamental role of functional integration in the context of a broad-scale segregation of anatomical structures and functions. In particular, the main integrated and higher functions of the mammal brain turn out to be behavior, learning, and memory. Actually, although eminently functions of the brain, they involve the whole organism and its relations with natural and social surroundings. In a less complete sense these functions are already present in lower animals. As mentioned, these functions can be displayed when at least a couple of brain systems are connected and interact. In such a case, I would speak of the activity of the motor, sensory and emotional systems and of the functions of behavior, learning, and memory. The higher functions can be schematized as in Fig. 13.11. In the next chapter, we shall consider the three basic brain systems.

# 14

## Decisional, Emotional, and Cognitive Systems

---

In this chapter we shall study the three systems of the mammal brain, the decisional, the emotional, and the sensory systems. The latter is considered this time in its fundamental dynamism. Cerebral systems have proper procedures, which do not possess survival value (have no vital meaning). The concept of system turns out to be a sophistication of the notion of structure. As such a system does not need to be localized but can be distributed across several brain areas. Instead, functions (like behavior, learning, and memory) are (by definition) aimed at satisfying vital needs [Sec. 12.2 and Subsec. 12.5.3].

### 14.1 The Decisional System

#### 14.1.1 Attention

I assume that attention is the empty *alert state* thanks to which we enhance perception and the further elaboration of certain items and decrease or suppress elaboration of other items [Sec. 13.3]. As I have stressed, while the iconic aspect of representation consists of the continuous wave-like production of structures that can be associated with events and objects, the attentional aspect, which is strictly related to the referential act, represents the discontinuous, point-like feature of the representational activity and is also the connection with movement and action<sup>1</sup> [Sec. 12.2 and Subsec. 12.5.1]: We can speak of an internal representational space and maps when we deal with the iconic aspect of representation, whereas attention is always concerned with external objects (in the case of humans this is done even with abstract objects) and events. Evidence for this character of attention comes from studies showing that attention from one object to another does not sweep along a space in a continuous way but rather moves ballistically from one location to another.<sup>2</sup> Further evidence for the point-like nature of attention is the following: It is impossible for a human to fix her attention on the same unchanging object or situation for a long time. For instance, after 300 msec, humans disengage attention from a given situation, so that, if a certain stimulus is repeated after this time window, a human subject needs additional time to come back to the previous attentional situation; an effect called inhibition of return.<sup>3</sup> However, attentional states may be accompanied by representations, and in this case they also go together with some wave-like

<sup>1</sup>One has spoken here of a central supervisory activating system [BADDELEY 1990, pp. 91–101].

<sup>2</sup>[KWAK *et al.* 1991].

<sup>3</sup>[POSNER/COHEN 1984] [WILLINGHAM 2001, pp. 133–4].

activity: They are in particular characterized by an increase in gamma rhythm<sup>4</sup> [Subsecs. 3.3.3 and 12.5.1; see also Subsec. 6.5.3].

Therefore, attention is the hypothetical agency or activity which produces some selectivity or has the central facilitation of a perceptual activity<sup>5</sup> [Sec. 3.5]. Attention increases the amplitude of orientation tuning curves without systematically changing their width (length) [Sec. 1.2]. These findings suggest that attention selectively strengthens the responses of the cortical neurons that convey behaviorally relevant sensory signals without changing their stimulus selectivity.<sup>6</sup> It is therefore a *supplementary selection*: Attention may alter the cortical representation of the visual scene in a way that is equivalent to increasing the contrast (the differences in intensity) of those stimuli currently having behavioral relevance.<sup>7</sup> Evidence for this is represented by the fact that there can be (bottom-up) intrusion of unattended information in an attentional state, which shows that we must also somehow process the apparently discarded information although with less stress.

There is also evidence that an increase in alertness improves the speed of processing events. The trade-off between improved speed and reduced accuracy of warning signals implies that alerting does not act to improve the build-up of information concerning the nature of the target but, instead, acts on attentional states to enhance the speed of actions taken toward the target. This is very important, as we shall see, for the issue of whether consciousness is quick or slow.

The previous findings find anatomic support. Anatomically, we may distinguish between the source of attention (those areas that are specific to attention) and the site of attention (where the computations involved in the task are usually performed). The latter deals with contents and receives inputs from other brain systems. Attention, at least in humans, is often employed to control cognition independently of the effector systems. Searching memory, combining thoughts, selecting information for storage, and monitoring behavior are important attentional operations that require no effector system. As a result of the activity within the attention network, the relevant brain areas will be amplified and/or irrelevant ones inhibited, leaving the brain to be dominated by the selected computations. If this is the correct theory of attentional control, one would expect to find the source of attention to lie in systems widely connected to other brain areas but without a unique and uniform structure. As pointed out by Goldman-Rakic, this indeed appears to be the basic organization of frontal midline networks. Anterior cingulate connections<sup>8</sup> to limbic, thalamic, and basal ganglia pathways would distribute its activity to the widely dispersed connections we have seen in order to be involved in cognitive computations.

An important point is to distinguish between passive and active attention<sup>9</sup>:

- *Passive attention* is generated by the effect of external stimuli and is always reflexive, immediate and bottom-up [Fig. 8.4 and comments]. For instance, an unexpected impediment of an automatic activity in general makes the subject aware of this activity,<sup>10</sup> as witnessed by the human language.<sup>11</sup> The stronger the stimulus, the more peaked the attentional reaction is.<sup>12</sup> It is here that we have contributions coming from the emotional system.
- *Active attention* is both autonomous and top-down, but it is always derived since we never attend to an object except for the sake of a remote interest or end. Past experience will then

<sup>4</sup>[ENGEL/SINGER 2001].      <sup>5</sup>[HEBB 1949, p. 102].      <sup>6</sup>[MAUNSELL/MCADAMS 2000].

<sup>7</sup>[MACALUSO *et al.* 2000a, MACALUSO *et al.* 2000b, MACALUSO *et al.* 2002].      <sup>8</sup>See also [HAN *et al.* 2003].

<sup>9</sup>[JAMES 1890, v. I, p. 402–58].      <sup>10</sup>[FOURNERET/ JEANNEROD 1998].

<sup>11</sup>[VYGOTSKY 1986, pp. 30, 48–9].      <sup>12</sup>[HERBART 1824–5, Par. 47].

**442** Decisional, Emotional, and Cognitive Systems

be condensed in the motives generating active attention. Therefore, we have contributions here coming from the cognitive system.

This distinction is also justified by neurological studies in humans<sup>13</sup>: It is indeed possible to distinguish the type of active attention needed to segregate the target from a more general awareness of the background. Damage to the attention network involving the parietal lobe and associated thalamic areas produces a kind of loss in focal awareness that is needed for active attention. In particular, neglect induced by parietal lesions may leave the patient unconscious of this lack of awareness, just as the split-brain person is unaware that the visual world has lost integration.

Therefore, attention involves both top-down purposeful or voluntary processes and bottom-up reflexive mechanisms.<sup>14</sup> Brain states receive inputs from error units located both at the same level and at lower levels of the brain's hierarchy while error units receive inputs from both the same level and above.<sup>15</sup> This mechanism is fully in accordance with the complex systems' dynamics [Subsec. 6.3.2]. Error units are responsible for feedforward connections arising from sensory inputs in the brain, while feedback connections are forms of Bayesian (proto-)inferences allowing for error correction [Subsec. 7.6.2 and Sec. 12.2]. Evidence of active top-down processes has been found in rats during the exploration of their environment: In this case there is an endogenous production of dopamine (that is, not depending on external stimuli) but recruited as a consequence of those attentional processes.<sup>16</sup>

Therefore, when some searching (for instance, in the case of exploring behavior) is activated we can also speak of a programming-attentional network.<sup>17</sup> This system displays two important overall functions:

- (i) It is informed about the processes taking place within the organism (proprioception [Subsec. 5.3.2]). A system that would be related to our experience of *focal attention* would clearly play this function about a subset of current (sensory) and stored (memory) information. There are reasons for relating anterior cingulate activity to focal awareness of the target.
- (ii) A second function is to exercise some *control* over the organism. Again, the anatomy of the anterior cingulate cortex provides pathways for connecting it to both the posterior parietal area and the anterior areas that are active in humans during language tasks.

While it is important to distinguish between attentional top-down processes (requiring focal attention) and cognitively induced top-down processes,<sup>18</sup> it is also crucial to consider that it is not always possible to know which of the two processes has taken place or which has been predominant in determining what we see. In normal conditions, the two processes operate together, but in extreme conditions they can go on almost independently from each other.

Summing up, the attentional processes carry at least three functions<sup>19</sup>

- (1) Maintaining the alert state (this is the generic attentional state),
- (2) Becoming aware of relevant events (passive attention),
- (3) Orienting the sensory stimuli, especially location in visual space (active attention).

<sup>13</sup>[POSNER 1994]. <sup>14</sup>[GAZZANIGA *et al.* 1998, pp. 244–99].

<sup>15</sup>[FRISTON/KIEBEL 2009]. <sup>16</sup>[KANDEL 2006, pp. 314–15].

<sup>17</sup>[POSNER 1994]. See also [GREENWALD 1992]. <sup>18</sup>[TREISMAN/GELADE 1980]. <sup>19</sup>[BAARS 1997a].

### 14.1.2 Decisions

The decisional-motor system does not only give rise to an empty alert state but should be able to produce decisions. Any decision is a break in the continuous (wave-like) chain of cerebral or mental activity, which, in its autonomy, is rhythmic [Subsec. 3.3.3]. I wish to point out that the brain both operates as a closed, self-referential system like the heart and also as an open system.<sup>20</sup> In its information-processing activity, when it builds iconic patterns constituting representations, it is like a closed system. As a decisional system it is both open and closed,<sup>21</sup> trying to catch external inputs and create expectations, respectively. As said, the break of the brain's rhythm can come from outside but also from inside, in the way the decisional system manages the received information.

One of the biggest difficulties when dealing with cognitive abilities is to understand this point correctly. We have already seen that in quantum mechanics the final detection event (the decider) is random [Subsec. 2.3.2]. Mostly, one takes the source of any information acquisition, and therefore any information processor, to be random. This is a mistake. Relative to any form of information acquisition, the source need only be *unknown* and not random. On the contrary, the selection that is performed when the information is finally acquired *always contains a random element* (at least relative to the source), otherwise it would not be a selection among several possible different messages, but reducible to a mechanical action (or a manipulation) of the source on the receiver. For instance, the cell's membrane is a piece of physics recruited by the organism without any connection to the genetic information processor [Subsec. 7.6.1]; we have also seen that during epigeny, in cell signaling the final reaction of the receptor depends on the receptor cell's state [Subsec. 11.3.1].

To understand this crucial point, let us come back to Wiener's contribution<sup>22</sup> [Subsec. 3.2.2], according to which, one can classically prepare a system in a certain way and then measure it, but not *vice versa*. The reason is that *selection follows preparation* [Subsec. 1.3.1], and the very act of selection always consists in an actual reduction out of a space of possibilities. In order to reverse this sequence, we should spread a single event in a multidimensional possibility. The result would be the annihilation of the event as such, i.e. the transformation of its specificity (and space-time localization) to a sort of generality. This would however contradict both thermodynamics and information theory, since in any transmission of a signal the entropy grows, which implies that a selection process always results at the end of the information exchange, even in terms of a mere dispersion producing a loss of a part of the initial information content [Subsec. 2.3.1 and Sec. 12.6]. In this sense, the purposive or goal-directed act of selection by the receiver represents only one of the possible "reductions" of the initial amount of information. In other words, due to this irreducible function many-to-one that characterizes any relation between initial information-processing and final information-selecting and to the irreversibility of such a connection, any decision has an intrinsic random component, which is the result out of a set of many possibilities.

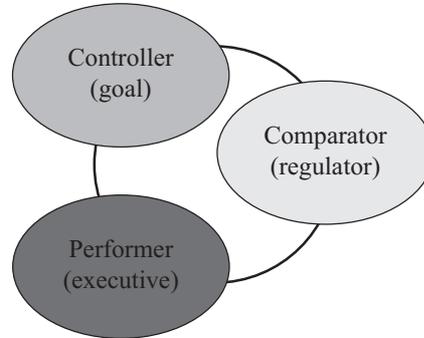
Obviously, for very elementary life processes, the decider has a strong random component, as it happens in chemotaxis [Subsec. 8.3.1]. It is only with the growing capability by the organism to control information, which also implies growing complexity [Sec. 9.10] that the decider, working together with other subsystems of the organism (namely the regulator and the information processor), can support its selection process in a way that makes it less random. However, the fact remains that a random component is deeply rooted in any decisional process and can never be fully effaced; this is the case even for higher organisms. This random aspect, when considered together with the separation between self and non-self [Sec. 8.4] is what ultimately makes any decisional

<sup>20</sup>[LLINÁS 2001, pp. 1–51] [CHANGEUX 2002, p. 25].

<sup>21</sup>[JEANNEROD 2009, p. 86].

<sup>22</sup>[WIENER 1948].

## 444 Decisional, Emotional, and Cognitive Systems



**Fig. 14.1** Feedback circuit Controller–Comparator–Performer. This is the general form of any decisional system. Note that, though the system runs in a specific direction (here, counterclockwise), each of the subsystems is either directly or indirectly sensitive to the other two through the action of the comparator: Indeed, the comparator receives inputs from both the performer and the controller and sends a feedback signal to the latter.

process performed by organisms not dependent on external sensory inputs and is therefore truly *endogenous* representing the root of the antifeedback mechanism through which bacteria also actively act on or resist to the environment [Subsec. 8.2.1 and Sec. 12.2].

#### 14.1.3 Planning, Programming, and Executing

The cognitive permeability—whose necessity has been stressed by Pylyshyn [Subsecs. 3.7.2 and 6.1.2]—of the neural motor network shows that it cannot be totally reducible to its physical (anatomical and neurophysiological) properties.<sup>23</sup> Then, we must postulate that there are programs ruling the different movements. R. Held<sup>24</sup> assumed that the same programs that were at the origins of active movements were also used as representations of the relations between the body and its environment. If it is true that in the CNS there are exact formulae of movements (the so-called *engrams*) and a certain autonomy of these motor programs must be acknowledged,<sup>25</sup> it is also clear that high-order factors such as the goal, the context, and the probable knowledge of the result of the action seem to be able to influence not only duration and velocity but also the intrinsic kinematic structure of the movements [Sec. 5.3]. The organism is a system aiming at a homeostatic equilibrium and can do this only through reference to—and (through the sensory–cognitive system) representation of—its environment [Subsecs. 7.6.2 and 8.2.7]. Therefore, while the decisional system can probably generate only a generic attentional or alert state, in order to operate true decisions it is necessary to focus the attention on specific items implying the contribution of the cognitive and emotional systems.

In other words, the decisional system cannot be taken to be a fully independent programming instance. The necessity of feedback and control in organisms and especially in motor systems arises from the fact that biological systems are non-linear systems acting in a complex environment, and therefore the effects of their actions cannot be perfectly foreseen<sup>26</sup> [Fig. 14.1]. Von Holst and Mittelstaedt<sup>27</sup> proposed that each time there is an outflow of information for producing a movement, a copy (the *efferece copy*) is sent to the short-term memory (which I shall elaborate

<sup>23</sup>[JEANNEROD 1988, pp. 1–40].

<sup>24</sup>[HELD 1961].

<sup>25</sup>[JEANNEROD 2009, pp. 54–9].

<sup>26</sup>[JEANNEROD 2006, pp. 16–21].

<sup>27</sup>[VON HOLST/MITTELSTAEDT 1950].

on in the next chapters). Reafferent sensory feedback plays a role in the comparison process: The reafferent signal (the feedback) generated from the movement is *compared* with the efference copy. We see here again the principle that, apart from the issue of the source, there is no distinction between endogenously generated or external, stimulus-evoked representations<sup>28</sup> [Subsec. 3.3.1]. Von Holst and Mittelstaedt's theory should be compared with the equilibrium-point theory of Feldman and Bizzi,<sup>29</sup> which focused on single points of equilibrium that need to be stepwise reached in the course of a movement execution. Again we see, in the equilibrium point hypothesis, a tendency of the brain to minimize prediction errors. In this instance, the error pertains to the distance between the current position of the motor plant and that predicted under prior expectations representing the intended target position. It is important to understand that Von Holst and Mittelstaedt's model addresses the control of complex movements while the equilibrium point hypothesis addresses single motor segments or outputs. Both aspects are relevant.

Therefore, to refine our analysis, we may distinguish between goals, intentions in action, purposes, and prior intentions:<sup>30</sup>

- *Goals* only demand a behavior that addresses the self-maintenance of the system [Subsec. 8.2.2] and can be found at any level of semiotic activity.
- *Intentions in action* are those goals that tend to give rise to a coordinated sequence of actions and single movements with a practical and immediate result. We may assume that reptiles are fully able to work at this level [Subsec. 13.1.3]; when those intentions express, or are related to purposeful and interactive behavior,<sup>31</sup> they require mammal complexity [Subsec. 12.7.2].
- *Prior intentions* are connected with true plans and demand in general abstract aspects and items that are not (neither spatially, nor temporally) immediately present. This level, at least in a very rudimentary form, is present in nonhuman primates.

The idea of purposeful action in humans (as well as in other mammals) imposed itself when it was observed that muscular contractions in response to stimulation were not localized in single muscles but consisted in the coordinated contraction of several muscles.<sup>32</sup> As we have noted, Hebb arrived at the conclusion that behavior cannot be produced by a reaction to what was previously occurring in the sensory system [Subsec. 5.3.2]. Intentions in actions already determine motor equivalence classes (although the single components of motor sequences are hardwired and ruled by engrams), and this is much more true where there are interactive sensorimotor schemata relying on back-reactions of other organisms.<sup>33</sup> For this reason, as already mentioned, Hebb considered attention as being able to design the activity that controls the form, speed, strength, or duration of a response, and does not depend only on previous excitation or reception.<sup>34</sup> This was also remarked by J. J. Gibson,<sup>35</sup> who examined the possibility of a central autonomous process (that is, relatively independent from stimuli). In the words of Hebb, attention shows that a factor not coming from the sensory system must be present in the cerebral action, even though he left the issue of what this could be as open.<sup>36</sup> On the other hand, the notion of prior intentions is quite important here and reprises the notion of Bayesian inference, both from the free energy treatment of informational selection at the cellular level and in terms of hypothesis testing as a framework for understanding decisions

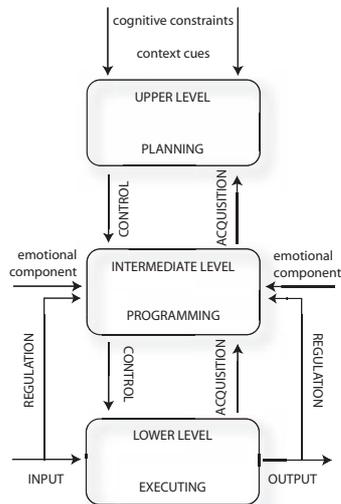
<sup>28</sup>[MATURANA 1970, pp. 22–23].      <sup>29</sup>[FELDMAN 1966] [BIZZI *et al.* 1971].

<sup>30</sup>[SEARLE 1983, pp. 79–111]. Prior intentions display a form of intelligent top-down causation (demanding a symbolic level) that should be kept distinct from goals (non-adaptive information control in Ellis' language) and intentions in action (adaptive information control in Ellis' language) [ELLIS 2008a].

<sup>31</sup>[HEBB 1949, pp. 144–45].      <sup>32</sup>[JEANNEROD 1999a].      <sup>33</sup>[MENZEL/HALPERIN 1975].

<sup>34</sup>[HEBB 1949, pp. 4–7].      <sup>35</sup>[GIBSON 1941].      <sup>36</sup>[HEBB 1949, p. 79].

## 446 Decisional, Emotional, and Cognitive Systems



**Fig. 14.2** Jeannerod's distributed model for the generation of actions. It is likely that inputs mediated by the affective system also go to the intermediate level. Adapted from [JEANNEROD 1999a, p. 62].

[Subsec. 7.6.2]. In the present context, it suggests that the genesis of purposeful movements could, in part, be ascribed to prior expectations (intentions) about proprioceptive signals. Movement or action can then be seen as suppressing the ensuing prediction errors, in accord with the equilibrium point theory above.

A study of Shadlen and Newsome showed that the activity of neurons in the lateral intraparietal region (LIP) precedes the motor decision a monkey will take.<sup>37</sup> In the context of a discrimination task, the decision process is simply a mechanism whereby sensory information is evaluated in order to guide the selection of an appropriate motor response: The primary finding is that neurons in LIP carry signals that predict the decision a monkey will make in a two-alternative (forced choice) direction discrimination task. These signals typically arise early in the trial during the presentation of the random dot stimulus and are sustained during the delay period following the disappearance of the stimulus. Thus, predictive activity can arise several seconds in advance of an eye movement that indicates the monkey's decision. The data also suggest a neural process that integrates weak, slowly arriving sensory information to generate a decision. The coherent motion signals are distributed randomly throughout the stimulus interval. When coherent motion is strong, a substantial amount of motion information arrives quickly and decisions can be formed earlier in the trial and with greater certainty.

The whole decisional system, at least in humans and in a very rudimentary form in primates, can be represented as in Fig. 14.2, as a planning-programming-executing system. Cognitive constraints are exerted on planning (in the seat of the sensorimotor prefrontal integration system) while emotional inputs contribute to the intermediate level: Programming [see also Fig. 5.10]. Programming happens in the premotor cortex and receives inputs from three main sources<sup>38</sup>

<sup>37</sup>[SHADLEN/NEWSOME 1996].

<sup>38</sup>[SAPER *et al.* 2000].

- (1) The motor nuclei in the ventroanterior and ventrolateral thalamus that receive inputs from the basal ganglia and the cerebellum. The cerebellum provides the basic maps that are used for forming segmental motor schemes [Subsec. 5.3.1].
- (2) The primary somatosensory cortex and the parietal association cortex. This is the reafferent information about the ongoing motor response as well as from the emotional sphere [Sec. 13.2].
- (3) The prefrontal association cortex, the upper-level planning system.

The motor aspects are crucial for life and I recall that the brain itself has originated from a selective-decisional system [Subsec. 11.4.4]; even representations find their evolutionary source in nonreactive actions [Subsec. 12.3.1].

## 14.2 The Emotional System

As mentioned, my main argument here is that the emotional system can only produce a generic *arousal state*,<sup>39</sup> which is then specified according to the crossing contributions of the cognitive and decisional systems [Sec. 13.3]. The arousal state is characterized by specific bodily reactions: Freezing, increase of blood pressure, production of stress hormones, and the startle reflex. In this sense, the basic arousal state is a derivation of a pure metabolic substrate and, in its fundamental significance for the organisms, is a manifestation of a distress of the metabolic system that is not totally disruptive. Therefore, the emotional system is a further development, at a higher level, of the tendency of the organism to preserve its homeostasis. As this state of the metabolic system is connected to an external event (through both the peripheral and central sensory system), it becomes information relevant for the organism, and in particular a generic sign of alarm, since, as we know [Sec. 12.2], a negative feedback not having a disruptive power but rather meaning a novelty represents a hint for the organism to react appropriately.

### 14.2.1 Some General Remarks

It is interesting to observe that many of the traditional interpretations of emotional states presuppose a stimulus giving rise to a step-by-step reaction process.<sup>40</sup> Obviously, the issue here is to understand whether emotions or cognitive-sensorial aspects come first. Traditionally, it was believed that a sensorial elaboration of the stimulus comes first, and emotions follow. William James was the first scholar to propose that emotions can precede cognitive features and that somehow they embody a subsequent representational activity or interpretation in the case of humans.<sup>41</sup> This theory has been confirmed by other scholars, like Carl Lange. W. Cannon, on the other hand, pointed out that the emotional state without cognitive components is too undetermined to give rise to specific reactions.<sup>42</sup> It is true that Zajonc<sup>43</sup> showed that preferences can be mechanically elicited through mere exposure to stimuli: In other words, affective reactions can occur before and independently of cognitive processes, as predicted by W. James. However, this only shows that affect and cognition respond to two different systems that can interact and influence each other, which also seems to be true for interactions between the decisional and emotional systems:

- Indeed, at an elementary level, the connection of the emotional system with the decisional system is displayed especially by the fact that the organism shows an expectation, a perceived tendency toward external events, even before their full onset.<sup>44</sup>

<sup>39</sup>[SCHACHTER/SINGER 1962].      <sup>40</sup>[LEDOUX 1998, pp. 42–72] [IVERSEN *et al.* 2000b].      <sup>41</sup>[JAMES 1894].  
<sup>42</sup>[CANNON 1927].      <sup>43</sup>[ZAJONC 1980].      <sup>44</sup>[ARNOLD 1960].

**448** Decisional, Emotional, and Cognitive Systems

- The connection with the cognitive system, on the other hand, is basically shown in humans in the way interpretations (both memorized and current ones) determine the emotional experience.<sup>45</sup> Moreover, I remark that there are degrees in cognition, since the relations between cognition and emotions are far more complex than a two- or three-step process. Also the idea that the emotional system is characterized by automatic responses, while cognitive processes are flexible,<sup>46</sup> is not completely true, since information processing is to a certain extent automatic. As we shall see, it is only the cooperating functions and processes in different brain's systems that are flexible [Sec. 13.3].

Summarizing, I assume that the emotional system is constantly (in a feedforward and feedback process) interacting with cognitive and decisional processes. Therefore, it seems appropriate to distinguish between (a) the emotional arousal state, (b) emotional or even body reactions, which, immediately connected with the arousal state, are elementary, mostly hardwired and very rapid (for instance, a startle reflex takes less than one-hundredth of a second to start), but can be accompanied by an expectation, and (c) emotional *responses*, which are more articulated and integrated forms of *behavior*, also important for future guidance.<sup>47</sup>

*Emotions*, in the proper sense of the word, can be defined as the processes by which the brain determines or computes the value of a stimulus.<sup>48</sup> This is called *appraisal*. In this sense, the emotional system, as it happens for the attentional state [Subsec. 14.1.1], contributes to a *supplementary selection* of any stimulus as a biologically significant sign, and emotions accompany the stimulus, especially when it is particularly relevant or shocking (thus representing a novelty or at least a discrepant event). After a first evaluation, emotion is eventually followed by (1) an emotional reaction or response, or even emotional behavior, and (2) by the eventual occurrence of a feeling. Feelings occur not only in humans<sup>49</sup> and represent the *subjective experience* of emotion.<sup>50</sup> Feeling then integrates (1) a stimulus (and its cognitive or sensorial aspects), (2) longterm memories about that stimulus or similar stimuli, (3) emotional arousal by the amygdala (which starts a possible reaction or behavior).

**14.2.2 Neural Machinery**

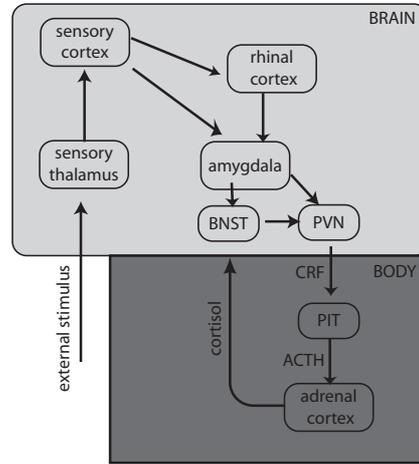
Emotions are essentially indexical [Subsec. 12.5.1], since any arbitrary event that simply *occurred* at the time and place (i.e. in the same context) of a violent emotion (and therefore became *associated* with this emotion) will thereafter provoke a *similar* emotional reaction on the subject when it manifests itself. We shall see below the role of emotions as memory amplifiers, due to this essential indexical aspect. Amygdala is crucial in this action of memory amplification, by allowing for the release of hormones by the adrenaline gland [Sec. 13.2]. In turn, the amygdala is a target of these body hormones especially when there are stressful stimuli like those provoking fear perception [Fig. 14.3], one of the most studied emotions. Thanks to its connections with the hippocampus and regions of the memory system [Fig. 13.9], the amygdala modulates the consolidation of explicit memories not only about dangerous situations (it does not necessarily reflect a state of fear and anxiety). Emotions come to monopolize consciousness, as it is evident with fear, when the amygdala comes to dominate the so-called working memory (and therefore it also makes rational thinking difficult, by allowing more automatic and unconscious reactions).

The neural mechanism underlying the emotional system is the following<sup>51</sup>: The amygdala is central for both the *reception* of emotional stimuli and the emotional *responses*. When a stimulus

<sup>45</sup>[LAZARUS 1966, LAZARUS 1984]. <sup>46</sup>[LEDOUX 1998, p. 69]. <sup>47</sup>[PANKSEPP 1998, p. 33].

<sup>48</sup>[LEDOUX 2002, pp. 206–29]. <sup>49</sup>[DAMASIO 1999, pp. 53–6]. <sup>50</sup>[CLARK 1993b].

<sup>51</sup>[LEDOUX *et al.* 1986, LEDOUX *et al.* 1990] [LEDOUX 1998, pp. 138–78].



**Fig. 14.3** In the presence of stressful stimuli, for instance fear, the central amygdala activates the paraventricular nucleus of the hypothalamus (PVN) through either a direct way or through the stria terminalis (BNST). Corticotropin-releasing factor (CRF) is released by axons from the PVN into the master hormonal gland, the pituitary gland (PIT), which in turn releases hormones ACTH into the blood stream joining the adrenal cortex. On the other hand, the adrenal cortex releases cortisol, which travels to many body locations, including the brain. Inspired by the figure in [LEDOUX 2002, p. 223].

is perceived as emotionally charged, the lateral nucleus is activated, whereas when a response is produced, the central nucleus is activated through the basal ganglia with the fundamental contribution of the motor-decisional system [Fig. 14.4]. This also means that the emotionally charged sensory stimulus takes two paths simultaneously, a lower and a higher one, the latter through the cognitive system of the sensory cortex [Fig. 14.5]; this explains the partial independence of emotional outputs relative to cognitive aspects and clarifies the discussion, developed in the previous subsection, about the relation between the emotional and cognitive systems. LeDoux and his coworkers discovered this mechanism by making use of Pavlovian conditional stimuli.

Previously, it was thought that a damaged amygdala implied a failure of perception of fear as such. Antonio Damasio and coworkers<sup>52</sup> showed that it is rather a failure to direct one's gaze on a dangerous target (for instance the eyes of a menacing person), which is crucial for fear perception. This also implies that the amygdala contributes in stimulating and moderating attention (again the decisional-motor system is involved). In other words, it was discovered that the neural circuitry underlying fear perception may not be as well localized as was previously thought.

### 14.2.3 Basic Emotional Behaviors

J. Panksepp<sup>53</sup> has distinguished four basic emotional subsystems that organize specific *behavioral* (well articulated) sequences and reactions [see Fig. 14.6]:

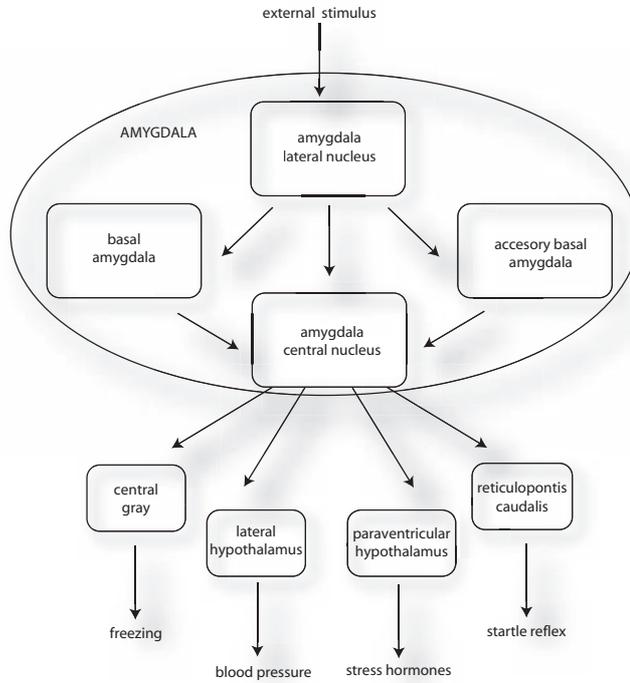
- The seeking system.<sup>54</sup> This is characterized by an autonomous search activity (for feeding sources, predation) and exploration. It involves the lateral hypothalamus. It is an anticipatory behavior.

<sup>52</sup>[ADOLPHS *et al.* 2005].

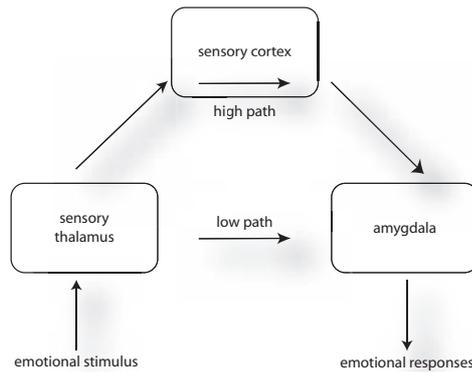
<sup>53</sup>[PANKSEPP 1992] [PANKSEPP 1998, pp. 50–56]. For a different point of view see [ORTONY/TURNER 1990].

<sup>54</sup>[PANKSEPP 1998, pp. 144–63].

**450** Decisional, Emotional, and Cognitive Systems



**Fig. 14.4** Amygdala’s outputs according to LeDoux’s model. The outputs are produced by the central nucleus of the amygdala in which both an emotional input (amygdala lateral nucleus) and a motor input (basal ganglia and accessory basal ganglia) concur. The motor system was already informed by inputs coming from the lateral nucleus of the amygdala.



**Fig. 14.5** The two paths of the emotionally charged sensory stimulus: The emotional response can be very quick, bypassing the sensory-cognitive cortex.

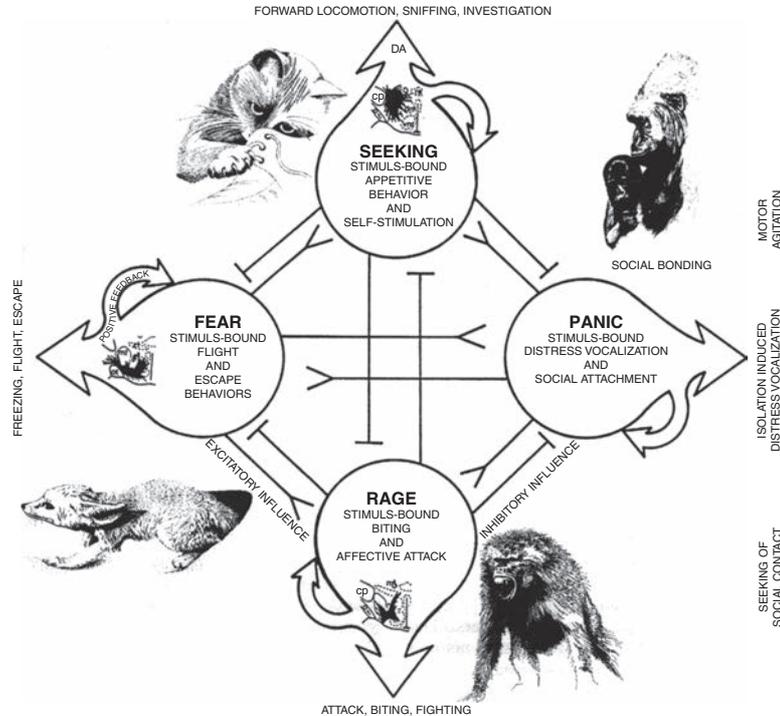


Fig. 14.6 Panksepp's four emotional systems. Adapted from [PANKSEPP 1998, p. 53].

- The fear system,<sup>55</sup> fundamentally an escape behavior.
- The panic system. This represents the breakdown of social relationships.<sup>56</sup> Young mammals (and even birds) in isolation emit distress vocalizations (“crying”). It is localized in the midbrain.
- The rage system.<sup>57</sup> To a certain extent this is the opposite reaction relative to fear: Instead of avoidance, it represents an aggressive approach to the target, and can also be determined by frustration.

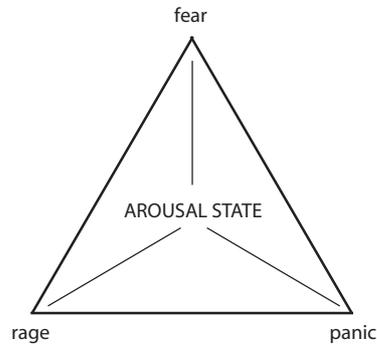
It is likely that the latter three emotional systems arose from the specification of the general arousal state we have previously mentioned, and are strongly dependent on the way in which the nonself has an impact on the self<sup>58</sup> [Fig. 14.7]. This is therefore related to the issue of passive (bottom-up) attention, considered in Subsec. 14.1.1. *Fear* and *rage* are probably founded on two basic states, which correspond to two different cerebral circuits<sup>59</sup>: An *approach system* for positive emotion and a *withdrawal* one for negative ones.<sup>60</sup> These two systems are probably basic to any animal and do not necessarily involve emotions in their fundamental manifestations. In their evolved form in neomammals, they involve a neural machinery located in the prefrontal cortex (PFC), with a subdivision between dorsolateral, ventromedial, and orbitofrontal sectors. The distinction between

<sup>55</sup>[PANKSEPP 1998, pp. 206–222]. <sup>56</sup>[PANKSEPP 1998, pp. 261–76]. <sup>57</sup>[PANKSEPP 1998, pp. 187–203].

<sup>58</sup>[LEVENSON 1992]. <sup>59</sup>[DAVIDSON/IRWIN 1999].

<sup>60</sup>Therefore, this kind of proposal could fit to a certain extent with the analysis developed in [LADYGINA-KOHTS 1935, p. 227] according to which there are three basic emotions in chimpanzees: Anxiety (the first and weaker manifestation of panic), joy (an active behavior like rage, but much more sophisticated), and sadness (a higher psychological manifestation of fear).

## 452 Decisional, Emotional, and Cognitive Systems



**Fig. 14.7** Relation between the arousal state and the fundamental three induced emotional states: Fear, rage, and panic. These three states are induced by the impact of the external environment on the self. On the contrary, the seeking system responds to an autonomous activity of the self.

left and right PFC is also relevant. Several studies show that the left PFC is responsible for the approach part of the emotional system, while the right PFC is responsible for avoidance. *Panic* is, instead, a form of paralysis and even a breakdown of the decisional system deriving from the conflict between opposite tendencies, essentially fear and rage. The true mark of panic is indeed when the organism remains frozen in its previous position and behavior, unable to attack or to flee, or when the reactions are fully uncontrolled and often showing dissonance with the context.

An interesting case of aggression is represented by male–male intraspecific competition. Here, we probably have a mechanism in which both the fear and the rage systems are active, but rather than producing a paralysis (as in panic), they give rise to an increase in arousal instead. In general, the more limited the territories are in number and quality, the stronger the selection for extravagant male characters and sexual size dimorphism [Subsec. 12.5.3 and Sec. 12.6]. The degree of territory heterogeneity (both size and quality) can also affect the strength of sexual selection: The more heterogeneous, the stronger the selection. The reason for this is that female choice is more important if differences exist in their reproductive output.

The *seeking system*, on the contrary, is strongly dependent on the endogenous decision of the self to start a complex series of actions, and it is therefore rather connected with the issue of active (top-down) attention [Subsec. 14.1.1]. Obviously, this seeking behavior is not indifferent to metabolic-homeostatic inputs [Fig. 14.8]. Nevertheless, this should rather be considered in terms of the general influence that the hormonal system exercises on the decisional system. As a matter of fact, predation is a very diffuse behavior that is present in reptiles, amphibians, and fishes, i.e. all animals that do not seem to possess an emotional system. My guess is that this is the source of noninteractive sensorimotor schemata [Subsec. 12.7.1]. It is important to stress here that aggressive behavior can not only be induced by the rage system but also be the consequence of a predation activity, although their modalities are very different: While predation is in general against members of other species, rage has a wider range of possible activation mechanisms, and is usually much stronger when there is a social bond with the animal or person toward which it is directed. Rage and seeking systems can also interact, when for instance a predation act fails, giving rise to frustration.

It is important to stress that the existence of an elementary emotional nuclei does not mean that there is a code of emotions.<sup>61</sup> Actually, emotions are not discrete units that can be combined

<sup>61</sup>[EKMAN *et al.* 1985]

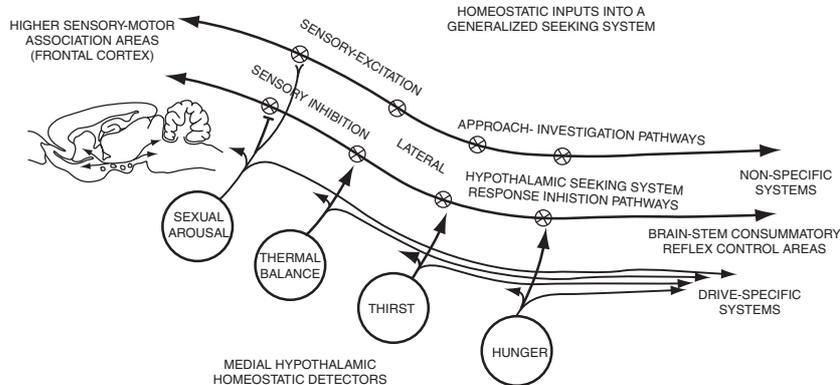


Fig. 14.8 Homeostatic inputs to the seeking system. Adapted from [PANKSEPP 1998, p. 167].

in a linear way [Subsec. 13.1.2], notwithstanding the fact that, at least in humans (but to a certain extent also in other primates), there are seven universal facial expressions: Anger, happiness, disgust, surprise, sadness, contempt, and fear.<sup>62</sup>

#### 14.2.4 Other Emotional Behaviors

According to Panksepp, there are three further (social) emotional behaviors<sup>63</sup>:

- *Interactive sexuality.* Sexual activity is very ancient and starts with eukaryotes [Subsec. 7.5.2]. The novelty with mammals is that it acquires a new social and therefore interactive significance [Subsec. 12.7.2]. It is here that the audiovisual communication reaches a new level of complexity [Subsec. 13.1.4].
- *Care and nursery.* It is interesting to observe that in mammals, the ancient molecules that control reptilian sexuality and egg laying have evolved into the oxytocin and arginine-vasopressin social circuits of the brain. The infant's suckling activates the paraventricular nucleus (PVN) of the hypothalamus, which induces release of oxytocin by the pituitary gland into the circulation. It is similar to the dynamics induced by opiates.
- *Play.* Play and aggression follow completely different rules. Play can have a function during waking similar to REM during sleeping. Actually, all play is a closed-circuits activity, since it is self-satisfactory and does not aim at other purposes apart from playing.

Play involves communication, intention, role-playing, and cooperation. It can be defined as<sup>64</sup> a motor activity performed postnatally that *appears* to be purposeless, in which motor patterns from other contexts may often be used in modified forms and altered temporal sequences. It is a form of stimulus-free, endogenous behavior. Play is difficult to define because it is a relative term. It is also clear that play has separate motivations. However, the heterogeneity is not a specific problem of play (also predation and sexuality are heterogeneous). Three general classes of functional hypotheses about play appear to have current standing in biology: (a) Training (but in this way one cannot explain adult play), (b) developmental rate (but the adaptive significance in

<sup>62</sup>[EKMAN *et al.* 1985] [KANDEL 2006, pp. 385–90] [GAZZANIGA *et al.* 1998, pp. 537–45].

<sup>63</sup>[PANKSEPP 1998, pp. 225–59, 280–97]. See also [ELSTER 1999a, ELSTER 1999b].

<sup>64</sup>[FAGEN 1981] [ALLEN/BEKOFF 1997, pp. 88–92].

## 454 Decisional, Emotional, and Cognitive Systems

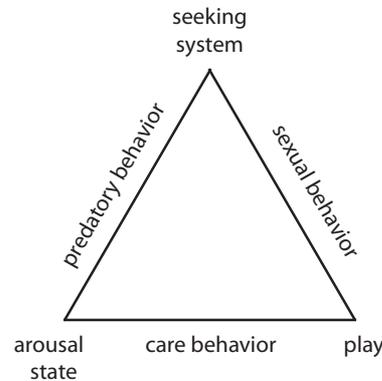


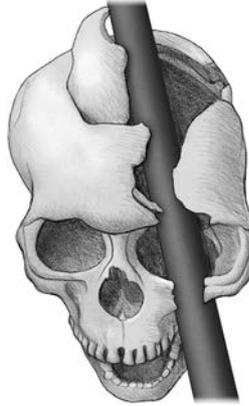
Fig. 14.9 Synoptic summary of emotions.

terms of fitness and evolutionary stability is seldom clear), and (c) cohesion, including short-term recognition (one explains adult and courtship play but fails to address play as such). Play is a behavior that has the function to develop, practice, or maintain physical or cognitive abilities and social relationships, including both tactics and strategies, by varying, repeating, and/or recombining already functional subsequences of behavior outside of their primary context.<sup>65</sup> It is a matter of taste considering some forms of behavior that do not simultaneously satisfy all these criteria as play; Fagen shows that 15 possible nonempty sets (categories) may be built out of the four types of criteria listed here. Bateson<sup>66</sup> stressed that an animal's play-signals serve to communicate contextual information about its subsequent behavior: They help to discriminate between pure mood-signs experiencing emotionally charged states [Subsec. 12.7.2] and simulations of the same.

It is important to remark that play is a costly behavior, whose benefits can be perhaps summarized as follows: It helps to develop and preserve physical ability, including strength, physical skill, endurance, and cognitive skill. It is also useful for learning specific information. Play fighting and real fighting indeed develop in parallel and interactively.<sup>67</sup> An animal at any given level of fighting ability is always potentially capable of improving this ability by playfighting, although the amount of possible refinement decreases with the age. Animals are in general capable of recognizing this duality: The animal's own attitudes and purposes are crucial here. Play has different forms depending on whether it is a play with inanimate objects and nonconspecifics, or with some conspecifics. It is interesting to note that baboon infants play mostly with infants of the highest-ranking female in their mothers' subgroup and juvenile and subadult baboons play mostly with infants whose mothers ranked higher than their own. In other words, choice of playmates may be tuned to potential benefits of support in later life.<sup>68</sup>

Probably, all emotional behaviors can be cast as in Fig. 14.9. I assume that the seeking system is the most primitive form of "emotion" and that it is already present in reptiles. I also assume that predatory behavior is very ancient and surely premammalian, even if in this case it may rather be connected with fixed schemata of reaction and action. In mammals (and in part in birds) a true affective system is constituted, in which two new dimensions are added: The arousal state and play, the former dealing with any external event, especially dangerous situations, the latter

<sup>65</sup>[FAGEN 1981, pp. 33–69].<sup>66</sup>[BATESON 1955].<sup>67</sup>[FAGEN 1981, pp. 248–495].<sup>68</sup>[CHENEY 1978].



**Fig. 14.10** The lesions reported by Phineas Gage after an accident in which an iron bar traversed his skull.

dealing especially with nonaggressive social relations with conspecifics. These two fundamental affective states enrich the sexual behavior as a high social activity and also determine a new form of behavior, namely the care of infants.

#### 14.2.5 The Influence of the Affective System on the Decisional System

Damasio's contribution showed that, in humans, rational, moral, and social decisions also depend on the emotional sphere.<sup>69</sup> In fact, the human patients affected by bilateral damage to prefrontal cortices including the ventromedial sector, do not feel their own situation, and they also fail in realizing a system of values. The most studied subject is the famous Phineas Gage<sup>70</sup> [Fig. 14.10]. In fact, these patients have access to social knowledge and they pass tests very well at a purely cognitive level; but the problem is the decision process itself. In other words, to act properly one needs some emotional engagement, not merely to possess an abstract knowledge about the involved actors. Thus, we may say that without the concurrence of the emotional system, knowledge does not constitute a sufficient motivation to act.

As a matter of fact, cognitive and emotional impulses intersect in the ventromedial prefrontal cortices. The point is that decision is confronted with such a huge amount of data that it would take an inordinately long time. Moreover, attention and working memory have a limited capacity. Now, the *somatic-marker hypothesis* consists in the fact that attention is forced to focus on a possible negative outcome to which a given action may lead and function therefore as an alarm signal. It is a more sophisticated system of marking [Subsec. 8.2.3]. This implies an emotional engagement that leads to the rejection of the negative course of action and consequently other alternatives are chosen. Somatic markers are acquired in social and environmental experiences.

This is confirmed by recent studies.<sup>71</sup> Patients with focal bilateral damage to the ventromedial prefrontal cortex, a brain region necessary for emotion generation, produce abnormal utilitarian behavior patterns that are unable to take into account moral aspects influenced by the emotional involvement of the person.

<sup>69</sup>[DAMASIO 1994].

<sup>70</sup>[DAMASIO *et al.* 1994].

<sup>71</sup>[KOENIGS *et al.* 2007].

### 14.3 The Brain Acquires Information Dynamically

We have extensively dealt with the cognitive system (schemata and representations). Let us now proceed to a more specific examination of the dynamic way in which representations are built by the cognitive system.

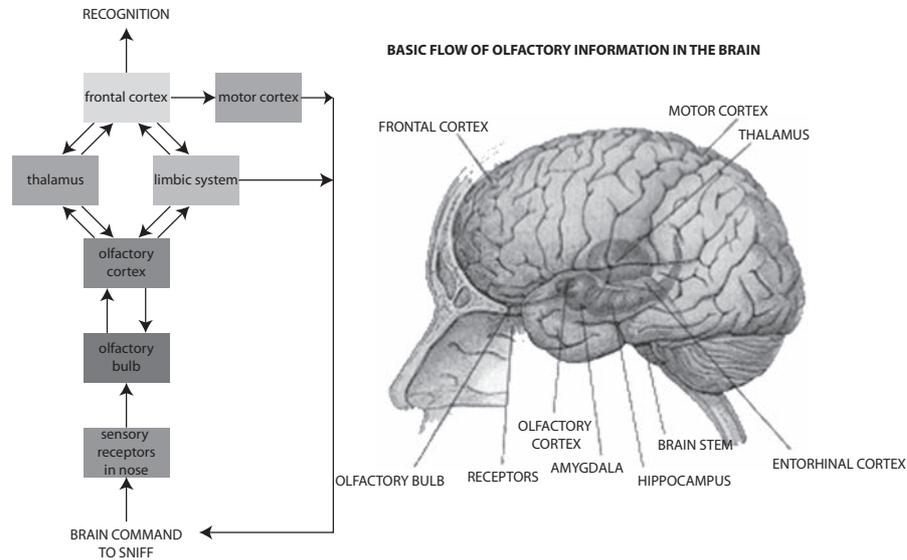
Given that the brain receives its information in a point-like form, how can it integrate this disorder to generate order (structured perception)? According to a passive, computational, bottom-up view of perception, the initial step is the perturbation of receptor neurons that through a biochemical bottom-up cascade causes the release of generator currents. These currents initiate action potentials of first-order neurons in the peripheral nervous system, exciting second-order neurons in the brain stem or spinal cord. They transmit to third-order neurons located in the thalamus, which finally inject the information into the primary sensory cortex. The final step is integration from different areas of the brain (from different perception modalities). This perspective may turn out to be true for nonmammals but cannot provide a complete account of the mammal sensory-cognitive system and must be integrated with an active and top-down view.

According to the *active and top-down view*, at least in mammals, the process begins with the emergence of a *pattern of expectancy* within the brain,<sup>72</sup> even if some form of elementary expectation can be found in every organism [Sec. 12.2]. This is particularly relevant in exploratory or active attentional behavior, but also when a first perception of a new item, requiring specific, top-down attention, has already occurred [Subsecs. 12.5.1, 14.1.1, and 14.2.3]. Indeed, such expectancy finds its roots in the necessity of the organism (especially when quick motor performances are necessary) to give rise to an *anticipatory behavior*; which cannot exclusively rely on feedback mechanisms for error correction.<sup>73</sup> Especially for higher organisms, there is indeed a (dangerous) delay from the moment in which an afferent signal is detected to the moment in which the source of disturbance is individuated. For this reason, the perturbation is also anticipated before the process is initiated in order to perform the correction *before* the perturbation has actually occurred. Such a control mechanism is a feedforward control (antifeedback) consisting in anticipating environmental negative feedback. This shows perfectly how such a dynamic view cannot be supported if we do not consider the motor–decisional–attentional system or the emotional system as being in cooperation with the sensory system, as mentioned before [Fig. 14.11].

The neural and physical basis for expectancy can be described as a set of attractor landscapes: I recall what has been said about equifinality [Subsec. 6.3.3] as well as about the necessity for a new language of physics [Sec. 8.1 and Subsec. 11.2.3]. The limbic-emotional system appears to act on the primary sensory cortices by modulating the attractor landscapes, so as to enhance selected basins of attraction. The selection process is called refference [Subsecs. 5.3.3 and 11.5.3], since the initiation of exploratory behaviors (such as sniffing, orienting, or searching by movement of the head, eyes, and ears) are accompanied by corollary discharges guiding the sensory cortices to those (equivalence) classes of sought stimuli. The attractors are preferred modes of global activity, which can be expressed in spatial patterns of amplitude modulation (AM) of an oscillatory wave form that is shared over the whole field of the primary sensory cortex. AM is a way to transmit information through a wave by varying the amplitude (intensity) of the transmitted signal [Fig. 1.1].

<sup>72</sup>[FREEMAN 2000b]. In other words, mammals' information acquisition is framed from the start in the second-order and third-order representational systems [Subsec. 12.7.2–12.7.3]. Bruner had already pointed this out, at least in the case of humans: Perception is characterized by readiness, that is, the accessibility of categories to the afferent stimulus inputs [BRUNER 1957].

<sup>73</sup>[GRILLNER 1999].



**Fig. 14.11** Flow of olfactory information in the human brain [see also Figs. 13.5–13.6, and 14.5]. Interaction between the olfactory cortex and the bulb as well as feedback from other parts of the brain are essential for the maintenance of control of variability in the olfactory system. Adapted from [FREEMAN 1991].

The spatially distributed interactions of neurons in the cortical populations give rise to a spatially coherent complex or chaotic attractor,<sup>74</sup> explaining the way in which self-organization is the basis of representational-pattern formation [Secs. 6.3 and 12.4]. This means that individual neurons participate in many different assemblies. In general, it is a Hebbian mechanism where some connections result strengthened<sup>75</sup> [Subsec. 3.8.2]. It is perturbation from which order emerges thanks to previous constraints [Subsec. 8.2.7]. The field of noisy chaos is the foundation of perception, because the distributed interactions among cortical neurons that arise in nerve cell assemblies “enslave” the mass of cortical neurons. They lead to (point-like) symmetry breaking of the white noise upon the formation of AM patterns by phase transitions. Sequential phase transitions give rise to a so-called chaotic itinerancy,<sup>76</sup> which expresses preferred or habitual trajectories of behavior but with a slight shift. Recent models<sup>77</sup> show that the classical attractor concept is not needed at all here, and that it is possible to deal with the problem with the so-called stable heteroclinic channel. A stable heteroclinic channel is defined by a sequence of successive metastable (“saddle”) states (a system is said to be in a metastable state when a slight fluctuation can drive it out of equilibrium). Under the proper conditions, all the trajectories in the neighborhood of these saddle points remain in the channel, ensuring robustness and reproducibility in a wide range of control parameters. Such dynamics can be described by the so-called generalized Lotka–Volterra equation [see the Appendix to this chapter], and are very common in complex systems. The mechanism for the readout of cortical AM patterns facilitates the extraction of those patterns as “signals” from background “noise.” Unlike the topographic mapping by which sensory input patterns are injected from receptor arrays into primary sensory cortices, the output of the cortex typically comes from

<sup>74</sup>[LARIMER/STROWBRIDGE 2007]. <sup>75</sup>[CASSENAER/LAURENT 2007].

<sup>76</sup>[TSUDA 1991]. <sup>77</sup>[RABINOVICH *et al.* 2008].

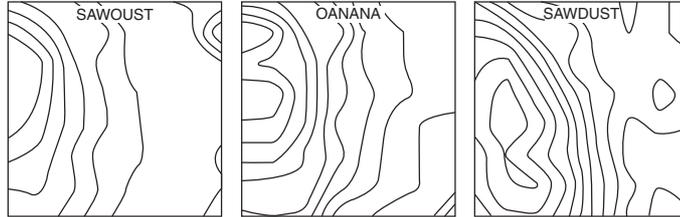
**458** Decisional, Emotional, and Cognitive Systems

divergent/convergent pathways. Each projection neuron diverges its pulses to many target neurons. With convergence, each receiving neuron sums the dendritic currents triggered by neurons that are broadly scattered over the transmitting cortex. This is how brains carry out spatial ensemble averaging. No storage of traces is required here [Sec. 12.4 and Subsec. 12.5.2]. The only activity that survives this spatial integration is that which has the same instantaneous frequency and phase over the spatial extent of the integration. The consequence is that the raw sense data, which can be regarded as a representation of each particular stimulus, is deleted by the brain as noise, and the self-organized AM pattern is (in the framework of the general context to which memory and other factors also contribute) accepted as the signal, that is, the (in terms of survival) significance (affordance) of the equivalence class of the particular stimulus [Subsecs. 8.2.3–8.2.5, Sec. 13.3]. I stress again that this would be impossible (at least at the plastic and high cognitive level of mammals) without contributions from the emotional system. Moreover, it is another instance of supplementary selection [Subsecs. 14.1.1 and 14.2.1], this time due to the central contribution of the sensory system. Summing up, we can say that it is only *supplementary selected stimuli* to possess survival value.

The *basin* of each attractor involved here is determined by the collection of past sensory stimuli of a certain kind, that were transmitted to the cortex during the time when that kind of stimulus was learned under reinforcement. Because the AM pattern is formed after the phase transition is induced by a sensory stimulus of a certain class, the AM pattern, at the time of its discovery, was thought to “represent” the stimulus in the brain (according to the traditional model of pure information-processing). This turned out not to be the case as the AM pattern was found to lack invariance with respect to a constant stimulus<sup>78</sup> [see also Subsec. 12.3.3]. For example, in serial discriminative conditioning, when animals were trained to respond to a succession of stimuli with the same conditioned response, a new instance of a previously learned stimulus resulted in a new AM pattern (it is a token! [Sec. 12.1]) and not in a recurrence of the old one [Sec. 12.2]. Then, the AM pattern (the icon, here) could be considered as the physical basis for the perceptual significance of a stimulus (the referent) to the biological self [Sec. 8.4], in a certain context, when taking into account the history and the action (the whole semiotic and representational mediation) toward which the stimulus attracts the individual having the AM pattern as an expression of the brain state. However, as a wave-like itinerant pattern, it cannot represent the stimulus as such without the contribution of spike-like phenomena and attentional–anticipatory processes to build a perceptual path. Thus, the problem of integration arises, that is, of dynamically modifying the attractors and their basins in order to integrate the new stimulus into the wave-like pattern.

Let us focus on some experiments that Walter Freeman has conducted on rabbits, even if the main conclusions can be correctly extended to higher mammals. Freeman found that the sensory input simultaneously excites a subset of the roughly 100,000 projection neurons, the mitral cells, to which the receptor axons converge. The mitral cells interact by reciprocal excitatory synapses. If the input is accompanied by reinforcement, then according to Hebb’s rule [Subsec. 3.8.2] the synapses between coactive neurons are strengthened.<sup>79</sup> A pair of mitral cells is changed by learning to fire together just as the pair of receptors fired together, but with the difference that, in later sniffing experiences, that pair of mitral cells will again cofire, even if only one of the two receptors is activated. As explained, this represents a generalization over equivalent stimuli. Again, there is no one-to-one relation between an odorant chemical and an AM pattern of neural activity since there are multiple processes that contribute to the pattern formation, including the entire past history of the individual and the state of arousal, as well as the process of refference that brings attention to focus [Fig. 14.12].

<sup>78</sup>[SKARDA/FREEMAN 1987].<sup>79</sup>[FREEMAN 1995].



**Fig. 14.12** Contour plot at the left emerged consistently from bulbar EEG's of a rabbit that had been conditioned to associate the scent of sawdust with a particular reinforcement. After the animal learned to recognize the smell of banana (middle), however, reexposure to sawdust led to the emergence of a new sawdust plot (right). Adapted from [FREEMAN 1991].

The fact that familiar sniffs in rabbits self-organize not just as a responses to the specific odor but in a complex context,<sup>80</sup> is only explainable if there is no fixed memorized schema but a dynamic assembly that is always a function in global activity [Subsecs. 12.4.3 and 12.5.2]. Nerve cell assembly that is repository of past associations participates in the global excitation. As I have said, it is not the shape of the carrier wave (the excitation wave conveying the information) that reveals the identity of an odor. Indeed, the wave changes every time an animal inhales (this is a token), even when the same odorant is repeatedly sniffed. The identity of an odorant is reliably discernible only as a trajectory in the bulbwide spatial pattern of the carrier-wave amplitude. Whenever an odorant becomes meaningful in some way, another attractor can be added and all the others undergo slight modification<sup>81</sup>: In short, the brain does not copy a stimulus but perception is a step in a dynamic trajectory preserving the memory of several past experiences (organization and reorganization of the brain).

#### 14.4 Concluding Remarks

In this chapter, we examined the three main functional systems of the brain and found:

- The decisional-motor system gives rise to a generic alert state and is organized into a lower level for execution, a middle level for programming, and a higher level for planning. To this higher level there are contributions coming from the cognitive and emotional systems. Planning formulates prior intentions embedding specific actions that are purposefully programmed and executed.
- Emotional stimuli can take both a lower path directly from the sensory thalamus to the amygdala to produce quick emotional responses or a high path passing through the sensory cortex. The emotional system taken separately can only give rise to a general arousal state. In order to produce specific emotional responses either the sensory-cognitive system or the decisional-motor system must take part. In this case, three basic emotional reactions can be produced: fear, rage, or panic. When play and the seeking system are added, we also have predatory behavior, care behavior, and sexual behavior.
- The sensory system as such can only give rise to information-processing. When on the contrary the decisional or emotional systems are involved, we have an anticipatory behavior that is loaded with expectations.

<sup>80</sup>[THELEN/SMITH 1994, pp. 215–45].

<sup>81</sup>[FREEMAN 1991, pp. 85–6].

**460** Decisional, Emotional, and Cognitive Systems

- Only *supplementary* selected stimuli are meaningful for survival. They are (a) stimuli whose content is particularly stressed (when the attentional alert states come into play), (b) stimuli that are emotionally appraised and charged (when the arousal state contributes), or even (c) stimuli that are accepted and acquired on a background noise (when the sensory information-processing activity is involved).

With this and the previous two chapters we have established the general framework in which the main mammal brain *functions* and activities are displayed: Behavior, learning, and memory. This is the subject of the next three chapters.

**Appendix: Generalized Lotka–Volterra Equation**

We have introduced a very basic equation (10.6) for describing the relation of a species with its ecological network. Another, more specific way to consider the problem is represented by the so-called Lotka–Volterra equations<sup>82</sup>:

$$\frac{dN}{dt} = N(a - bP), \quad \frac{dP}{dt} = P(cN - d), \quad (14.1)$$

which relates the prey population described by  $N(t)$  and that of predators described by  $P(t)$ . What these equations tell us is:

- In the absence of any predation, the prey population grows unboundedly due to the presence of the  $aN$  term in the first equation;
- The effect of predation is to reduce the prey's per capita growth rate by a term proportional to both the prey and predator population, that is the  $-bNP$  term in the first equation;
- In the absence of any prey, the predator population results in exponential decay due to the presence of the  $-dP$  term in the second equation;
- The prey's contribution to the predators' growth rate is given by the  $cNP$  term in the second equation.

A generalization of the previous equations is represented by<sup>83</sup>:

$$\frac{dA_j(t)}{dt} = A_j(t)F \left( \sigma_j(I_k) - \sum_{i=1}^N \rho_{ji}A_i(t) + \eta_j(t) \right), \quad (14.2)$$

where  $j = 1, \dots, N$  and  $F(0) = 0$ . The variables  $A_j(t) \geq 0$  can represent the amount of biomass in ecological networks, the probability to use a particular strategy in game theory, but also the firing rate of excitatory neurons. The variable  $\eta_j(t)$  represents the perturbation of the system,  $I_k$  the environmental stimulus and  $\sigma_j(I_k)$  the gain function that controls its impact,  $\rho_{ji}$  determines the interaction between the elements (for  $\rho_{ji}$  all agents compete with each other). Without noise the generalized Lotka–Volterra equation has many fixed points; whether these are saddles or nodes depends on the values of  $\rho_{ji}$ .

<sup>82</sup>[VOLTERRA 1926a, VOLTERRA 1926b] [LOTKA 1925] [MURRAY 1989, pp. 63–8].

<sup>83</sup>[RABINOVICH *et al.* 2008]. See also [MURRAY 1989, pp. 161–6].

# 15

## Behavior

---

In this chapter we shall deal with the first fundamental function connecting the three main subsystems of the brain (sensory–cognitive, emotional, motor–decisional): Behavior. After having studied the classical behaviorist approach to the problem and considered some of its intrinsic limitations, we shall deal with the school of ethology, which has preferred observational studies on the field to laboratory experiments. Finally, I shall present the true revolutionary discovery of mirror neurons that allows us to treat the high forms of behavior as determined by the third-order representational system in a very precise and rich form.

### 15.1 Behaviorism

*Behavior* can be preliminarily defined as a coherently organized sequence of motion executions or postures that might be finalized to a task and determined by some emotional component. For this reason, behavior cannot be identified with movement [Ch. 5]. It can also consist in the absence of external action since the suppression of movement can still be a behavior<sup>1</sup> (for instance when a predator does not want to be apperceived by the prey), but it necessarily consists in a structured cluster of single motion acts or postures. This is the reason why we cannot completely remove representational aspects when we deal with the external world<sup>2</sup> [Subsec. 5.3.4]. On the other hand, movement may be reflexive [Subsecs. 5.3.1 and 8.3.1], while, though behavior can have a reflex component, it is never purely reflexive.

Let us first consider the theory that is centered on behavior and that brought its relevance to the attention of the scientific community. *Behaviorism* is a form of experimental associationism [Secs. 3.8, 12.2, 12.4]. Although acknowledging from the start the possibility that many human acts are uncertain,<sup>3</sup> it proposed a science of behavior understood as the organism's *external* and *controllable* manifestation. Behaviorism searched for the causes (functional dependencies) of behavior by discarding inner causes (also neural ones) because, at that time (until the 1970s), the knowledge of related neural mechanisms was still insufficient. The important point for behaviorists was to evaluate the probability (the frequency) of a certain behavior given certain variables (context or initial conditions), and to ascertain the relevant variables (the so-called control parameters [Subsec. 6.5.4]) able to change the state of an organism from without. However, most behaviorists also (wrongly) assumed that it would never be possible to achieve a knowledge that is able to predict and control behavior on the basis of the knowledge of *internal* processes.

<sup>1</sup>[ALLEN/BEKOFF 1997, pp. 41–8].

<sup>2</sup>[REES/FRITH 2001].

<sup>3</sup>[SKINNER 1953].

### 15.1.1 A Historical Background

Historically, behaviorism began with the work of Ivan P. Pavlov, a Russian scholar, as well as with that of John B. Watson and Edward L. Thorndike, two American psychologists.<sup>4</sup> Pavlov studied animals' (actually, mammals') responses to conditioning: In particular, dogs had been conditioned in Pavlov's laboratory to salivate at the sound of a bell (the triggering stimulus) previously associated with presentation of food. Watson's work was based on the experiments of Ivan Pavlov. Both Watson and Thorndike claimed that psychology was not concerned with the mind or with human consciousness (subjects that are irrelevant for this study<sup>5</sup>) but only with behavior: This approach could provide scientific grounds to psychology.<sup>6</sup>

Like Pavlov, Thorndike also conducted several experiments on animals (again mammals) and humans. I recall here the paradigmatic experiment where he put a cat in a cage with a latch on the door and a piece of salmon outside of the cage. After first trying to reach through the cage and then scratching at the bars of the cage, the cat finally hit the latch on the door and opened the door. With the repetition of this experiment, the amount of time and effort spent by the cat on the futile activities of reaching and scratching decreased and the releasing of the latch occurred sooner. Thorndike's analysis of these findings was that the behavior producing the desired effect became dominant and therefore occurred faster in the subsequent experiments.<sup>7</sup> He then argued that more complicated forms of behavior were influenced by *anticipated* results, not by a triggering stimulus as Pavlov had still supposed [Sec. 14.3]. This idea became known as the *law of effect*, and it also provided the basis for Skinner's operant conditioning. The law of effect can be formulated as follows<sup>8</sup>: Of several responses to the same situation, those which are accompanied or closely followed by the satisfaction of the animal's purposes, all things being equal, are more firmly connected with the situation, so that when the latter recurs, their recurrence will be more likely. On the contrary, the responses which are accompanied or closely followed by discomfort to the animal's expectation and purposes, all things being equal, have their connections with that situation weakened, so that recurrence will be less likely. The greater the satisfaction or the discomfort, the greater the strengthening or the weakening of the bond.

Behaviorism was later associated with the name of Burrhus F. Skinner, who made his reputation by experimentally testing Watson's theories and realizing some of the experiments he proposed. Skinner's studies led him, like Thorndike, to reject Watson's (and Pavlov's) almost exclusive emphasis on reflexes. He argued that people and animals respond to their environment, but they also actively operate on the environment in order to produce certain consequences [Sec. 10.3 and Subsec. 12.3.1]. Skinner developed the theory of *operant conditioning*, according to which we behave the way we do because this kind of behavior has had certain consequences in the past. Operant conditioning must evolve together with a susceptibility to reinforcement by certain kinds of consequences and a supply of behavior less specifically committed to eliciting or releasing stimuli. According to Skinner, this *selection by consequences* made its appearance with the first cell that was able to reproduce itself, because in this way it could become a causal mode by determining and selecting *new* behavior.<sup>9</sup> Moreover, when sexual contact and food-gathering become reinforced, new

<sup>4</sup>[THORNDIKE 1898, THORNDIKE 1931] [WATSON 1925]. To a certain extent C. Lloyd Morgan could be considered a precursor [LLOYD MORGAN 1896, LLOYD MORGAN 1900]. For historical reconstruction see [BOAKES 1984] [REZNIKOVA 2007, pp. 11–19]. *Be* Thorndike see also [CORDESCHI 2002, Sec. 2.4].

<sup>5</sup>[WATSON 1914, pp. 4–7].

<sup>6</sup>Pavlov considered it necessary for psychology to go in this direction in order to be considered a science at all [PAVLOV 1927, pp. 3–4].

<sup>7</sup>He relied here on the theoretical grounds established by Lloyd Morgan [LLOYD MORGAN 1896, pp. 149–50].

<sup>8</sup>[THORNDIKE 1898, p. 244]. See also [THORNDIKE 1931, pp. 30–63]. <sup>9</sup>[SKINNER 1984].

forms of behavior are also reinforced and maintained, which are no longer necessarily adaptive.<sup>10</sup> In this way, operant conditioning is not so far away from the Baldwin effect [Sec. 9.10]. It is very important to realize that the neo-Darwinian synthesis did not make an adequate treatment of this problem possible, especially with regard to the connection and reinforcement mechanisms between plastic-interactive or intelligent behavior and natural selection. G. Romanes<sup>11</sup> understood quite well that plastic or even intelligent adjustments help the results of natural selection to become really adaptive. Notwithstanding his criticism, according to Skinner Pavlovian respondent conditioning continues to work together with operant conditioning.

Behaviorism substituted the old procedure of observation and collection of anecdotal facts about the behavior of animals and humans in a true-life context with a verifiable experimental procedure.<sup>12</sup> It originated in the field of psychology but it has had a much wider influence. Its concepts and methods were also used in education and sociology, in the form of sociobiology.

### 15.1.2 Two Schools

Behaviorism, in both Pavlov's and Skinner's views, was centered on serial learning (subjects are asked to remember both items and their serial order). The stimulus–reaction (S–R) theory consists in the idea that reinforced repetitions of an S–R pair gradually build up their associative strength toward a maximum<sup>13</sup> [Subsec. 5.1.1]. As we shall consider in the following, Skinner's approach is rather centered on a R–S sequence since it focuses on the reinforcement of an operant response; however, it is still serial. Experiments have shown that recognition tests are easier to pass than recall tests (the recall threshold is higher than the recognition one). Three consequences emerge: (1) Any S–R (or R–S) pair that is recalled can also be recognized. (2) There is an accumulation in strength at a “subthreshold” level before the association is recalled. (3) Repetitions beyond the point of recall (overlearning) continue to strengthen the habit.

Notwithstanding this serial explanation, I stress that no form of conditioning deals with pure reactive behavior [Subsec. 8.3.1]. Reactive behavior is a unimodal response to a very specific stimulus, and cannot show the minimal plasticity and variety of possible reactions that are necessary for conditioning. A very important result of conditioning (in any form) is indeed the following: Its inhibition of instincts and previous associations. An animal who has become habituated to get out of a box by pulling a loop and opening the door will do so even though the hole at the top of the box is uncovered, whereas, if in early trials a hole is left uncovered, it would take the instinctive way and crawl through it.<sup>14</sup>

Let us now consider the main differences between conditioned (Pavlovian) reflex and operant (Skinnerian) conditioning:

- Pavlov's respondent conditioning<sup>15</sup> consists in transferring (a) involuntary (reflex-like) responses to an unconditional stimulus to (b) conditioned responses to an associated conditional stimulus.<sup>16</sup> In Pavlov's famous conditioning experiments with dogs, the presentation of food was the unconditional stimulus because it evoked the natural (unconditional) response of salivation in preparation for eating.<sup>17</sup> Thus, spontaneously salivating to food, a dog has been conditioned to salivate when it hears a bell (the conditional stimulus), even in the absence of food [Fig. 15.1]. An unconditional stimulus is a stimulus that is directly (and naturally) related to a response

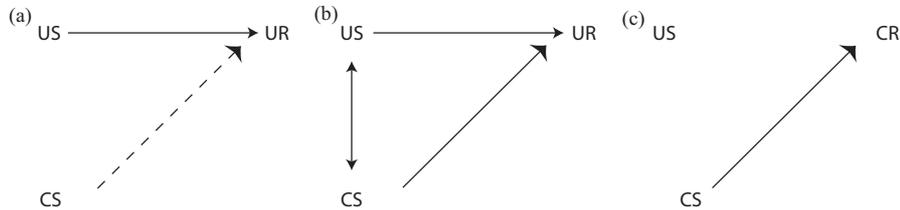
<sup>10</sup>[SKINNER 1981].      <sup>11</sup>[ROMANES 1884, pp. 177–8, 200–3, and 219].      <sup>12</sup>[PREMACK 1988].

<sup>13</sup>[BOWER 2000].      <sup>14</sup>[THORNDIKE 1898, pp. 142–3].

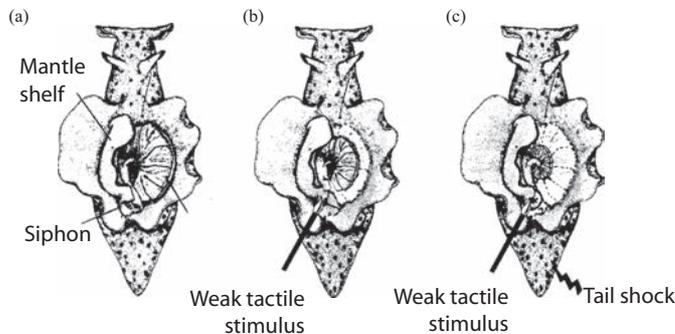
<sup>15</sup>[PAVLOV 1927, pp. 33–47].      <sup>16</sup>[REZNIKOVA 2007, pp. 43–5].

<sup>17</sup>Respondent conditioning could also occur spontaneously in nature [FROLOV 1937, p. 38].

## 464 Behavior



**Fig. 15.1** The three steps of Pavlovian conditioning. (a) An unconditional stimulus (US) provokes an unconditional response (UR). (b) A connection between an unconditional and a conditional stimulus (CS) is established. (c) Now the conditional stimulus alone is able to elicit the response that in this way becomes a conditional or conditioned response (CR). The critical issue here is: Are UR and CR equal?

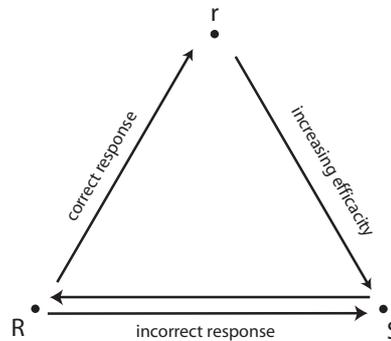


**Fig. 15.2** A primitive and still very much associative form of learning. (a) The gill, through which *Aplysia* breathes, is normally relaxed. (b) The gill withdraws into the mantle cavity for protection when the snail is touched on its siphon. Even this simple response can be modified through habituation, sensitization, and conditioning. (c) After repetition of a weak touch on the siphon, habituation is induced. However, when this weak stimulus is paired with a noxious one, a strong gill-withdrawal reflex is produced as a response even to a weak touch alone. However, the two signals must occur in a precise sequence: When the siphon is touched just before the tail is, the sensory neurons will fire action potentials just before they receive signals from the tail. Adapted from [KANDEL 2006, p. 191].

(for instance, food is related to salivation) and is therefore an affordance [Subsec. 12.5.3], while the conditional stimulus (for instance, the sound of a bell) is originally unrelated with the pair unconditional stimulus–response and is set up in connections between sensory and motor areas in the cortex. Another and more elementary example of respondent conditioning is shown in Fig. 15.2.<sup>18</sup> In my opinion, Pavlovian behaviorism could be summarized as *old responses to new stimuli* (unconditional responses to conditional stimuli). This obviously raises the question whether or not the response is still unconditional. The biological basis of this behavior is to be found in the teleonomic processes of adaptation<sup>19</sup> [Subsec. 8.2.1].

<sup>18</sup>See also [CASTELLUCCI *et al.* 1970] [KUPFERMANN *et al.* 1970] [PINKSER *et al.* 1970] [?].

<sup>19</sup>[GERHART/KIRSCHNER 1997, pp. 201–9].



**Fig. 15.3** The mechanism of operant conditioning: A stimulus S determines a response R. If the response is incorrect, we start the process again with a new equivalent stimulus. If it is correct, it will produce a reinforcement r that strengthens the efficacy of the stimulus. It is a typical negative–positive feedback circle.

- In Skinner’s operant conditioning, which as mentioned represents a further development of Thorndike’s work (but also with some important differences, as we shall see), a reinforcement (positive or negative) will occur only *upon* an *appropriate* response of the organism (and not simply to accompany another stimulus in an associative way), and in this sense it represents a feedback [Fig. 15.3]: The consequences of an action determine the likelihood that this action will be performed in the future. Successive approximations in shaping a behavior will eventually produce the desired behavior.<sup>20</sup> Skinnerian behaviorism could be summarized as *new responses to old (already presented) stimuli*. Obviously, this raises the question whether or not the reinforced stimulus is the original one. The biological roots of this behavior can again be found in the teleonomic emergence of new functionalities [Subsecs. 8.2.2 and 9.5.3], but its immediate basis is much more in the teleologic ability of the individual organism to control environmental information and thus deploy new functionalities [Subsec. 8.2.7].

In *respondent* conditioning, the conditional stimulus is a sign that the unconditional stimulus is imminent, it is an inductive-like procedure: A previous class of stimuli (the unconditional or natural ones) is enlarged so as to include other stimuli (the conditional or unnatural ones) that may actually have nothing in common with the previous items, apart from the fact that, *having been associated* in the past with the natural stimuli (i.e. having occurred in spatial and temporal *contiguity*, which is a pure indexical relation), they *provoke* certain reactions. This is strictly associated with the first-order representational system [Subsec. 12.7.1], in particular with perceptual schemata, as far as a past stimulus may provoke new forms of reaction when associated with a new stimulus. A crucial issue here is whether reptiles, which lack an emotional system, are able to show such a behavior as well or whether Pavlovian conditioning is limited to the first type of representation in mammals and birds. Experiments like those shown in Fig. 15.2 seem to lead to the conclusion that such a behavior is not confined to mammals or birds. Accurate tests on this point would be crucial for our understanding of these mechanisms.

<sup>20</sup>[GAZZANIGA/ HEATHERTON 2003, pp. 174–97] [REZNIKOVA 2007, pp. 47–54].

## 466 Behavior

*Operant* conditioning is an abductive-like establishment of a connection. While through respondent (Pavlovian) conditioning, unconditioned responses already available as results of natural selection could come under the control of new stimuli, through operant conditioning, *new responses* could be strengthened (reinforced) by events which immediately follow them. According to Skinner, operant conditioning in this way could ontogenetically replace the phylogenetic work of natural selection and contribute to evolution. As a matter of fact, it is strictly connected with ontogenetic aspects like niche construction [Sec. 10.3]. There are in fact several adaptive advantages: Its consequences are almost immediate and it can give rise to new adaptive solutions. For instance, eating has produced different ways of gathering and cultivating food through reinforcing. This form of behavior and conditioning is strictly connected with the second-order representational system [Subsec. 12.7.2] as far as it amounts to the evaluation and expectation of the consequences of one's own action on both social and natural environment, and therefore, for being fully displayed, requires interactive sensorimotor schemata.

An important consequence of this examination is the following: With operant conditioning we already have a form of psychology. Animals showing these types of behavior have a personality, a character, a specific level of intelligence, and refer socially to other beings in forms that are, at least at a rudimentary level, individual. This obviously does not imply that dogs or primates have a mind. However, the presence of the affective system and the consequent social behavior makes mammals *psychological animals*.

Unfortunately, behaviorism has not considered a third form of behavior associated with the third-order representational system [Subsec. 12.7.3]: When an agent is able to evaluate the consequences of the behavior of *another agent*. This is typical of primates and their communication system: Indeed, it is impossible to communicate information to a partner about something without understanding how the partner (in general, a conspecific) will behave in certain circumstances [Subsec. 12.7.3]. Behaviorists probably did not consider this form of behavior, since here conditioning does not seem possible. This explains in part the failures of the behaviorist method when dealing with human communication, as we shall see later on.

We can generalize Pavlov–Skinner's model of behavior in this form: The behavior function is characterized by reaction or action (in the wide sense of the word, which does not necessarily imply external manifestation), correction, and reinforcement [Fig. 15.4], where nothing is said here about the mechanism of correction and reinforcement, which, as already explained, can be different.

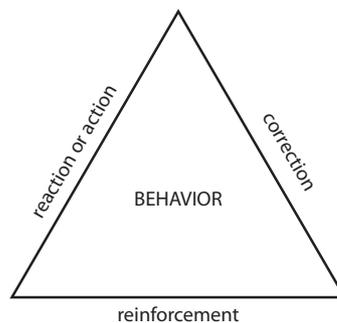


Fig. 15.4 Behavior as an integration process of the components: Action, error correction, and reinforcement.

### 15.1.3 Some Criticisms

An astonishing fact about behaviorism is that for the most part it does not actually deal with behavior, but only with the *results* of individual acts.<sup>21</sup> Furthermore, it is often more interested in learning than in behavior as such, probably because it assumed from the start that learning is a form of behavior; but I will show that this is not the case. Moreover, even if animal and human behavior often consists somehow in a structured and complex series of actions, the behaviorist school only focused on single responses or on basic sequences of single stimuli (this was one of the points of Chomsky's criticism of behaviorism, as we shall see).

A criticism coming from the cognitive school is the following: Block and Fodor<sup>22</sup> stressed that what an organism does or is disposed to do at a given time is a very complicated function of its beliefs and desires together with its current sensory inputs and memories. It is thus very unlikely that it is possible to link together a behavioral predicate with a psychological one in the way that behaviorism requires since, according to behaviorism, an organism might be considered in a certain psychological state if and only if a specified behavioral predicate is true for it at that moment. This suggested to the authors that behaviorism is likely to be wrong simply because of its empirical consequences and independent of its plausibility.

What should be relevant for behavior is the environment as it is interpreted by the agent (rather than as it is described by physics) and actions performed with certain purposes.<sup>23</sup> Many physically different successions of events can be interpreted as instantiating the same behavior (they represent an equivalence class of that behavior [Subsecs. 5.3.2–5.3.3]) whereas identically performed physical actions can instantiate different behaviors, so that certain physical or biological properties are neither sufficient nor necessary for the occurrence of a specific behavior. An organism can respond selectively to properties of the environment (for example beautiful or fearful aspects) that are not physically specifiable. We come back here to the problem of affordances [Subsec. 12.5.3]. As a matter of fact, the behaviorist vocabulary is full of cognitive terms such as stimuli, responses, reinforcers, and so on.

Apart from the general criticism of behaviorism, there are also specific remarks. Pavlovian respondent conditioning was soon criticized. In particular, Thorndike pointed out<sup>24</sup> that *temporal contiguity* cannot lead to learning, because the acquired reflex declined with habituation, contrary to what happens for true learning. For this reason, he considered Pavlov's theory as a theory of excitation rather than of learning. Consequently, according to Thorndike, Pavlov's conditioning theory could not represent an adequate basis for explaining behavior or learning, but was rather concerned with the very first steps of memory consolidation.

Concerning operant conditioning, Hebb pointed out that, in situations where the correct choice is rewarded, mistakes occur frequently, and at the beginning the same mistake is repeated more frequently than the correct behavior up to a point where that mistake vanishes and the correct reaction is produced.<sup>25</sup> Moreover, it was found that punishing the right response by electric shock may help the subject to learn as much as punishing wrong ones. Tolman<sup>26</sup> showed that learning may take place without reinforcement at all and that rats have cognitive maps of their maze (latent learning), a fact that cannot be accounted for in terms of reinforcement but requires endogenously generated goal-directed behavior.

According to Hebb, a theory of associations between autonomous central processes should be substituted for a theory of the association between afferent and efferent processes [Subsec. 14.1.3].

<sup>21</sup>[KELSO 1995, p. 30].      <sup>22</sup>[BLOCK/FODOR 1972].

<sup>23</sup>[PYLYSHYN 1984, pp. 1–48].      <sup>24</sup>[THORNDIKE 1931, pp. 19–25, 101–18].

<sup>25</sup>[HEBB 1949, pp. 174–80]. See also [BRELAND/BRELAND 1961].      <sup>26</sup>[TOLMAN 1932].

## 468 Behavior

Something similar was also stressed by the Gardners<sup>27</sup> on the basis of a yoked control experiment: Here, two boxes are used, one where incentives are delivered in a way that is relative to some specific criterion response, and the other where the same quantity of incentives is delivered, but randomly and independently of the criterion response. The Gardners' point was that one always fails to distinguish between pure contiguity and contingency, so that the principle of response-contingent reward and punishment is without any scientific justification. This is a failure in principle. We have seen the same confusion between contiguity and contingency in the Pavlovian paradigm. What generally happens is that a stimulus serves to evoke obligatory or at least biologically relevant responses rather than to reinforce arbitrary responses, so that conditioning can only work if one aims to reinforce a response which is similar to, or compatible with, the natural or compulsory response evoked by the stimulus. However, if there is some incompatibility instead, the response-contingent stimulus *depresses* the performance.

I have already mentioned that not all forms of behaviorism fit very well with the primates' communication system. The result with the chimpanzee Washoe, in the framework of Gardners' research program, was that the hungrier the chimpanzee and the more attractive the food was, the more disastrous the teaching session was (the reward is here a distractor). Moreover, very few signs of American Sign Language (ASL) could be grasped by pure trial and error.<sup>28</sup> Actually, it seems more parsimonious to say that the trainers' communicative, social responses to Washoe evoked communicative, social responses from her. Terrace's Nim,<sup>29</sup> instead, showed a grabbing behavior because his trainers provided conditions that evoked grabbing rather than communication. In fact, without considering here the issue of whether chimpanzees are *able to learn* human language, we have a failure in those cases in the methodology of *teaching* language to a chimpanzee.

An important point is represented by the experimental procedure used by behaviorists. I have mentioned that it was an important step for controlling some forms of behavior. However, both in the case of animals and in that of humans,<sup>30</sup> it has become increasingly clear that the behavior in a laboratory is mostly very different from the counterpart behavior in a natural context.<sup>31</sup> In other words, behaviorism underestimated the role played by the environmental context (I have, instead, stressed in this book the relevance of environment [Sec. 4.1, Subsecs. 8.2.1, 10.3.2, and 11.2.4]). In a laboratory, animals lack motivation to perform what is requested.<sup>32</sup> Field experiments have two further advantages with respect to the laboratory: 1. Problems of human training are circumvented. 2. Free-ranging mammals daily encounter similar social and nonsocial problems. The trouble with the behaviorist canons of scientific evidence is that they virtually ruled out the description of anything but what is often repeated, and this is precisely the sort of behavior which reveals no particular plasticity and intelligence at all.<sup>33</sup>

#### 15.1.4 Input–Output?

The above examination casts general doubts over the whole of stimulus–response theory or, in more general terms, the input–output theory [Subsec. 12.4.1]. It is very difficult to deal with these problems, since

- Some of the criticisms of behaviorism (especially those of the cognitive school) have been raised by considering behaviors that are much more complex than those explicitly taken into account by

<sup>27</sup>[GARDNER/GARDNER 1988]. <sup>28</sup>[FOOTS/MILLS 1997].

<sup>29</sup>[TERRACE *et al.* 1979]. <sup>30</sup>[KAGAN 2002, p. 35].

<sup>31</sup>[POLLEY *et al.* 2004]. <sup>32</sup>[CHENEY/SEYFARTH 1985]. <sup>33</sup>[DENNETT 1988].

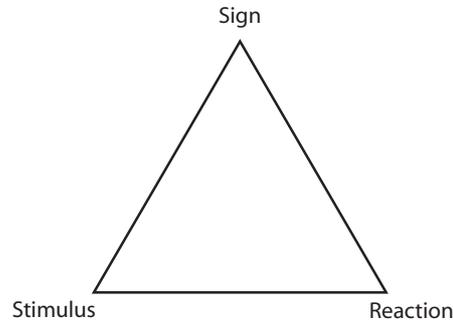


Fig. 15.5 Vygotsky's model of behavior [see also Fig. 8.1].

behaviorists; in particular, the authors of those criticisms have considered behaviors connected with the third-order representational system or characterizing humans.

- There is a problem at a lower level: Information-processing, not to mention information-acquiring, is not at all considered by behaviorism (given the refusal of dealing with internal brain processes). This means that even simple representational tools like spatial maps (as we have seen before for rats) become incomprehensible from this perspective.
- Finally, most critics have understood behaviorism as a theory of reflexes, which is not wholly true.

Consequently, the question arises as to whether or not it is possible to reformulate some aspects of this theory in more general terms in order to take into account some correct criticisms without missing the basic tenets of behaviorism.

Concerning the conditioned reflex, Hebb<sup>34</sup> pointed out that the conditional stimulus does not evoke the unconditioned response, as Pavlov implied. The conditioned response *is not a duplicate* of the unconditioned, but something different [Fig. 15.1]. Attention is needed to fill the gap between the unconditioned and the conditioned stimuli [Subsec. 14.1.1].<sup>35</sup> Furthermore, both neuronal and behavioral learning predominantly occur when there is an error in the prediction about the reward. In fact, learning is not determined by the pure connection stimulus–reward,<sup>36</sup> but needs a *discrepancy* (negative feedback) between the actual and the predicted reward<sup>37</sup> [Sec. 12.2]. For this reason, behavior (as well as learning) is intrinsically *semiotic* [Secs. 8.1 and 12.1; Subsec. 12.5.1]. This semiotic nature of behavior and learning was stressed by Vygotsky.<sup>38</sup> He assumed that, while some very elementary functions are ruled by the S–R (or even R–S) mechanism, for behavioral processes in which some representation is involved there is a triadic structure where the relation between stimulus and reaction is mediated by signs [Fig. 15.5]. In other words, previously remarked difficulties stem from the fact that behaviorism is apparently based on (or at least consistent with) a classical information theory of input–output (S–R). Instead, learning, even when associative and conditioned, always shows a semiotic (under this respect, nonmechanical) nature, and the same is true for behavior [Sec. 12.4].

We could therefore try to reformulate behaviorism in semiotic terms or to take signs more seriously into account when interpreting behavior. However, if we take into account a fundamental criticism coming from Thorndike, the consequence seems a simple deconstruction of behaviorism.

<sup>34</sup>[HEBB 1949, p. 175].

<sup>35</sup>[CLARK/SQUIRE 1998] and [KOCH 2004, p. 191].

<sup>36</sup>[PIAGET 1967, pp. 25–7].

<sup>37</sup>[WAEELTI *et al.* 2001].

<sup>38</sup>[VYGOTSKY 1978].

**470** Behavior

Though a behaviorist himself (at least at the onset of his research), Thorndike was the first psychologist to assume that S–R or input–output direct association was not correct. First, he noted that no animal can form an association leading to an act unless an *autonomous* impulse leading to the act is *already included* in the association.<sup>39</sup> In other words, without a specific *predisposition* of the animal to act in a certain way, no conditioning would be possible. We have already considered this aspect in the previous subsection. However, the consequences drawn by Thorndike were more general than those previously taken into account. In his famous experiment with cats [Subsec. 15.1.1], he considered the relation between the following actions:

- (a) The sense impression of his movements and voice when giving the good signal to the cat.
- (b) The sense impression of his movements in taking a fish, rising, walking to box, etc.
- (c) The cat's act of climbing up, with the impulse leading thereto.

The question was whether, after a while, *a* would remind the cat of *b* and cause him to do *c before* he got the actual sense impression of *b*. If *a* leads to *c* through a memory of *b*, the path from *a* to *c* goes through an association of *representations*<sup>40</sup> and not of mere reflexes. A reason for assuming that animals do indeed have representations and images was found by Thorndike in the longer time taken to form the association between the act of licking or scratching and the consequent escape.

Obviously, the issue here is whether Thorndike's criticism focuses on behavior or learning, provided that these are, to my modest opinion, two different functions. My guess, as I shall explain in the following, is that expectancy is concerned with *learning* and not directly with behavior. However, since learning often constitutes the *antecedent* of behavior and in many cases contributes to it, a highly developed behavior cannot be fully detached from a high-level form of learning. In this sense, Thorndike's criticism is fully correct and points out the fact that mammals have a form of interactive behavior precisely because they show the ability to anticipate certain events.

Therefore, when speaking of representations, behavior is understood from the start as a relation between a *certain expectation* based on past experience (the iconic aspect, here) and a *certain event* to which this expectation is referred<sup>41</sup> [Sec. 8.1 and Subsec. 12.5.1]. In the first-order representational system this is still very rudimentary [Subsec. 12.7.1], but becomes fully developed in the second-order one [Subsec. 12.7.2], as the emotional system of mammals is tightly connected with expectations and with interactive sensorimotor patterns. Tolman had actually already discovered that rats develop some forms of expectancy about the possible reward<sup>42</sup> [Sec. 14.3]. Also, according to Bruner,<sup>43</sup> we should consider behavior in terms of antecedent and subsequent conditions from which we make our inferences or protoinferences. A beautiful manifestation of expectancy at least in some birds and mammals is extrapolation, the ability to cope with a situation in which, say, a trajectory is partly hidden, for instance, to expect that an object rolling down a tunnel will emerge on the other side.<sup>44</sup> Now, coming back to the experiments performed by behaviorists, it is evident that, when we try to condition a mammal (like a dog or a cat), we stimulate an organism with some appropriate input, and it responds by relating the input to some general class of things or events of which it has a representation. A conditional response is a manifestation of the functional equivalence [Subsec. 8.2.3], within a certain tolerance window, of the conditional stimulus with

<sup>39</sup>[THORNDIKE 1898, pp. 100–13].<sup>40</sup>[LLOYD MORGAN 1891, pp. 417–18].<sup>41</sup>[REZNIKOVA 2007, pp. 93–5].<sup>42</sup>[TOLMAN 1932, TOLMAN 1948]. See also [BURGHARDT 1985b].<sup>43</sup>[BRUNER 1957].<sup>44</sup>[REZNIKOVA 2007, pp. 125–8].

the unconditional one, and it can bear the signature of the consummatory behavior relevant to the resource that ensues.<sup>45</sup> Therefore, even if one could try a reformulation of behaviorism in semiotic terms, the question arises of whether behaviorism, even under Skinner's account, which is based on the reinforcement of an operant response (centered on an R-S sequence instead of an S-R one) but is still an input-output theory, is able to truly account for such expectancy. Reinforcement of an *appropriate* behavior seems to go in this direction. However, it also appears to be a very rudimentary form of expectancy, so that it may be said that behaviorism as a whole is not fully able to catch the richness of behavior and its motivations<sup>46</sup> connected with the second-order representational system.

Behaviorism was also experimentally tested in order to ascertain whether or not mammal behavior is indeed dominated by some form of expectancy. I recall here two experiments performed by Capaldi, Nawrocki, and Verry.<sup>47</sup> In the first experiment, all rats received a number of pellets divided into two series of three runs each: 10-0-10 and 10-0-0. This means that the first series terminates with 10, 0, and 10 pellets, and the latter with 10, 0, and 0 pellets. The rats were trained each day, receiving the two series once, always occurring in the same order. That is, the rats could determine which of the two series was to occur. The results of this experiment showed that the rats ran faster on run 2 in the series 10-0-10, a result that could be explained by anticipation of the third run (where they get 10 pellets). The same result was confirmed by a second experiment, where two groups of rats were employed: Group A which received the series 0-0-20 and 20-0-0, and Group D which received the two series 20-0-20 and 0-0-0. The results of this second experiment confirm the above interpretation and also exclude the simultaneous contrast hypothesis, following which the more the contrast between the two runs, the slower the rats are on the second run. However, the results are opposite, thus confirming the anticipation hypothesis. Moreover, these experiments exclude the serial-chaining view, according to which a stimulus S1 causes a response R1 which results in S2, and so on; and confirm the serial mapping view, following which the rats, which are provided with S1, *anticipate* S2. This is a sort of S-S (representational) view that does not fit very well with a behaviorist paradigm but perfectly corresponds to the views advanced by Tolman and Vygotsky.<sup>48</sup>

## 15.2 Ethology

Ethology is the study of behavior patterns from an evolutionary and ecological perspective. It tries to somehow overcome the problem arising from the artificial environment of laboratories. It is therefore observational and not experimental although it also shows some common points with behaviorism. Since it is impossible to give an account of this truly wide field of research (many specific results can be found in different places in the present book), here I will only cover some basic points.<sup>49</sup>

### 15.2.1 The Objectivist School

Both Lorenz and Tinbergen based their work upon the notion of phylogenetic adaptations in behavior (centered on instinct, internal drives, and energy models of motivation)<sup>50</sup>: This was an important conceptual change relative to behaviorism. This school treated behavior as one among

<sup>45</sup>[RUMBAUGH/WASHBURN 2003]. <sup>46</sup>[REZNIKOVA 2007, pp. 54-6]. <sup>47</sup>[CAPALDI *et al.* 1983].

<sup>48</sup>[REZNIKOVA 2007, pp. 66-7]. <sup>49</sup>[CAMPAN/SCAPINI 2002] [HINDE 1982] [TINBERGEN 1963].

<sup>50</sup>[TINBERGEN 1951] [LORENZ 1971].

**472** Behavior

other inherited phenotypic characteristics. Lorenz discovered fixed action patterns (like imprinting<sup>51</sup>) that are triggered by a specific sign stimulus in a specific critical period. Action patterns are instinctive behavioral sequences that are indivisible and run autarchically to completion once started. It may be useful here to introduce the definition of instinct given by Lloyd Morgan<sup>52</sup>:

Instincts are congenital, adaptive, and co-ordinated activities of relative complexity, and involving the behaviour of the organism as a whole. They are not characteristic of individuals as such, but are similarly performed by all like members of the same more or less restricted group, under circumstances which are either of frequent recurrence or are vitally essential to the continuance of the race. While they are, broadly speaking, constant in character, they are subject to variation analogous to that found in organic structures. They are often periodic in development and serial in character. They are to be distinguished from habits which owe their definiteness to individual acquisition and the repetition of individual performance.

It is very important to stress that the instinctive patterns of action commonly found in vertebrates do not reduce to elementary reflexes or reactive behavior [Subsecs. 8.3.1 and 12.5.2]. Indeed, in the 19<sup>th</sup> century Romanes had already pointed out that instincts are not pure reactions<sup>53</sup>: While reactions can be associated with sensations, instincts are connected with perception<sup>54</sup> [Subsec. 3.3.1]. Since instincts are strongly associated with species specific fixed and schematic forms of behavior, they are consequently connected with species-specific affordances [Subsec. 12.5.3]. Therefore, instinct, as it is understood by the school of ethology, should be associated at the very least with what I have called the first-order representational system.

Lorenz's studies on imprinting as a form of learning independent of the consequences of behavior, are now widely acknowledged. One of the most famous examples was that of newly hatched ducks following the first moving object they see (in some cases even the scholar himself). Lorenz pointed out that the fixed action pattern is not directed toward any goal but that it constitutes the end toward which the preceding appetitive behavior is directed.

Lorenz and Tinbergen have shown that behavior is characterized by spontaneity, that is, when external conditions are held as constant as possible, the animal's behavior changes as time passes, hinting at an endogenously generated modification.<sup>55</sup> However, it became evident later on that explanations based on dynamical concepts such as "outward-flowing nervous energy" (Lorenz) or "motivational impulses" (Tinbergen) were too simplistic.<sup>56</sup>

**15.2.2 Instincts and Learning**

Lehrman criticized the separation between genetics and learning introduced by Lorenz and Tinbergen.<sup>57</sup> According to them, innate behavior is consistent enough to be a clearcut category. Obviously there are stereotyped behaviors, as we have seen. However, an animal raised in isolation from its conspecific is not necessarily isolated from the environmental stimuli that could be necessary for behavior *maturation*, which is a fundamental teleonomic process [Subsec. 8.2.1], especially when it is species-specific invariant, while the concrete execution of a behavior has a more teleologic nature [Subsec. 8.2.7]. Moreover, seen in developmental terms, the antinomy inherited/learned even disappears. Indeed, I have remarked that development, and especially epigeny, is a process

<sup>51</sup>[REZNIKOVA 2007, pp. 279–86].      <sup>52</sup>[LLOYD MORGAN 1896, pp. 27–8].

<sup>53</sup>[ROMANES 1882, pp. 10–17]. The behaviorist school prefers to consider instincts as composed reflexes [FROLOV 1937, p. 35]. The problem is how to understand this statement. Strictly taken, it leads to characteristic errors like believing that motor centers of the cortex are not effector organs but receptors of reflexes, the starting-point of which lies in the contraction of the muscles [FROLOV 1937, p. 68].

<sup>54</sup>[ROMANES 1884, p. 159].      <sup>55</sup>[HAILMAN 1977, pp. 21–155].      <sup>56</sup>[ALLEN/BEKOFF 1997, pp. 29–31].

<sup>57</sup>[LEHRMAN 1953].

in which *both* teleonomic and teleologic processes are active [Subsecs. 10.3.3, 11.2.3 and 11.7.2]. Therefore, we may say that behaviors *develop*: They are neither fully learned, nor fully inherited.<sup>58</sup> There is indeed always an interplay between genetic endowment and environment through the phenotype and its interaction with the environment.

Therefore, not only is the antithesis between learning and instinct incorrect, but also the idea that instincts are fixed and immutable seems to be inconsistent.<sup>59</sup> Obviously, there are genetic programs that are more elastic and others that are more closed to external variations. However, this is a matter of degrees that does not touch the essence of the argument. Animals are provided with an instinct to learn specific things. For instance, among birds, even if there is great variability in the sensitive period for learning and even in its individual style, they always show a preference for learning the song of their species. Nevertheless, it is impossible that an instinctive behavior be inherited as such: Instincts cannot be present as such in the zygote (the *homunculus* problem). The essence of development, as we have understood it, is the emergence of new structures from existing structures that may in principle be functionally separated and independent of the former ones [Subsecs. 7.4.5, 8.2.1, 8.2.4, and Sec. 11.4]. In other words, what we have discovered for other phenotypic characteristics is also true for imprinting. It is unfortunately a common error to extrapolate the source of a process without understanding the constraining influence of the context necessary to fully develop it, in a complex process of self-organization starting from that source [Sec. 2.1].

Other exponents of the ethological school have tried to overcome the traditional antithesis genetic inheritance/learning. According to Huber,<sup>60</sup> learning is only one of the several ways an organism copes with an unpredictable environment. In considering the coupling with the environment, Riedl<sup>61</sup> distinguished between the principle of correspondence between external and internal (mutual information, which determines a specific capability to detect some stimuli) and the principle of coherence with the internal conditions of the system (representational autarchy of the system) [Secs. 12.2 and 12.5]. The latter is due to the need for the organism to defend itself from random perturbations (a conservative function). As I have often stressed, external influences are only a type of *trigger* of internal modifications. In this process, both behavior and environment are constructed together rather than given, so that one can speak of an ecological intelligence<sup>62</sup> [Sec. 10.3 and Subsec. 11.7.2]. Both the negative feedbacks from environment (treated, at least initially, on the basis of genetic inheritance) and acquired behavioral capabilities (coming also from nurture, e.g. acquiring hunting skills from parents) converge in highly complex and dynamic developmental processes so that the nature–nurture problem can be overcome. Such kinds of problems are often taken into account for the human case, but they can also be very relevant for all those animals that possess a proper affective system and for whom parental care is important [Subsecs. 12.7.2 and 14.2.5].

### 15.2.3 Reflex, Instinct, Awareness

It could perhaps be said that there is a certain rigidity of species-specific behaviors that results from teleonomic processes of coadaptation, but also that individual variability is always important for behavior, especially when ascending the ladder of complexity in evolution [Secs. 10.2–10.3]. The individual actions of animals always constitute a history not only for themselves but also for their offspring as well as for other species (and therefore for whole ecosystems). It is very helpful here to recall that according to Lloyd Morgan, lower vertebrates already *accompany* their

<sup>58</sup>[HINDE 1982]. <sup>59</sup>[MARLER 1970, MARLER 1991]. See also [REZNIKOVA 2007, pp. 240–9, 254–6, and 267–78].

<sup>60</sup>[HUBER 2000]. See also [AVITAL/JABLONKA 2000]. <sup>61</sup>[RIEDL 1995]. <sup>62</sup>[REZNIKOVA 2007, pp. 261–6].

## 474 Behavior

individual behavior with at least a certain sort of awareness of their environment.<sup>63</sup> In this case, the animal shapes its actions in *cognitive consistency* with past experiences, as experiments on the delayed reaction procedure show<sup>64</sup> (to a certain extent this has also been Skinner's contention with operant conditioning [Subsec. 15.1.2]). This cognitive consistency has also an impact on subsequent generations. In this sense, there is a *difference* between behavior and teleonomic and teleologic processes occurring during epigeny, in which awareness is never involved. The biological origin of this awareness is to be found in the fact that many vertebrates develop sophisticated mechanisms of proprioception [Subsec. 5.3.2] that allow for the continuous checking of both internal and external states and therefore for increasing control of behavior. Then, the roots of awareness can be found in attentional processes [Subsec. 14.1.1]. These mechanisms can initially be reflex-like, but their composition can give rise to new functions [Subsec. 9.5.3] like cognitive consistency.<sup>65</sup> Awareness of environment is in accordance with what I have called the first-order representational system [Subsec. 12.7.1]. We should therefore distinguish this form of awareness from the basic sense of the biological self [Sec. 8.4] that already goes together with a pure reactive behavior.

It is also important to understand that the awareness accompanying behavior has phylogenetically evolved. Mammals (and in part birds) have awareness also of *their own* actions and their impact on other animals [Subsec. 12.7.2], which is much more than mere awareness of the environment.<sup>66</sup> However, it is no less important to understand that in *any* behavior there is some instinctual component, so that the interaction between instinct and learning is one of the most important but difficult issues.

This is precisely Lloyd Morgan's approach, which was rejected by Watson, who tried to prove that there is no awareness at all (nor in the first modality) and that any instinctive behavior could be reduced to a sum of congenital and reactive reflex segments.<sup>67</sup> For Watson, habit formation can also be reduced to this reflex-like activity. For this reason, Watson was sympathetic to Lamarckian theory.<sup>68</sup> Obviously, Watson was fully aware that an organism can react in different and individual ways to certain stimuli.<sup>69</sup> However, he assumed a compositional principle of causality [Subsec. 2.4.2], according to which these different forms of behavior can be interpreted as the results of several physiological processes operating in parallel in any organism and determining different reactions in various individuals. These different psychological processes reduce ultimately to basic stereotyped reactions. Indeed, Watson maintained that reactions can be fully controlled and predicted, and therefore behavior can be too.<sup>70</sup> This went together with the use of an experimental procedure for testing behavior [Subsec. 15.1.1].

However, what we seem to lose here is the insight suggested by ethological studies: That a true organism in a true environment is able to have a real individual (not repeatable) history. It is likely that Watson's explanation is sufficient or partially sufficient for invertebrates and lower vertebrates, where the distinction between memory, learning, and behavior is not yet clear (as we shall see in the next chapter). But it is certainly not an adequate explanation for developed animals like birds or mammals. Here, the organism cannot be taken in isolation from the external environment or its

<sup>63</sup>[LLOYD MORGAN 1896, pp. 126–31 and 147] [LLOYD MORGAN 1900, pp. 100–6]. This was also perfectly clear to Bichat who, in this regard, spoke of two kinds of life [BICHAT 1800, pp. 46–58].

<sup>64</sup>[TINKLEPAUGH 1928, TINKLEPAUGH 1932] [HARLOW *et al.* 1932] [GRIFFIN 2001] [REZNIKOVA 2007, pp. 89–92].

<sup>65</sup>Understood perfectly in [SHERRINGTON 1906, pp. 129–32 and 355–6].

<sup>66</sup>Thorndike denied the existence of such a social awareness [THORNDIKE 1898, pp. 102–3]. If we consider *social* awareness of others' behavior, I think that this should be confined to primates, on which point Thorndike would perhaps agree.

<sup>67</sup>[WATSON 1914, p. 184] [WATSON 1925, p. 13]. <sup>68</sup>[WATSON 1914, p. 174].

<sup>69</sup>[WATSON 1914, pp.107–8]. <sup>70</sup>[WATSON 1914, p.10] [WATSON 1925, p. 16].

conspecifics. It is unnecessary to reduce the instinctive behavior and the accompanying awareness to a sum of reflexes. In short, what we risk losing is the systemic totality of the organism and its environmental and social situatedness that is the *raison d'être* of ethology.

Considering Watson's (and also Pavlov's, as we have seen) stress on behaviorism as a scientific treatment of psychological processes, some words on science's methodology seem appropriate here. Science always begins with models that introduce strong simplifications. This is methodologically necessary, otherwise we would not be able to explain what is more complex. At this early stage of science a pure reductionist and even mechanist explanation is the only methodological tool at our disposal. However, as science continues, a level of complexity is sooner or later attained where such methodology appears insufficient. The science of the 20<sup>th</sup> century, with quantum mechanics and the theory of complexity, has brought us to this awareness. It is now time for biology, neurology, and psychology to go along this path.

### 15.3 Mirror Neurons and Purposeful Operations

As I have stressed [Subsec. 15.1.2], behavior determined by the third-order representational system was not at all considered by behaviorism, and neither was it considered by the school of ethology at its first stages. Now, as we have seen [Subsec. 12.7.3], new studies in the ethology of primates have contributed to changing this approach.<sup>71</sup> Nevertheless, an important contribution for understanding this form of behavior does not come from studies in ethology but from studies in neurology, performed by a team of Italian scholars.

Let me first develop some general considerations on the way behavior is selected and monitored in primates. When the same stimuli are involved in *several* possible behaviors in different contexts, the task-specific activity could give a signal that allows ambiguous or conflicting sensory information to be mapped to the appropriate motor output.<sup>72</sup> Task-specific activity in the PF cortex could function, via top-down signals, to bias the activity of sensory systems towards the representation of information relevant to the *task*.<sup>73</sup> It is a kind of high-sophisticated reentrant mechanism [11.5.3]. When sight is involved, stimulus or saccade-direction selectivity whose magnitude differs with the current task indicates that some PFC neurons do not simply reflect single stimuli or isolated forthcoming actions. Rather, this suggests that *behavioral context* (i.e. information associated with the cue or saccade that is unique to a *particular task* or the manner in which an object is used) modulates PFC activity. For example, a neuron that is apparently selective during a "pure" object memory task (the so-called "object task") does not necessarily exhibit selectivity for the same objects in other contexts (e.g. the "associative task"). The fact that many neurons do reflect a given object or saccade regardless of task indicates that both sensory information and convergence towards motor output are indeed present in the PFC [Subsec. 14.1.1]. But the existence of *task-specific selectivity* suggests that the PFC also has information about what is "in between," i.e. the mechanisms for mapping sensory input to motor output [Sec. 13.2], and is able to modulate this connection. The responses of neurons in the lateral PFC and frontal eye fields to a visual target can therefore dramatically differ depending on the rule used to deal with the target. Together, these results support the notion that the information conveyed by PFC neurons is not limited to discrete sensory events or to specific motor executions. Rather the dynamical behavioral context in which the animals are engaged has a pervasive influence on PFC activity, and in this situation we can speak of purposeful behavior [Subsec. 14.13].

<sup>71</sup>[TARTABINI 2003].

<sup>72</sup>[ASAAD *et al.* 2000].

<sup>73</sup>See also [OBAYASHI *et al.* 2001].

## 476 Behavior

Therefore, the sequence of events that characterizes behavior can be changed by varying the information that is available to the agent.<sup>74</sup> This explanation simply captures relevant generalization and makes predictions that cannot be accounted for by a lower level of explanation. In primates, the aim of the motor system is at least the making of internal copies of actions for both generating purposeful actions (motor acts) and understanding purposeful motor events.<sup>75</sup> At a basic level, the motor system deals with intentions in action, consisting of single action patterns (shake a hand, push a stone, and so on) that primates can not only execute but also follow and recognize (in their task significance). Sometimes they can even follow short sequences of action patterns, which normally already express a higher level, namely that of programming.

The discovery of mirror neurons in the frontal lobes of macaques and their implications for human brain evolution is one of the most important findings of neuroscience in the last decades and allows us to understand primates' behavior in a new way. Rizzolatti and colleagues at the beginning of the 1990s found that in monkeys' cortex areas F1–F7 there were mirror neurons, i.e. neurons that discharged when an action was *performed* and also when it was *only observed*. Mirror neurons are therefore active when the monkeys perform certain tasks, but they also fire when the monkeys watch someone else performing the same specific task. One of the first findings in this area was the discovery that a monkey's brain shows the same spatial patterns which the hand motion will follow when the action is executed a fraction of second later, which could mean that the animal was *programming* the movement before carrying it out<sup>76</sup> [Subsec. 14.1.3]. In particular, the neurons of area 6a $\beta$  are not influenced by the location of the object, nor whether it was grasped or not. Their brain activity changed before the arm movement and continued until the end of it.<sup>77</sup>

There is evidence which shows that a similar observation/action matching system exists in humans. One of the most relevant points is that among the involved areas F1–F7, the area F5 corresponds to the Broca's language area (Brodmann's areas 44 and 45) in humans. This is very interesting, due to the fundamental *interchangeability of roles* that characterizes language and the ability to *share intentional acts* that are common to any symbolic exchange, as we shall see in the next part of the book. Mirror neurons are probably the basis of a gestural protolanguage as they display forms of reciprocal acknowledgment.<sup>78</sup>

It is important to understand that mirror neurons are not activated when the action is only pretended and not truly performed: It must be a truly purposeful operation. Moreover, there is a strong task-dependence, since different neurons are activated when making the same gesture for eating or for putting a piece of food in a box<sup>79</sup> [Fig. 15.6]. For this reason, mirror neurons represent the bridge between a basic replicative behavior and true imitation.<sup>80</sup>

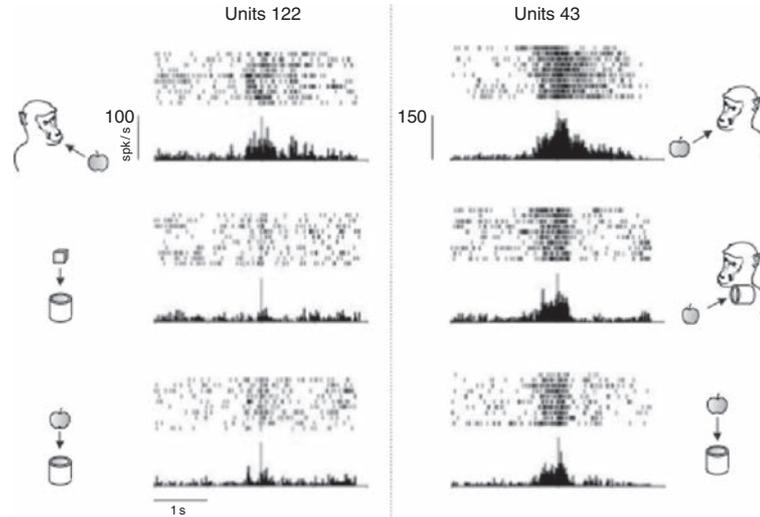
To be able to acknowledge single action patterns, and eventually elementary combinations of them, requires the identification of the specific situation, which in turn implies a categorial level (a perceptual connection between events and in particular an understanding of causal relations), while no abstract concepts, or mind-reading are necessary here, as is often assumed. However, it also demands the ability to understand the fact that the other agent is dealing with a specific object or event. As we have seen, this is typical of the third-order representational system, but does not necessarily imply a true imitation. Obviously, many actions performed by chimpanzees seem to imply a capability to imitate, since they combine the ability to understand the functionality of

<sup>74</sup>[SHIMA *et al.* 2007]. <sup>75</sup>[RIZZOLATTI *et al.* 2000].

<sup>76</sup>[GEORGOPOULOS *et al.* 1989] [GRIFFIN 2001, pp. 149–50]. <sup>77</sup>[RIZZOLATTI *et al.* 1990].

<sup>78</sup>[RIZZOLATTI/ARBIB 1998][ARBIB 2005]. <sup>79</sup>[FOGASSI *et al.* 2005].

<sup>80</sup>This also explains why human children acquire the distinction pretend/real *before* the distinction apparent/real [FLAVELL *et al.* 1987].



**Fig. 15.6** Discharge of two IPL neurons during active grasping. Unit 122 strongly discharges when the monkey grasps a piece of food to eat (top), whereas it does not respond when the monkey grasps an object (center) or a piece of food (bottom) to place. Unit 43 strongly discharges when the monkey grasps a piece of food to eat (top), whereas the discharge is significantly weaker (12) when the monkey grasps a piece of food to place into a container positioned near the mouth (center) or near the grasped object location (bottom).

Rasters and histograms are synchronized with the moment when the monkey touched the object to be grasped. Red bars, monkey releases the hand from the starting position; green bars, monkey touches the container;  $x$  axis, time;  $y$  axis, discharge frequency. Adapted from [FOGASSI *et al.* 2005]. (This figure is reproduced in color in the color plate section.)

objects (a pure categorical problem) with the social understanding witnessed by mirror neurons [Subsecs. 4.5.2–4.5.3]. Again, we see the principle that the lower level of dealing-with-information is somehow able to anticipate and almost to mimick higher levels [Subsec. 12.7.1].

## 15.4 Concluding Remarks

Behavior is born in the intersection between emotional and decisional systems:

- Traditionally, behaviorism was the most influential school for studying behavior. Behaviorism has two main varieties. In the Pavlovian variety (respondent conditioning), involuntary responses to an unconditional stimulus can become conditioned to an associated conditional stimulus. It consists in old responses to new stimuli.
- In the Skinnerian variety of behaviorism (operant conditioning) a reinforcement (positive or negative) will occur only with an appropriate response of the organism (and will not simply accompany another stimulus in a purely associative way). It consists in new responses to old stimuli.
- Though behaviorism is very good at explaining reactive behavior and behavior based on perceptual schemata, it is nevertheless inadequate at dealing with the visceral and interactive sensorimotor schemata of mammals and especially with the behavior of primates. In general it

**478** Behavior

strongly underestimates the fact that in any behavior the reaction to a stimulus is mediated by a semiotic process, in which expectancy plays a major role.

- Behaviorism has underestimated the role of environmental factors and especially the awareness of environment, shown by most vertebrates, and the awareness of the consequences of one own's actions, typical of mammals.
- Primates are able to perform purposeful operations, which are reflected by the recent discovery of mirror neurons that fire both when an action is executed or watched.

Having established these general features of behavior, in the next chapter I shall consider another of the brain's functions: Learning. We shall see that, even if learning shares many elements with behavior, it is a different brain functionality.

# 16

## Learning

---

In this chapter, after some introductory remarks and an exposition of the general features of learning, I shall introduce associative and conditioned learning. Next, we shall focus on the difficult issue of learning causal relations. We shall see that old mammals like rodents are already able to do that. The next step will be the examination of learning through categories, especially focusing on primates' understanding of analogies and whether they are able to treat categories as true models of the world.

### 16.1 What Learning Is

Learning and behavior are often confused, as we have seen [Subsec. 15.1.3], since both are selective actions and are very tangled. In fact, learning can modify behavior as much as behavior can have effects on learning. However, they are different functions of the organism. Learning is a postnatal developmental mechanism that, like reaction diffusion mechanisms or positional interactions during epigeny [Subsec. 11.3.2], even if at another level and with other means, acts to establish correlations among traits in order to give rise to a self-reinforcing process between the environment and the phenotype: The better equipped individuals are those that will have more rewards, which again reinforces this acquired equipment.<sup>1</sup> In this way, learning mimics natural selection and is the true center of the ontogenetic activity of the phenotype. Here, behaviorism, especially in Skinner's variety, has individuated a relevant point [Subsec. 15.1.2].

The organism is able to acquire new forms of skilled behavior on the background of already existing capacities, which are not necessarily innate [Subsecs. 15.2.1–15.2.2].<sup>2</sup> The extent to which specific behavioral information cooperates or competes with spontaneous self-organizing tendencies determines the resulting patterns and their relative stability. The interplay between these two sources of information is a selection mechanism. In other words, as happens in epigeny [Sec. 11.2], we have an interplay between learning requirements (based on environmental inputs) and intrinsic organizational tendencies.

It may help to consider learning as taking the form of a phase transition. This approach supposes the ability to transfer knowledge to *other situations*, that is, to establish appropriate linkages [Subsec. 15.2.3]. Since in certain situations it can be shown that the transfer of a new skill, for instance from one hand to the other, occurs with the same timing relations but with an inverse ordering (left–right), then what is learned is a phase relation that is apparently *independent* of how the operation is specifically instantiated [Subsec. 3.3.3 and Sec. 14.3, as well as Fig. 12.13]. The phase relation is then an *order parameter* determining the arising of intrinsic organizational

<sup>1</sup>[WEST-EBERHARD 2003, p. 338].

<sup>2</sup>[KELSO 1995, pp. 159–285].

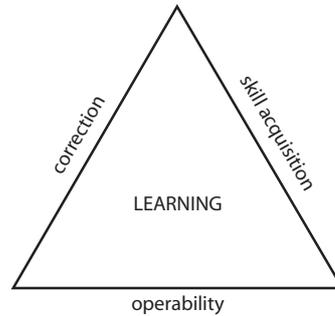


Fig. 16.1 Learning as a process consisting in operability, error correction, and skill acquisition.

tendencies [Subsec. 6.5.4]. This also shows that learning only consists of shared information with control but not of a physical transfer.

Let us consider this problem from a more general point of view. Learning is not only bound to the decisional center but also to the cognitive one [Sec. 13.3]. For this reason, in contrast to behavior, learning is necessarily bound to *some generalization* (which does not imply that learning is instructed in any sense<sup>3</sup> [Subsec. 9.2.1]). In other words, learning is not the *material* transfer (the immediate content) of information from one cerebral circuit to another, but the transfer of a general form, pattern, or principle. This is beautifully shown by an experiment performed by Rumbaugh and coworkers.<sup>4</sup> A monkey able to manipulate a joystick with its feet only, once allowed for the first time to do it with its hands, succeeds better than individuals who have never manipulated joysticks (either with hands or with feet). This shows that it had learned the *general* structure of the movement and not the *specific* performance with its feet. For this reason, reinforcement is not the reason for learning, but is only a form of reward, and can eventually be a form of motivation [Subsec. 15.1.3]. Reinforcement is nothing but a salient stimulus (a supplementary selection) that has a function of eliciting behavior needing other salient stimuli to start [Subsec. 15.1.4]. At most, it informs the organism about contextual resources and how they can be accessed by appropriate behavior (due to the mechanism of operant conditioning).

In the examples provided here we are concerned with operative skills, for which visuomotor representations are particularly relevant [Sec. 5.2]. This is very important. Indeed, when dealing with learning in its generality (as something characterizing many animal species), we should avoid the danger of being overinfluenced by a human school-like type of learning (that cannot be generalized to all human cultures either). As we shall see, this is rather a very specific and complicated case that can confuse the main issue at stake here.

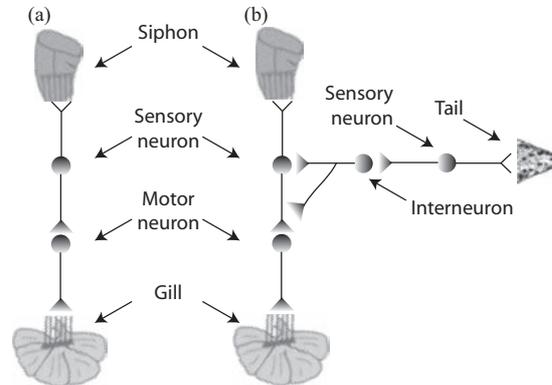
Summarizing, *learning* can be defined as a postnatal activity aiming at *acquiring operative skills* through *correction* due to interaction with an environment (providing negative feedback). It is therefore characterized by correction, skill acquisition, and operability, where by *operability* I mean that after the stage of learning, the subject is able to perform goal-directed or purposeful operations for which that skill is fundamental or relevant [Fig. 16.1].

## 16.2 The Problem of Learning

The main problem of learning is the following: Most regularities only enjoy an attenuated existence in a body of training data, which means that in a cluster of training data, regularities (especially

<sup>3</sup>[PIATTELLI PALMARINI 1989].

<sup>4</sup>[RUMBAUGH/WASHBURN 2003, pp. 199–236].



**Fig. 16.2** A very schematic overview of the learning circuit in *Aplysia*. (a) Mediating circuit of learning. (b) Cooperation between mediating (vertically arranged) and modulating (horizontally arranged) circuits of learning [see also Figs. 12.7 and 15.2]. The interneurons involved here release serotonin. Neurons are represented by circles, synapses by triangles. Adapted from [KANDEL 2006, p. 224].

if they have a high complexity) are not evident and very difficult to individuate. This is due to the underdetermination of any regularity relative to data [Subsecs. 2.4.3 and 4.4.5, Sec. 12.2]. These are therefore regularities whose statistical visibility depends on some systematic recording of the data. However, the space of possible recording is infinitely large since it is the number of possible ways to record. Therefore, mappings based on such attenuated regularities are very difficult to achieve by a brute-force search (i.e., the systematic enumeration of possible solutions, each of them being eventually checked as appropriate or not).<sup>5</sup> As we shall see, this is also the so-called induction problem.

The traditional cognitive program has tried to solve this problem by assuming that learning mechanisms are computationally specialized for solving particular kinds of problems.<sup>6</sup> It is true that living organisms use different systems for learning. However, it is probable that the specialized systems rather reflect the *result* of learning, that is, the transfer of learned abilities to circuits that can perform their behavioral activity in a more mechanical and automatic way. These hard-wired circuits can contribute to subsequent learning processes, so that they only constitute a part of the problem. Indeed, there are at least two kinds of neural circuits involved in any form of learning<sup>7</sup>: A mediating circuit and a modulating one [Fig. 16.2]. *Mediating* circuits directly produce behavior and genetically and developmentally represent determined components of behavior, whereas *modulating* ones serve as teachers and are not directly involved in producing a behavior but rather in strengthening certain connections. As with the immune system [Subsec. 12.4.4], we have two main functions here: Information-processing and regulation. As I have said, interneurons [Subsec. 3.3.2] induce changes that last some minutes (allowing for a very rudimentary short-term memory to work). This is due to the different mechanisms ruling release of serotonin relative to other neurotransmitters.<sup>8</sup> Consequently, one should try to find a middle way between (genetic) innatism and models exclusively based—such as behaviorism in its respondent-conditioning form—on at least some kind of inductive-associative learning<sup>9</sup> [Subsec. 15.1.2].

<sup>5</sup>[CLARK/THORNTON 1997].

<sup>6</sup>[HOLYOAK 1991][GALLISTEL 2000].

<sup>7</sup>[KANDEL 2006, pp. 222–34].

<sup>8</sup>[CASTELLUCCI *et al.* 1980].

<sup>9</sup>[CARNEVALI/PATARNELLO 1987].

**482** Learning

For higher animals equipped at least with schemata, learning presupposes more complex teleonomic circuits, i.e. species-specific affordances, which are properties of the environment as they are related to animals' capabilities of using environmental tools for their needs<sup>10</sup>: Recall that affordances require reciprocal co-adaptation between the organism and the environment, and are strictly related to niches that are unique to species [Sec. 10.3 and Subsec. 12.5.3]. In this case, we speak of *guided learning*.<sup>11</sup> Consequently, changes in the environment or changes induced by the organisms are fundamental for extracting information, as they allow for both the continuous shifting of viewpoints on a stable background and the constitution of relatively stable frameworks included in their remolding.

Therefore, learning can arise from specific forms of *goal-directed activity* when facing new kinds of problems.<sup>12</sup> To perform learning of some specific type, the learner must be adequately oriented by its own inbuilt goals. One learns by doing: *Consequences of action* and not antecedents constitute the essence of learning<sup>13</sup> [Subsec. 12.5.1]. We have seen that this was also Skinner's main concern with his theory of selection by consequences. Here is the main difference with *behavior*, in which *both* antecedents and consequents of action play a role, as well as with *memory*, in which *only* antecedents play a role. For this reason, learning is about the predictive relations between stimuli<sup>14</sup>: Having focused on the consequences of some action, learning is also able to guide future action in a feedback circuit.

These conclusions are in accordance with the results of the previous chapter. As we have seen, Thorndike<sup>15</sup> came to the conclusion that the mere repetition of the same situation as well as the mere frequency of a certain response do not cause learning [Subsecs. 15.1.3–15.1.4]. In order to learn, there must be a sense of belonging (that is, the attribution of a certain value to a certain sequence of events, including the evaluation of one's own response as right), which also demands a specific attention (a supplementary selection [Subsec. 14.1.1]) toward the stimulus and the aspects that are considered to be salient.<sup>16</sup>

### 16.3 Associative and Conditioned Learning

At a previous stage of evolution, learning and memory have probably developed as a differentiation from an initial very undifferentiated brain organization, whose main subsystems consisted of a sensory and a motor system [Subsec. 13.1.3]. At this level, we have a pure associative organization of the brain that will evolve in evolutionary time. In particular, animals dominated by noninteractive and automatic sensorimotor schemata [Subsec. 12.7.1], are at a stage in which behavior is not clearly distinguished from learning and memory. This stage might very roughly correspond to Pavlov's *respondent* conditioning. As time flows, consequences of behavior become increasingly important and *operant* conditioning, although not completely absent in more rudimentary life forms, becomes crucial. In this context, it is therefore important to distinguish between pure association (which probably cannot be called learning for the reasons already explained and which is, on the contrary, fundamental for memory<sup>17</sup>) and the discovery of regularities [Subsec. 12.4.1] that demands some kind of ability to respond to new patterns, which is typical of mammals. The law of effects, being the basis of operant conditioning [Subsec. 15.1.1], can help the transition to a differentiation between behavior and learning.

<sup>10</sup>[GIBSON/PICK 2000, pp. 14–44].      <sup>11</sup>[GOULD/MARLER 1987] [REZNIKOVA 2007, pp. 250–66].

<sup>12</sup>[NEWELL 1990, pp. 310–22].      <sup>13</sup>[DEWEY 1929, p. 154].

<sup>14</sup>[RUMBAUGH/WASHBURN 2003, p. 39] [POGGIO/BIZZI 2004].      <sup>15</sup>[THORNDIKE 1931, pp. 6–29, 64–81].

<sup>16</sup>[WILLINGHAM 2001, pp. 185–7].      <sup>17</sup>[WUNDT 1907, Sec. 16].

In order to understand this point, let us reconsider some concepts that are relevant to the behaviorist theory of learning:

- (i) Facilitation by practice and *sensitization*—a form of reference to a salient or unexpected event.<sup>18</sup> Sensitization is the process by which the organism has learned that what will follow a first stimulus is not trivial. This is the specific aspect leading to learning.
- (ii) Association (*habituation*,<sup>19</sup> habit, traumatic avoidance, and imprinting) [Subsec. 3.3.2]: it is the formation of a linkage between two nervous processes that were hitherto not causally connected [Subsec. 15.1.2]. For this reason, it was assumed that it is based on a contiguity principle. Habituation is when nothing of relevance is expected to follow.<sup>20</sup> However, habituation is rather connected with *memory*.
- (iii) *Reinforcement*, a learning through success and failure which presupposes a feedback circle.<sup>21</sup> As we have seen, reinforcement is a behavioral category that does not necessarily reduce to the Pavlovian variety of conditioning. In the case of learning, it would be better to speak of skill acquisition [Fig. 16.1].

Habituation or sensitization alone do not constitute learning, but these two aspects together with reinforcement constitute the law of effects, which then becomes the bridge to *learning*.

According to the traditional (behaviorist) point of view, the most elementary stage of learning is trial-and-error.<sup>22</sup> This is learning by *trial* and accidental success, consisting in the strengthening of the connections between the sense impressions representing the situation and the acts—or impulses to act—representing our successful response to it, and by the inhibition of similar connections with unsuccessful responses. Its core is represented again by the law of effects. It is doubtful that trial-and-error can be considered as such as an autonomous form of learning. It is probably some basic form that recurs at any level of the evolutionary ladder and expresses the circumstance that many animals (including humans) are able to learn by some accidental consequences of their actions. But, as I have remarked, learning starts with (even innate) goals and not with accidental trial. We have indeed seen how instinct can be combined with different kinds of awareness [Subsec. 15.2.3]. Now the specific function of awareness is the following: It is what is *necessary for learning* in the proper sense of the word, precisely because learning is concerned with the consequences of an action. The plastic combination of learning and instinct can finally result in a constitution of habits, which will be the subject of the next chapter.

Therefore, considering the behaviorist theory of learning as providing the bridge to true learning or at least as describing an early stage of learning, we can make a start with Thorndike's classification of learning<sup>23</sup> but introducing some substantial modification:

- (1) *The associative stage* in which learning and behavior are not yet clearly differentiated [Subsec. 6.1.5], with two probable substeps represented by respondent and operant conditioning.
- (2) Learning by *understanding the consequences of our own actions*. This is connected with the second-order representational system and is common in old-mammals.
- (3) Learning by *understanding the consequences of others' actions*. This is connected with the third-order representational system and is typical of primates.

We also have symbolic learning, which shall be specifically treated in the next part of the book.

<sup>18</sup>[LAMARCK 1809, I, pp. 93–4].      <sup>19</sup>[DE BIRAN 1803, pp. 73–84].      <sup>20</sup>[REZNIKOVA 2007, pp. 41–2].

<sup>21</sup>[BITTERMAN 2000]. See also [BREMBS *et al.* 2002].      <sup>22</sup>[REZNIKOVA 2007, pp. 45–7].

<sup>23</sup>[THORNDIKE 1898, p. 174].

## 16.4 Learning Causal Relations

Old-mammals are able to understand (are aware of) the consequences of their purposeful actions on other animals [Subsec. 15.2.3], Point (2) of the previous section. This is strictly connected with the second-order representational system since here the animal is provided with interactive sensorimotor schemata enabling it to induce some reaction in an another agent [Subsec. 12.7.2]. Being capable of that, the animal can also understand causal mechanisms, which are, at this stage, strictly connected with the animal's modalities of action and its expected consequences [Sec. 14.3 and Subsec. 15.1.4]. We see here the fundamental *cognitive* relevance of the emotional system [Subsec. 14.2.5].

Let me consider first the problem at a purely abstract (system-theory) level. To understand causal relations amounts to being able to distinguish between associations where an event E1 produces an event E2 and other associations where E1 is not the cause of E2: Mammals can learn this distinction.<sup>24</sup> Abstractly and statistically speaking, in order to learn that E1 causes E2, the condition is that the probability that E2 occurs shortly after E1 is greater than that of E2 not occurring when E1 occurs. However, this purely associative mechanism only represents one of the conditions for individuating a cause. The problem is represented by the fact that in the real world there are far more potential causes of E2 than E1 alone. As such, we must understand which mechanism is the one that enables animals to isolate a specific cause and individuate causal relations [Sec. 16.2]. In a first approximation, conditions for learning causal relations seem to be (i) how causally relevant the two events are and (ii) how close the temporal relationship is between E1 and E2 [Subsec. 15.1.2]. Obviously, learning the correlation between E1 and E2 is simplified if the occurrence of E2 is surprising or unpredicted in the given environmental conditions (sensitization) [Secs. 12.2 and 16.3, Subsec. 12.5.2]. There are two possibilities in this sense:

- (a) The amount of information-processing induced by an event depends upon the extent to which the event itself is predicted. The traditional behaviorist Rescorla–Wagner model focuses on discrepancy and variations in the processing of E2<sup>25</sup>: To the extent that E2 is predicted, the animal fails to learn. This model provides an account of overshadowing (of E1 by other environmental factors) and blocking, but fails to account for observed variations in the processing of E1 that occur when the animal is exposed to this event prior to learning.<sup>26</sup>
- (b) Processing of E1 depends upon whether or not it has been an *effective predictor* of important events in the animal's *past experience*. Here there are again two possibilities. Mackintosh<sup>27</sup> suggested that E1 will be effectively processed if this event has been a *good* predictor of other events in the past, while, according to Pearce and Hall, such processing occurs only if E1 has been recently associated with an *unpredicted* E2. However, it is again difficult to see how the animal could pick out the *right* predictor in the past, considering that the probabilities of a chance prediction are near zero in a real environment.

As mentioned, in the context of the conditioning theory one deals with this problem by assuming a temporal contiguity between the involved events. However, this is not sufficient: Rats avoid the ingestion of food that makes them sick according to the rule of rejecting the last food eaten before becoming sick. Since this can happen even hours after consumption, the finding violates the cardinal principle of behaviorism, namely, that the association between stimuli should be coincidental or proximate<sup>28</sup> [Sec. 15.1]. These problems could be solved if E2 is a consequence of

<sup>24</sup>[DICKINSON 1980]. <sup>25</sup>[RESCORLA/WAGNER 1972][RESCORLA 1980].

<sup>26</sup>Wagner modified the model on this point but some problems still remain. <sup>27</sup>[MACKINTOSH 1974].

<sup>28</sup>[REZNIKOVA 2007, pp. 64–6] [LEDOUX 2002, p. 127].

an action *performed by the animal itself*. In this case, if it is able to understand the consequences of *its own* action on both its natural and social environment, it will also understand what is *unexpected* and be able to make appropriate predictions about the cause of the perturbation.<sup>29</sup> The reason is that a performed purposive action is a *decision* that lowers the number of possible causal factors by a great deal when the result is different from the expected one. A Pavlovian agent, lacking in the ability to act purposively, is at the mercy of a fortuitously occurring stability of the connection between its behavior and events in the world<sup>30</sup> (a conditioned reflex is only a substitute and has no predictive power in the strong sense of the word [Fig. 15.3]). This is also the weak point of Skinner's theory, as remarked above [Subsec. 15.1.4]. Indeed, Skinner takes the behavior of pigeons performing some arbitrary action after one has repeatedly presented food without any connection with their previous behavior, as a manifestation of the fact that only a coincidental "superstitious" relation has been established<sup>31</sup>: He seems to totally miss the relevant point here, namely that the pigeon manifests, still at a very elementary level, the exigency to connect in a causal way its *own* actions with a subsequent happening in the external world, which is tremendous progress relative to the schematic forms of action that still characterize reptiles or amphibians.

Turning the argument around, I suggest that purposeful actions presuppose (i) a representation of the *causal* relationship between a certain action (which is under the full control of the animal) and its outcome, (ii) a representation of the beneficial value of a certain specific outcome [Subsecs. 12.5.3 and 15.2.3]; (iii) an emotional interface between the two representations above: The *desire*. Desires are grounded in the affective reaction to potential targets.<sup>32</sup> When there is a purposeful behavior, a representation of a certain *causal* relation between the action and its outcome together with an appreciation of the *value* of this outcome and the *desire* for the outcome constitute the *antecedent* of a practical proto-inference [Sec. 12.7]. Studies show that there are similar psychological processes underlying human causal judgments and simple instrumental performances by rats. Not by chance, both rats and humans are sensitive to similar causal illusions. Given the existence of very similar causal representations in humans and rats, human causal judgments and animal action should not represent independent but convergent processes.

Studies on young children also confirm that simple coincidental co-occurrence (the ground of associations) is not the reason for learning causal relations<sup>33</sup> [Subsecs. 15.1.3–15.1.4]. The idea of cause may be induced even during a *single* experimental run. Repeated co-occurrence is essential only in *already learned* causal relations. Furthermore, one can be exposed to the repeated co-occurrence of two events without being able to form any idea of the cause. If the sense of causation were built through association, children, relative to adults, should have a weaker perception of causes, but experiments of Leslie and associates show the contrary to be true.<sup>34</sup> Therefore, associationism [Subsecs. 3.8.1 and 6.1.5], for instance Hume's theory of causes,<sup>35</sup> is not the correct explanation of causal understanding.<sup>36</sup> In other words, the fact that causal relations are *inferred* does *not* mean that they are *pure associative constructs*. A splendid proof of this has been found by T. Shultz<sup>37</sup>: Three-year-old children understand that the cause of extinguishing of a candle is a blower turned on but positioned far away from the candle rather than a blower turned off but positioned closer to it. That is, they prefer to adopt a principle of generative transmission at a

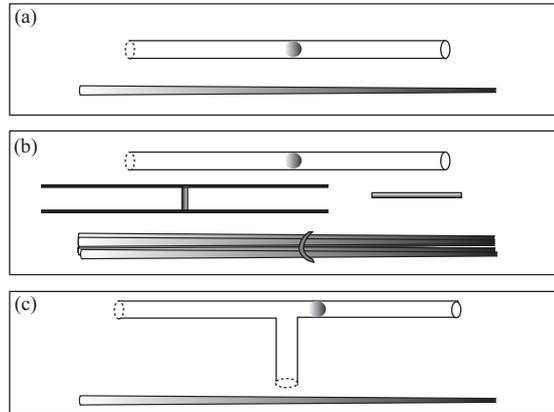
<sup>29</sup>Evidence for the fundamentality of this structure is the fact that in humans the most basic and ascertained form of self-awareness touches self-generated, intended actions relative to those externally generated [BLAKEMORE/FRITH 2003].

<sup>30</sup>[DICKINSON/SHANKS 1995]. <sup>31</sup>[SKINNER 1948]. <sup>32</sup>[DICKINSON/BALLEINE 2000].

<sup>33</sup>[PREMACK/PREMACK 2003]. <sup>34</sup>[LESLIE 1994, LESLIE 1995] [BARON-COHEN *et al.* 1985].

<sup>35</sup>[HUME 1777, pp. 32–47]. <sup>36</sup>[PLOTKIN 1997, pp. 179–84].

<sup>37</sup>[SHULTZ 1982]. See also [POVINELLI 2000, pp. 89–90].



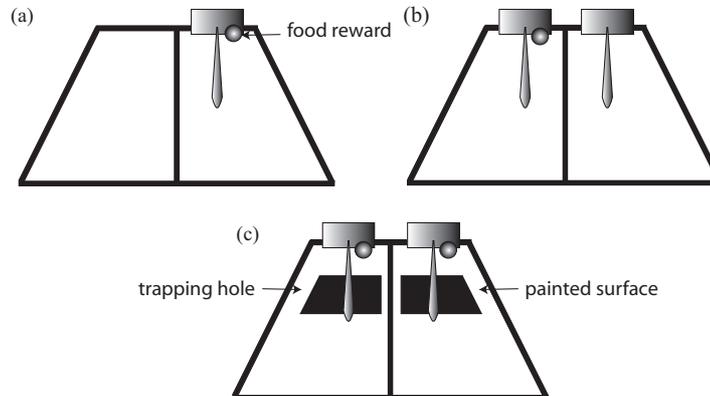
**Fig. 16.3** The test performed by Visalberghi and Limongelli. (a) The orienting test with a transparent tube with a reward inside and a stick. (b) The task here is to use the appropriate stick and to perform the appropriate actions for obtaining that treat. (c) The task here is to avoid the treat falling into the vertical segment of the tube.

distance instead of a pure assumption of spatial contiguity. Shultz's experiments could be interpreted as simply showing cultural familiarity to devices like blowers. However, Shultz successfully performed the same experiments on African children who were not familiar with these objects.

An important point is whether nonhuman mammals are able to grasp causal relations as totally *objective* or only relative to the agent's point of view. Some experiments on tool use can help us. Tool use is not completely independent from the agent but establishes a certain distance or mediation between the agent's action and effect, since the objective properties of the tools as well as the objective context are much more important than when the animal acts without intermediation. Chimpanzees, children, and capuchin monkeys were tested in three different tasks<sup>38</sup> [Fig. 16.3]: (1) To push a treat out of a transparent tube using a stick; (2) to push a treat out of a transparent tube but having a bundle of sticks at their disposal (a bundle that is too wide to enter into the tube), an H-shaped stick, and short sticks; (3) to push a treat out of a transparent trap tube that is T-shaped. The point of this latter test is to verify whether or not the animal is able to obtain the treat (the reward) and avoid it falling into the orthogonal tube. The first task was solved by all participants. The second task was solved by all chimpanzees and children but by only one monkey. Moreover, the performance of capuchins did not significantly improve across trials. The third task was solved by chimpanzees and children only. The conclusion here could be that capuchins seem to live in a world where every possible option can be tried, while humans and apes discard some possibilities beforehand. There is another and more interesting possibility, already mentioned in the previous section: Animals (including humans) have recourse to trial-and-error methods (like the capuchin monkeys in this experiment) when they have no framework to guide them in dealing appropriately with some situations. If I am right, trial and error is not an evolutionary early stage of learning but a *late ontogenetic form* of dealing with problems when there is *no previous habit* or appropriate instinct.

Further experiments with the trap-tube device<sup>39</sup> have even questioned the ability of chimpanzees to understand causal relations. Indeed, we shall see that they do lack a true ability

<sup>38</sup>[VISALBERGHI/LIMONGELLI 1996]. See also [REZNIKOVA 2007, pp. 196–205]. <sup>39</sup>[POVINELLI 2000, pp. 108–31].



**Fig. 16.4** A table is divided by a rail into two identical parts. (a) Orienting experiment in which chimpanzees learn to use a rake for obtaining food. (b) Here, there is a rake on each section of the table, one with and the other without a food reward. (c) Finally, both rakes have a reward, but on the left part of the table there is a hole in which the food can remain trapped, while on the right part the hole is only painted. Can the chimpanzee distinguish between these two situations and act accordingly?

to consider the problem in the most abstract and general terms, which demands a symbolic knowledge. This assumption seems to be confirmed by a series of experiments with another apparatus [Fig. 16.4]. A table is divided by a rail that separates two sections of identical area. In the orienting experiment, a chimpanzee showed its ability to pull a rake in order to obtain a reward. In a second run, for testing, there were two rakes (each one in a separate section), one with a reward, the other without. Finally, in the third series, both rakes had a reward, but, on one side the reward could fall into a hole, while on the other only a painted hole was present.<sup>40</sup> At least one chimpanzee succeeded consistently, so that it seems that chimpanzees are somehow able to understand objective causal relations. The experimenters have rightly pointed out that here the context is much more understandable for the chimpanzees than in the previous series of experiments, since both options are *simultaneously present*. Now, as I shall stress at the beginning of the next section, categorical reasoning supposes relations between objects and situations that are *immediately experienceable and perceivable*. Povinelli seemed to come to similar conclusions when analyzing other experiments on chimpanzee's understanding of causality in which the shape of the tool played a central role,<sup>41</sup> and his general conclusions are along the same line.<sup>42</sup>

## 16.5 Categories?

### 16.5.1 Relations and Categories

Categories are different from schemata. The latter involve perceptual features, categories allows to classify objects and display perceptual (i.e. observable) *relations* among objects that schemata

<sup>40</sup>[POVINELLI 2000, pp. 132–48]. <sup>41</sup>[POVINELLI 2000, pp. 204–5].

<sup>42</sup>[POVINELLI 2000, p. 298]. The only concern here is that Povinelli makes use of the term “concept” with a perceptual basis, instead of the term category, but I consider this to be merely a terminological matter.

**488** Learning

do not have, while concepts instantiate abstract class relations<sup>43</sup> [Subsec. 12.5.2]. This distinction is clearly supported by the fact that the so-called *delayed matching-to-sample* tasks are difficult for nonhuman primates (and cetaceans) which can successfully cope with a delay of at most few minutes.<sup>44</sup> Schemata are essential for the recognition of events that are meaningful to the species; categories for comparison between contexts and semiotic communication with conspecifics; concepts for classification, reasoning, and symbolic communication with conspecifics (but in principle with every symbolic species). It is also true that categories remain the basis of many concepts.

In the case of object categorizing, the true mark of a category is when it is acknowledged as equivalent to (as a model of) the objects it represents, in such a way that it is possible to understand true aspects and properties of an object by considering the category only, while a schema represents a very rough description of an object seen only from the perspective of a certain vital value that is species-specific relevant (affordance) [Subsec. 12.5.3]. Therefore, a category implies a certain interchangeability with the object of which it is a category. Animals generally have more difficulties in picture-object transfer than in object-picture transfer, since lower-level perceptual experiences as expressed in schemata suppose a signal going from the object to the animal [Subsec. 12.7.1]. However, only these two relations *together* build a true category, which consists, as said, in an equivalence between the representation and its referent: Recall that an equivalence is transitive (if A is a representation of item B and of item C, the properties that A attributes to B must also be possessed by C), symmetric (if A represents B, B could also be said to represent A), and reflexive (any item represents itself) [Subsec. 8.2.3].<sup>45</sup> In the following, I assume that apes but not monkeys (at least not in a general way, as the previous section shows) are able to reach a full categorical level, since monkeys rely more on feature sharing among objects than on perceptual equivalence.<sup>46</sup> Given such an equivalence, young children (like some primates) have a causal understanding of representations: A modification of the representation would produce a corresponding change in the represented object. Categories and only categories (not schemata) allow for understanding the functionality of certain objects, even if they are always connected with perceptual features [Sec. 4.5]. M. Jeannerod and P. Jacob seem very well aware of this when they say that the function and the categorization of an object stand as an interface between pure visual processing (schemata) and conceptual processing of information.<sup>47</sup>

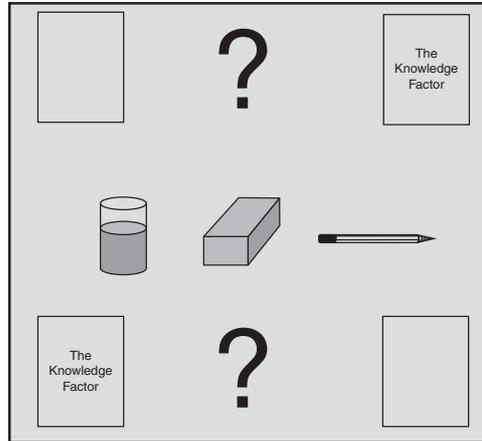
A proper causal action can be much better understood when the animal is provided with categories.<sup>48</sup> An *objective* causal relation can probably be understood only with categories—consider the problems analyzed in Sec. 16.4.<sup>49</sup> Indeed, any causal action can be represented as a sequence: An object in its initial state, some factor capable of changing the object, the object in its final state. Language-trained chimpanzees can fill the sequence, where either the terminal state of the object or the instrumental factor is not specified and each time a choice among different alternatives is presented. Another more difficult test is on the temporal order of cause and effect: In performed experiments, the particularly intelligent chimpanzee Sarah was requested to read a

<sup>43</sup>[KAGAN 2002, pp. 44–8].      <sup>44</sup>[REZNIKOVA 2007, p. 91].      <sup>45</sup>[AULETTA 2002].

<sup>46</sup>I therefore agree with the view presented in [THOMPSON/ODEN 2000]. The only difference is that the two authors speak here of the apes' understanding of equivalence as a conceptual level. The issue is probably terminological. For me, what is really crucial is that apes cope with perceptual situations while humans, making use of true concepts, also deal with nonperceptual equivalence. The authors seem to accept this difference towards the end of their paper. Everybody can agree that it is an issue that needs further clarification.

<sup>47</sup>[JACOB/JEANNEROD 2003, p. 78].      <sup>48</sup>[MICHOTTE 1963] [PREMACK/PREMACK 2003, pp. 159–76].

<sup>49</sup>Evidence for the fact that understanding of objective causality is more difficult than understanding of egocentric or social-induced causality can be found in the experiments cited in note 80 to the previous chapter.



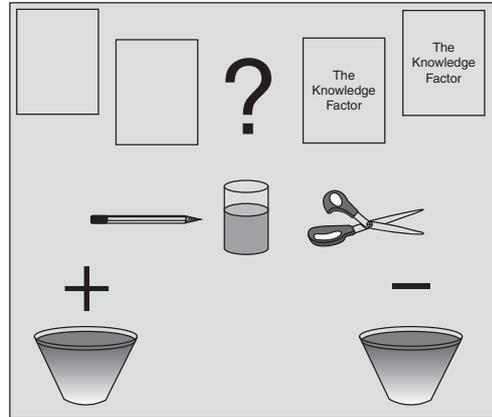
**Fig. 16.5** Premack's causality test for the reversal of one of the above actions. Below we have the sequence marked paper, blank, unmarked paper, and above unmarked paper, blank, marked paper, along with three alternatives: Container of water, pencil, eraser. In this way, the two sequences above and below show two different causal actions: Marking is shown on the top panel, erasing on the bottom panel. Adapted by [PREMACK/PREMACK 2003, p. 169].

sequence of three elements [Fig. 16.5] from left to right (marked paper, blank, unmarked paper) and from right to left (unmarked paper, blank, marked paper) along with three alternative tools (container of water, pencil, eraser): Water is irrelevant in both cases (it is a distractor), whereas eraser and pencil are relevant to the first and second sequence, respectively. Sarah was right on 40 of 60 runs, which is probabilistically very significant (66.6667%). Another experiment is shown in Fig. 16.6. Here, it is possible to act on paper in three different ways: Cutting, marking, or wetting. The objects actually used for a certain operation will be put in the bin on the left while the irrelevant ones go in the trash bin on the right. Let us consider a rather complex example: By observing a marked paper as input that ends as marked, wet, and cut as output, Sarah was able to put the scissors and container of water in the left bin and the pencil in the trash bin on the right. Therefore, the conclusion is that chimpanzees seem to be able understand *observable* and *objective* (i.e. not dependent on our actions) causal connections.

### 16.5.2 Metaphor and Metonymy

As I have stressed, primates, especially apes (and cetaceans) have categories because they are able to understand not only the consequences of their own actions but also of *others' actions* [Subsec. 12.7.3 and Sec. 15.3]. This means that apes (and cetaceans) are able to establish a *comparison* between their own behavior and the behavior of others. This is why categories are intrinsically relational. In particular, they are born and enlarged through a double process [Sec. 4.4.4]. Recall that when a new event is perceived, it can violate less essential features of a previous category (and it is a *discrepant event*) or alter essential features (and it is a true *novelty*) [Secs. 12.2 and 16.1, Subsec. 12.5.1]:

- In the first case, we enlarge the previous schema by a *metonymic* process (by weakening the constraint).



**Fig. 16.6** Premack's causality test for multiple transformations: Cutting, wetting, and marking, with separate receptacles (shown bottom) for relevant (on the left) and irrelevant (on the right) operators. Here, the relevant operation is marking (pencil involved). Adapted by [PREMACK/PREMACK 2003, p. 170].

- In the second case, we split a previous category or schema into two parts, giving rise to a new category defined through new features (it is a *metaphoric* process).

Let us now consider the concept of metaphors. *Metaphors* import the relational structure (and not the surface features) from the base domain to the target domain. Metaphors (at a purely abstract level of analysis) therefore have a vehicle, a topic, and a ground.<sup>50</sup> Any metaphor takes an essential feature of the vehicle for expressing a less essential feature of the topic. The feature is the common ground. For instance, let us consider the linguistic metaphor “Words are weapons.” Here *word* is the topic and *weapon* the vehicle. The feature is the *capability of doing damage*. Metaphors must have a rich schematic and categorical component in order to work. Summing up, through metaphor we consider a certain new stimulus as not being completely in the same category as a previous one. It is a fact that uncommon events sharing essential features of a category, but not sharing less essential ones, recruit more attention.<sup>51</sup> In this way we are forced to split a previous category in order to allow for the possibility of comprehending both stimuli. The new category is defined by a property or a feature that was unknown before, and metaphors can provide a bridge between the two.

Instead, when categories are enlarged by *metonymic* procedures, a part or an aspect stays for a whole or there is some relation of contiguity [Secs. 4.1 and 15.1]. In both cases, of metaphor and metonymy, the existence of previous categories is presupposed, which again confirms that any treatment of information consists in selection and not in true creation [Subsec. 2.2.3].

### 16.5.3 Metaphor and Metonymy in Apes

In the 1960s Piaget's developmental psychology methods (causal reasoning, logical-mathematical reasoning) were applied to animals and especially to great apes and human evolution.<sup>52</sup> Very

<sup>50</sup>[BLACK 1962] [MCNEILL 1992, p. 146]. <sup>51</sup>[KAGAN 2002, p. 117].

<sup>52</sup>[PARKER/GIBSON 1977, PARKER/GIBSON 1990] [DORÉ/DUMAS 1987] [RUSSON/BARD 1996]. For an analysis of the problems in this field see [ALLEN/HAUSER 1991] [VAUCLAIR 1996, pp. 29–52].

MODEL	MBUS	LBUS
TRANSFER	MYUT	LYUT
DISTRACTOR		LYMS

**Fig. 16.7** The transfer (equivalence) experiment performed by Premack. The ape must understand that the relation between MYUT and LYUT is the same as that between MBUS and LBUS.

important experiments were carried out many years ago. As is well known, one of the most striking abilities of a human child is analogical reasoning.<sup>53</sup> The chimpanzee Sarah was subjected to several experiments of analogical (metaphorical) reasoning (transfer from items A–A' to items B–B'), and solved them correctly.<sup>54</sup> In one of them the A series was of the type MBUS, the A' series printed words of the type LBUS, the B series the type MYUT, the B' series type LYUT, and the distractor C' series type LYMS. Having been trained to pair A–A' (MBUS–LBUS), Sarah correctly paired B' to B (MYUT–LYUT), discarding C' (LYMS) [Fig. 16.7].

A chimp can also learn to match proportions (half an apple to a half-glass of water) and to add proportions. For instance, Premack and Woodruff tested an adult and four juvenile chimpanzees about their knowledge of proportion and number with conceptual matching-to-sample tasks. The juveniles failed but the adult successfully matched exemplars of the proportions 1/4, 1/2, 3/4, and 1 and the numbers 1, 2, 3, and 4, when the sample and the options were highly dissimilar in shape and color, mass, area, length.<sup>55</sup>

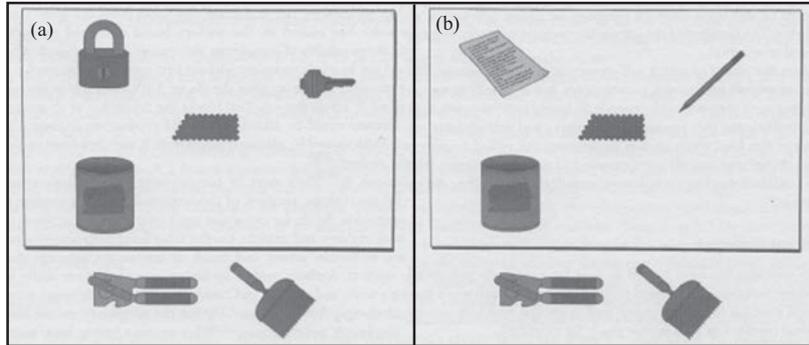
Language-trained animals only show a main difference relative to untrained animals: The perception of same/different. This may depend on the use of proper names (different relations of being alike or unlike). It is interesting to observe that human children are highly conservative: They assume that the object they are looking at now is the same that they saw earlier if there is no violation of the spatial-temporal relations<sup>56</sup> [Subsec. 12.5.2]. An important test of apes' ability to categorize and eventually to think abstractly was performed. The experiment is a comparison between language-trained animals (three African-born chimpanzees) and nonlanguage-trained animals (four African-born chimpanzees) in solving problems.<sup>57</sup> In the language training the animals were given a string of elements representing simple propositions and the goal was to fill an incomplete slot (it is a method for interrogation where a single word suffices as an answer). The animal was requested to identify the relation that a pair of arguments instantiated, to fill the slot of an argument or to answer a yes/no question. The effect of training seems permanent since the test occurred six years after the language training ended. The tests that discriminated the two groups of animals is about the *relation between relations* (as in analogies), whereas the test which could not discriminate them was about spatial location of items and inferences one might make to obtain an item. The second test is concrete, because spatial relations can be perceptually imagined. Instead,

<sup>53</sup>[LADYGINA-KOHTS 1935, pp. 295–316]. <sup>54</sup>[GILLAN *et al.* 1981].

<sup>55</sup>[WOODRUFF/PREMAC 1981]. Experiments with pigeons, for instance about the “category” of insiderness [HERRNSTEIN *et al.* 1989], do not seem to capture the more fundamental meaning of the term *category* involved here but rather seem to be concerned with pure perceptual relations or feature sharing. Studies on parrots about the “category” same/different [PEPPERBERG 1999] also seem to fall into the same basket. See also [REZNIKOVA 2007, pp. 152–66] for a review.

<sup>56</sup>[PREMAC/PREMAC 2003, pp. 177–226]. <sup>57</sup>[PREMAC 1983a].

## 492 Learning



**Fig. 16.8** Premack's test of the chimp Sarah consists of understanding the relations between: (a) A lock is to a key as a can is to a can opener. (b) Marked paper is to a pencil as a paint can is to a brush. To pass such a test implies the ability both to understand the functionality of objects and establish analogies between relations among different groups of objects. Both skills are based on the ability to build categories. Adapted from [PREMACK/PREMACK 2003, p. 192].

the first test is abstract, because an abstract relation can be instantiated in indefinitely many cases and here the invariance common to the individual cases cannot be represented in a schema-code. The same holds for judgments about the membership in functional categories. As an example of such analogy testing, consider Fig. 16.8.<sup>58</sup> Sarah, a language-trained chimpanzee, in this test was right on all 12 trials, while the nonlanguage-trained animals responded by chance, showing no progress. This is an evidence that chimpanzees are at the threshold of symbolic communication and understanding: They can judge the relationship “can opener: can” to be the same as the relationship “key: lock” and a cylinder 1/4 full to be the same as 1/4 apple, but different from 3/4 apple. Indeed, to understand relations between relations is the highest form of categorization, because more than just immediate perceptual aspects that are involved.<sup>59</sup>

Now, language training affects the concept of *sameness*. Since there are ambiguous situations where sameness can be confused with physical similarity, in order to test for sameness one needs to present two items that either are or are not the same, and to use an arbitrary item like a piece of plastic, meaning the words “same” in the one case and another meaning “different” in the other case. The ability of language-trained chimps to understand abstract relations was confirmed in an experiment in which the primates must make use of items instantiating the category of identity and nonidentity as such<sup>60</sup> [Fig. 16.9].

Metonymies are less studied. Evidence of their use by apes comes from Povinelli's studies on chimpanzees' ability to discriminate, at least in certain contexts, between pure contact and physical connection between objects.<sup>61</sup> This is an important metonymic category. Indeed, Povinelli has proposed as one of the basic “principles” of the chimpanzees' folk physics<sup>62</sup>: Contact is necessary and sufficient to establish covariation during an object's motion [Subsec. 5.1.3]. This is the reason why chimpanzees are not fooled by superficial changes of the task, like color, size, and so on [Fig. 16.4(c)].

<sup>58</sup>[PREMACK/PREMACK 2003].

<sup>59</sup>I again essentially agree on this point with [THOMPSON/ODEN 2000], although there are some terminological differences.

<sup>60</sup>[THOMPSON *et al.* 1997].

<sup>61</sup>[POVINELLI 2000, pp. 252–3 and 269].

<sup>62</sup>[POVINELLI 2000, pp. 305–8].

<b>AA</b>	<b>EF</b>	( I )	( NI )
<b>BB</b> <b>CD</b>	<b>BB</b> <b>CD</b>	( I )      ( NI )	( I )      ( NI )
<b>PHYSICAL STIMULI</b>		<b>CATEGORIES</b>	

**Fig. 16.9** Experiment performed by making use first of the identity (AA, BB) and nonidentity (CD, EF) relation between physical items represented by the two sets of letters on the left, and then by making use of other items instantiating the abstract relation of identity (I) and nonidentity (NI) (the two sets on the right). Adapted from [THOMPSON/ODEN 2000].

### 16.6 Concluding Remarks

Learning is different from behavior (although the relation is often confused) because it is grounded in the cooperation between the decisional and the sensory–cognitive systems requiring the involvement of attentional processes and of some form of awareness. Indeed we do not have learning without some kind of result of general validity:

- It is possible that most premammals essentially rely on conditioned learning, in which learning and memory or learning and behavior are not yet fully distinguished.
- A very elementary form of learning is represented by pure associative learning.
- Mammals are able to understand causal relations since they are accustomed to dealing with the effect of their own behavior (addressing, i.e. the second-order representational system).
- The ability of primates to learn is far more advanced (third-order representational system). In this case, primates and especially apes rely on categories which are grounded on perceptual relations. Here, metonymic and metaphoric extensions and relations are established.

# 17

## Memory

---

In this chapter, we shall examine the third fundamental functionality of the brain: Memory. After a short presentation of the nature of memory and of its fundamental features, I shall introduce the relevant neural mechanisms. This will allow us to deal with the fundamental difference between procedural and declarative memory, which turns out to be connected with two different functions, namely learning and memory. Finally, we shall discuss the problem of the stages of memory consolidation, starting with the traditional distinction between a short-term and a long-term memory. We shall see that it is convenient to establish a dichotomy between active and inactive memory.

### 17.1 What Memory Is

The memory function is taken to be characterized by three features<sup>1</sup>: Acquisition, consolidation, and retrieval. Retrieval is probably not sufficiently general and does not take into account emotionally charged memories that are generally recurrent without explicit retrieval. It would then be better to use the concept of memory *activation* [Fig. 17.1].

Memory is a special case of neural plasticity. It has been hypothesized that memory depends on a protein and therefore on the genetic code. But this seems unfounded. As we have seen, Hebb hypothesized a dual trace mechanism according to which the transient reverberatory trace may be reinforced by structural change, the so-called Hebbian rule<sup>2</sup> [Subsec. 3.8.2]. These changes are modifications in the synaptic connections and probably in the morphology as well as in the chemical composition of synapses. In other words, growth processes accompanying synaptic activity make the synapse more readily traversed. When two cells that were repeatedly active at the same time become associated, new structures will be built. This is also the associative basis of any mental activity.

Learning and memory are closely related: Learning is the process of acquiring operative patterns [Sec. 16.1] and it is often claimed that memory is the persistence of learning.<sup>3</sup> However, memory is also concerned with events that simply *happen* to an organism and in which therefore there is no learning involved, especially when those events are heavily emotionally charged. The connection between the emotional system and memory is particularly evident for dramatic events (the so-called flashbulb memory): For instance, recalling what one was doing when the Twin Towers collapsed.

<sup>1</sup>[SHERRY/SCHACTER 1987].

<sup>2</sup>[HEBB 1949, pp. 59–78].

<sup>3</sup>[SQUIRE 1987].

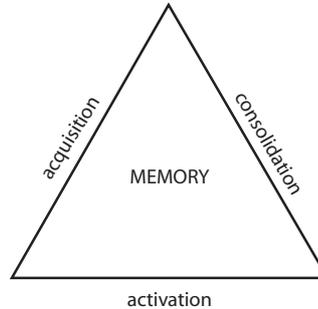


Fig. 17.1 Memory function as characterized by acquisition, consolidation, and activation.

Therefore, the purposive aspect that is typical of learning here fails: Memory, consisting of associations,<sup>4</sup> finds its specific ground in *habit* formation<sup>5</sup> [Subsec. 12.5.1], while sensitization does not play the central role that it has in learning [Sec. 16.3]. However, in accordance with what we have seen in the case of learning (pure repetition does not lead to results [Sec. 16.2]), memories are not organized along a single dimension only according to their strength.<sup>6</sup> Tulving showed that, in certain experimental contexts, memories wane and come back on successive trials. Given this proviso, it remains true that any organism, seen from the outside, is a bundle of habits.<sup>7</sup> In this context, I also recall Thorndike's law of exercise: Any response to a situation, other things being equal, will be more strongly connected with the situation in proportion to the number of times it has been connected with that situation and to the average vigor and duration of the connections<sup>8</sup> [Subsec. 15.1.].

It is possible to assume that habit already plays a role at a physical level (for instance, water follows existing channels, a fracture in some material follows paths that have been weakened). It can even play an important role in the initial differentiation of tissues and structures, both at an epigenetic and phylogenetic level.<sup>9</sup> However, habit is very important *for ontogeny* as far as it represents the necessary background for simplifying the procedures for obtaining a result, making them more accurate, thus diminishing energy expenditure as it implies a decrease in active attention: Automatization converts *previously learned* motor processes into routines that can be executed for specific tasks as wholes independently of other processes and at a lower cost. Let me also distinguish between habituation and automatization, at least for mammals: Automatization is driving a car; habituation is me driving a car to my office. Thus we should avoid thinking of habit as a sort of ready-made behavior concerned with congenital instinct<sup>10</sup> [Subsec. 15.2.3]. Another issue is whether habit can somehow be inherited,<sup>11</sup> a result that could be possible in the light of the Baldwin effect [Sec. 9.10]. In this case, even in the absence of a Lamarckian inheritance of

<sup>4</sup>[ROMANES 1884, pp. 37–8].

<sup>5</sup>[BICHAT 1800, pp. 37–46]. This important French tradition focusing on habit investigation has been developed in de Biran's work from the perspective of the influx of habit on cognition [DE BIRAN 1803].

<sup>6</sup>[TULVING 1976] [WILLINGHAM 2001, pp. 204–206].

<sup>7</sup>An honorable tradition has pointed out this aspect [LAMARCK 1809, I, p. 237] [JAMES 1890, v. I, p. 104–127].

<sup>8</sup>[THORNDIKE 1898, p. 244].

<sup>9</sup>[SPENCER 1864–67, v. II, pp. 339–55]. This is the way in which both Lamarck [LAMARCK 1809, v. II, p.181] and Spencer [SPENCER 1864–67, v. II, pp. 355–61] explained the formation of a nervous system. Recall, however, that without an element of novelty (for instance, chance) habit alone could not produce such a differentiation.

<sup>10</sup>[LLOYD MORGAN 1896, pp. 17 and 142–43].

<sup>11</sup>[LLOYD MORGAN 1891, p. 434].

## 496 Memory

habits, we could say that habits contribute to the canalization of the action of natural selection and therefore can be considered to be the nursery of instincts<sup>12</sup>: It has been shown<sup>13</sup> that when an action is automatized, reflex-like circuits are formed in the motor cortex, so that habit is the first stage toward automatization.<sup>14</sup>

Habit therefore is also the basis of any modularization [Sec. 3.6 and Subsec. 11.5.1]: In any succession of modular operations one can find the appropriate successor without the risk of taking a false alternative. Here, the action occurs not as a consequence of a thought or of a decision but of a sensation occasioned by some previous action. Indeed, in recalling some past events, there is always a prevailing element that captures the organism's interest.<sup>15</sup>

We may then define *memory* as the purposeless and emotionally charged process of the brain that through habituation determines the background of learning and behavior in terms of a reservoir of experiences and ontogenetic accommodations.

## 17.2 Past and Present

The new approach to the physiology of learning and memory in animals began with a study by Kandel and Spencer in 1968.<sup>16</sup> The collection of neural changes representing memory is called an *engram* [Subsecs. 5.3.2 and 14.1.3].<sup>17</sup> *Memory acquisition* is the process that converts an experienced event into an engram, but engrams are specified in terms both of their antecedent conditions (past memories) and present conditions (*memory activation* in a certain context). Therefore, one of the most distinctive features of engrams is their mutability. The activation of an original engram is governed by the similarity of interpolated events to the original event.

Memory processes are very ancient from an evolutionary point of view. For example the fly *Drosophila* shows the ability to discriminate and memorize parts of its environment independently of the retinal position during acquisition of the pattern (translation invariance).<sup>18</sup> This example shows that *previous* experience is always generic [Subsec. 12.5.2]. Only *actual* experience, especially when attentional processes are involved, is specific [Subsec. 14.1.1]; or better yet, the process of our actual experience is concerned with *individual sensation* [Sec. 12.2]. In this sense, actual things and events are always ahead of us, and we approach them with certain expectations, which represent the future [Sec. 14.3]. However, since we also (through antifeedback) try to efface the effects of environmental stimuli (which in principle are negative feedbacks for us), we will never completely preserve a trace of the event in its uniqueness. This is the basis of the *genercity* of our memories.<sup>19</sup>

Let us consider this mechanism. As I have stressed, one acquires memories when the object possesses some relevance for the organism<sup>20</sup> [Sec. 17.1]. By remembering, one makes use of schemata or categories (or even concepts) which are *active organizations* of past reactions. A new act is not something completely new, or something completely old, but rather stems from new and original interpolations of already-existing schemata or categories. The past determines the present but not massively, rather it is itself dominated by a particular present event or feature, by an occurring, actual spontaneous selection or decision [Subsec. 14.1.2]. Then, we build a scene afresh on the basis of the memory, and what dominates is the *current attitude* (emotional state or interest), and recall (or activation) represents rather a justification or a help for the current action or state. Summing

<sup>12</sup>[LLOYD MORGAN 1900, pp. 177–78].

<sup>13</sup>[EVARTS 1973] [BADDELEY 1990, pp. 86–95].

<sup>14</sup>[WILLINGHAM 2001, pp. 115–19].

<sup>15</sup>[JAMES 1890, v. I, p. 571–90].

<sup>16</sup>[KANDEL/SPENCER 1968].

<sup>17</sup>[LASHLEY 1950].

<sup>18</sup>[LIU *et al.* 2006].

<sup>19</sup>[DE BIRAN 1803, pp. 108–9].

<sup>20</sup>[BARTLETT 1932].

up, memories are constructions assembled *at the time of activation*. The remembered present is the application of potential information [Subsec. 2.2.2] coming from the past into a present whose meaning or result will emerge in the future (according to whether our expectations will be fulfilled or not).

Indeed, as we have seen, representations are not passive projections onto a mental screen but are self-organizing and dynamic constructions that process, share, and select information [Secs. 12.1–12.3]. Memory is an emergent property of the collective behavior of systems of neurons organized transiently (or more permanently) in functional networks of preferential coupling.<sup>21</sup> Only the transient organization, which is the support of an active memory, is the actual support of information [Sec. 12.4]. Static or inactive memory, leading to dormant or potential engrams, is characterized by the absence of specific activity in the network and therefore it cannot be taken as the support of information *per se*. It only provides the potential for reactivating the memory trace whenever a subset of the original information or related information is available. It is interesting to recall here that computationalism and cognitivism [Secs. 3.2, 3.5, and 6.1] instead maintain the idea that memory is a sort of library in which the item should be statically classified and cumulated.<sup>22</sup>

For this reason, unused areas of the cortex become occupied by other closely located functions. This is not simply a consequence of disuse, but means that there is a continuous *reorganization* of the cortex [Subsecs. 3.3.2 and 3.4.3], and that there is a real forgetting (even if there are always traces of the ancient functions and memories).<sup>23</sup> Since memory and representation are redundant and since memory involves, at least in higher animals, the representation of many features of an event, forgetting is an issue of *grades* (not all connections are lost). Moreover, due to the resculpturing of previous memories when new memories are added, forgetting occurs continuously, but so does strengthening of the memories that are preserved.<sup>24</sup> Memory depends on factors such as the alertness level at the time of memory acquisition and the nature of the events occurring just after information has been registered. If the consequences of an experience are trivial, the experience is quickly forgotten, while memories that show themselves to be relevant on many occasions are more and more consolidated.

In memory activation there are different levels to consider: Neurochemical changes, neural activity, and psychological experience, at least in the case of mammals.<sup>25</sup> Successful remembering is a combination of trace information (reflecting variables involved in memory acquisition) and cue information (reflecting activation variables). These cues have a true semiotic structure since they represent the signs (actually, the marks) referring to previous memories. Activation cues will be effective to the extent that information in the cues was incorporated in the trace event at the time of its original encoding (this is called the *encoding specificity principle*). It is also true that, at least in humans, memory is content-addressable, i.e. one can access information in the memory based on almost any attribute of the representation one may desire to retrieve<sup>26</sup> [Subsec. 12.4.3]. Since this activation generally deals with actions to be performed (behavior) and with learning activity, decisional and selective processes are very important, and for this reason memory retrieval in particular, being an active search process, is concerned with the future<sup>27</sup>: For *retrieval* to occur the brain must be in the retrieval mode: It is by definition an attentional top-down activity. Evidence of this is that recognition memory functions serially and not as a parallel network.<sup>28</sup>

<sup>21</sup>[LAROUCHE *et al.* 1995]. See also [MCCLELLAND *et al.* 1995]. <sup>22</sup>[SIMON 1969, p. 88]. <sup>23</sup>[GROSS 2000].

<sup>24</sup>[DE BIRAN 1803, p. 14]. <sup>25</sup>[CRAIK/LOCKHART 1972, LOCKHART/CRAIK 1990] [BROWN/CRAIK 2000].

<sup>26</sup>[MCCLELLAND *et al.* 1986]. <sup>27</sup>[DUDAI/CARRUTHERS 2005]. <sup>28</sup>[GAZZANIGA *et al.* 1998, pp. 99–100].

## 498 Memory

This also implies that memory acquisition and its activation can be separated. This has been proved for the *Drosophila*.<sup>29</sup>

### 17.3 Neural Substrate of Memory

Memory is stored in terms of changes in the same neural systems that originally participated in the perception and in the processing of the information that has been acquired<sup>30</sup> [Sec. 13.2]. Therefore,

- In its *access to the information source*, memory is both *localized*, since particular brain systems represent specific aspects of each event, and *distributed*, since many neural systems participate in reenacting a whole event.<sup>31</sup>
- In its *specific operational modality*, memory is again *localized*, since there are very specific systems participating with memory, and it is also *distributed*, since any sensory and even motor area of the brain can contribute to memory.

All neocortical sensory systems (the areas of neocortex where the different sensory information is conveyed) converge into the rhinal cortical areas known as the parahippocampal region, which integrates information before delivering it to the hippocampus<sup>32</sup> [Fig. 17.2]. The hippocampus and the rhinal cortex together constitute the medial temporal lobe (MTL) memory system. In this way rhinal areas are able to build representations that finally become independent from the original modality (whether visual, auditory, or somatosensory) of information-processing. The medial temporal regions do their job only during the consolidation time<sup>33</sup> [Fig. 17.1]. As a result, sights, sounds, and smells can constitute a global memory of the situation contributing to a *single* engram. In this way, categories and concepts beyond single perceptions also become possible [Secs. 12.5 and 13.2]. It is interesting to stress that, while primates' neocortex has several cortical convergence zones, the neocortex of old mammals shows few of such areas, which again is evidence that we cannot speak of true categories in lower mammals (and in birds).

The *hippocampus* seems needed for storage processes. The hippocampus and dentate gyrus are metabolically activated by the so-called working-memory task.<sup>34</sup> But a functional disruption of hippocampal activity can also affect the so-called long-term memory. The reason is that the hippocampus also participates, in the first stages, in the reinstatement of the pattern of cortical activation that occurred during the original experience.<sup>35</sup> However, each reinstatement slightly changes cortical synapses (a sort of inductive tuning and shifting), according to the dynamic understanding of memory presented in the previous section. Old memories are the result of the accumulation of synaptic changes in the cortex after many reinstatements. Eventually, the cortical representation comes to be self-sufficient and independent from the hippocampus<sup>36</sup>: The neocortex is indeed the place where long-term memories are “stored” (it is again a highly distributed system, depending on the mode of the memory and on its connections with other memories: This runs against a strict modularization).

The traditional understanding of the hippocampus' role in memory is founded on the distinction between an anterograde amnesia (for new events) and a retrograde amnesia (for past events). As I have stressed, memory is not fixed at the moment of acquisition but continues to stabilize or

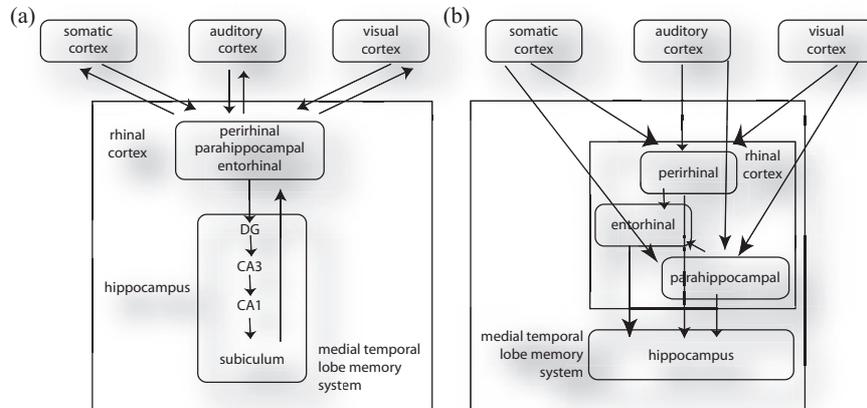
<sup>29</sup>[DUBNAU *et al.* 2001]. <sup>30</sup>[SQUIRE 1986].

<sup>31</sup>A compromise suggested by K. Lashley [REZNIKOVA 2007, p. 85].

<sup>32</sup>LEDOUX 2002, pp. 103–7]. <sup>33</sup>[ZOLA-M./SQUIRE 1990].

<sup>34</sup>[FRIEDMAN/ GOLDMAN-R. 1988][TRAUB/MILES 1991].

<sup>35</sup>See also [LOUIE/WILSON 2001]. <sup>36</sup>See also [NIELSEN/STENSTROM 2005].



**Fig. 17.2** Hippocampus and sensory systems. (a) The medial temporal lobe memory system showing the feedback circuits between the rhinal cortex and sensory systems, as well as between the hippocampus and rhinal cortex [see also Fig. 13.7]. (b) A detailed view of the rhinal cortex in its function of giving information from the sensory systems to the hippocampus. Inspired by [LEDOUX 2002, pp. 104–5].

consolidate with the passage of time, so that the initial acquisition of information is followed by two parallel processes: Gradual *forgetting* and gradually developing *resistance to disruption* of what remains, in a process of reorganization. Memory is also affected by rehearsal and by subsequent memory storage of episodes which may remodel the neural circuitry underlying the original representation. In this sense, there is always accommodation to new information. But, also for the same reason, *retrograde amnesia* cannot affect memories that have already been well consolidated. The role of the hippocampal formation in acquiring new information and in maintaining recently acquired information in memory until the consolidation process has ended<sup>37</sup> can explain the difference between the anterograde form of amnesia and the retrograde variety: The process of retrograde amnesia seems to occur because the role of the hippocampus in memory consolidation is still relevant.<sup>38</sup> A helpful model has been developed by McClelland and coworkers<sup>39</sup>: A connectionist net works better when new information is gradually incorporated into it—so-called interleaved learning—preventing new information from interfering with old memories [Sec. 12.4]. With interleaved learning, new information is slowly assimilated over many repetitions (nondisruptive induction), finally producing a new representation that it is well integrated with previous knowledge.

Recently, the role of hippocampus as a general relay station for the very first steps of memory consolidation has been partly redimensioned. Experiments on nonhuman primates show that rather than operating as a single functional unit, components of the medial temporal lobe make independent and dissociable contributions to object recognition and associative memory, sometimes in the absence of the hippocampus.<sup>40</sup> The perirhinal cortex plays a particularly central role in visual recognition (probably in the recognition of types). The rhinal cortex plays a central role in stimulus–stimulus association [Subsec. 15.1.4] and, according to Murray, is therefore critical for the formation of new memories involving objects, though, it is not necessarily the site of storage of these

<sup>37</sup>[BONTEMPI *et al.* 1999].

<sup>38</sup>[LEDOUX 2002, pp. 105–6].

<sup>39</sup>[MCCLELLAND *et al.* 1995].

<sup>40</sup>[MURRAY 2000].

**500** Memory

memories. In fact, as mentioned, stimulus representations are widely distributed throughout the neocortex, and the projections from the entorhinal and perirhinal cortex (Brodman's areas 28, 35, and 36), which are part of the ventromedial temporal cortex, back to modality-specific neocortical fields may provide a mechanism enabling this distributed long-term storage. The contribution of the rhinal cortex may be crucial for object identification: A part of the process through which different stimuli coming from different sensory modalities and are linked together, and through which environmental stimuli are invested of biological meaning. It is also true that experiments on macaques show that there is also no evidence for a separate class of perirhinal neurons involved in memory processes.<sup>41</sup> Rather, the mnemonic properties of perirhinal neurons exist simultaneously with their stimulus-selective properties.

This does not mean that the hippocampus is irrelevant. There are even aspects of memory consolidation in which it is very relevant. The hippocampus is probably crucial for spatial mapping and learning how to move to particular locations.<sup>42</sup> In particular the hippocampus may be properly responsible for sketch maps while the hippocampal dentate gyrus provides bearing maps<sup>43</sup> [Fig. 12.16]. On the contrary, the rhinal cortex does not seem to be involved in an egocentric frame of reference [Subsec. 5.3.3]. The hippocampus may also be important for event memory. Eichenbaum's results<sup>44</sup> suggest that the hippocampus maintains a local organization with respect to place fields, despite having no apparent large-scale isomorphism with the spatial environment. The organization of multiple, clustered place fields with correlated movement-tuning properties in small neural ensembles suggests the existence of functional neural ensembles serving to encode multiple sensory and behavioral aspects of a place or event. Such an organization is similar to that observed for neocortical association areas afferent to the hippocampal system.

In purposive behavior there is a multimodal sensory convergence, and the sensory input is then integrated over time and located in space.<sup>45</sup> The hippocampus provides a mechanism of attention as far as it does not select an input coming from the sensory cortex but arranges the output by biasing the sensory cortex so as to move in an appropriate direction for pattern construction, if the input contains the receptor activity that the command has predicted. I recall that [Subsec. 14.1.1], while in implicit memory storage the signal is reflexively recruited in a bottom-up way, in spatial memory dopamine is purposively recruited in a top-down way<sup>46</sup>: The cerebral cortex activates the cells that release dopamine, and dopamine modulates activity in the hippocampus.

**17.4 Experiential and Procedural Memory****17.4.1 The Meaning of Procedural Memory**

It is conventional to distinguish between a procedural and a declarative memory.<sup>47</sup> *Procedural* memory is a collection of abilities, is inflexible and bound to the learning situation, while *declarative* memory is about single events or situations and is flexible.<sup>48</sup> The former should be phylogenetically more ancient, implicit, and accessible to the same sensory modality in which material was presented

<sup>41</sup>[ERICKSON *et al.* 2000].<sup>42</sup>This has been crucially confirmed by recent studies on the hippocampus's growth in London taxi drivers [MAGUIRE *et al.* 2006].<sup>43</sup>[JACOBS/SCHENK 2003]. <sup>44</sup>[EICHENBAUM *et al.* 1989]. <sup>45</sup>[FREEMAN 1995, pp. 75–9].<sup>46</sup>[KENTROS *et al.* 2004] [KANDEL 2006, pp. 314–16].<sup>47</sup>[WARRINGTON/WEISKRANTZ 1968, WARRINGTON/WEISKRANTZ 1970].<sup>48</sup>[SQUIRE/KNOWLTON 2000].

initially. Declarative memory is more recent (it is typically present in mammals), elastic, and accessible from all perceptive modalities. Some scholars believe that it is present only in humans, an issue I shall deal below with. The declarative form enables integration of disparate, though relevant, items of information, while the procedural form allows for the translation of knowledge into action. In declarative memory, related memories are automatically activated,<sup>49</sup> which is not the case for the selective so-called procedural memory. It is a matter of fact that there is a double dissociation between declarative and procedural memory.<sup>50</sup>

The hippocampus is predominantly implicated in declarative memory and not in the procedural one.<sup>51</sup> There is therefore some evidence that the hippocampal system has two main functions: (1) The temporary maintenance of memories (reference to certain events and contexts), and (2) the processing of a particular modality of memory representation.<sup>52</sup> In the first case it is able to retain isolated terms at full strength for at least some minutes. During this time the second system begins to work and relations are fixed. The combination of these two aspects builds a declarative memory. As we have seen, the hippocampus receives inputs from and sends inputs to polysensory association areas of the cortex. The projections are combined but not completely superposed in order to allow for further processing.<sup>53</sup> However, in the intermediate time, it is especially the parahippocampal region that supports the formation of relations, as seen in the previous section.

Mishkin and Squire and coworkers<sup>54</sup> proved for the first time that the amygdala [Subsec. 14.2.2] lesion has no effect on specific learning tasks, such the so-called delayed nonmatching to sample, in which an animal must pick an item that was not shown among other items presented previously. This is evidence that learning can be dissociated from emotional subsystem too<sup>55</sup> [Fig. 13.9]. Since, on the other hand, the latter subsystem plays a major role in memory [Sec. 17.1], this shows that memory and learning can also be dissociated. In particular, declarative memory can be dissociated from learning,<sup>56</sup> while learning and procedural memory seem strictly related.

The previous examination therefore suggests a possible and rather radical solution<sup>57</sup>:

- The so-called *procedural memory* is actually part of the learning-processes circuit.<sup>58</sup> In other words, learning consists essentially in learning new *skills*, while memory is always memory of—more or less emotionally charged—*events* [Fig. 17.3]. The source of all complications here is that many examinations and discussions about this matter take *human* memory into account from the start (which has also been a necessity for a previous stage of neuropsychological studies). Now, human memory is *symbolic*, and symbols, as we shall see, have a strong *procedural* component that signs—and representations in particular—lack. Moreover, symbols are explicitly *learned* in a way in which categories are not, while schemata cannot be learned at all. Indeed, Warrington, Shallice, and McCarthy<sup>59</sup> showed that amnesia of terms referring to manipulable objects can mean an impairment of the motor system while amnesia of food names has more to do with problems with information storing [Subsec. 4.5.3].

<sup>49</sup>[COHEN/EICHENBAUM 1993, p. 62].

<sup>50</sup>[GABRIELI *et al.* 1995] [MARKOWITSCH 2000] [GAZZANIGA *et al.* 1998, pp. 332–3].

<sup>51</sup>[ZOLA-M. *et al.* 1986]. See [COHEN/EICHENBAUM 1993] for a comprehensive summary.

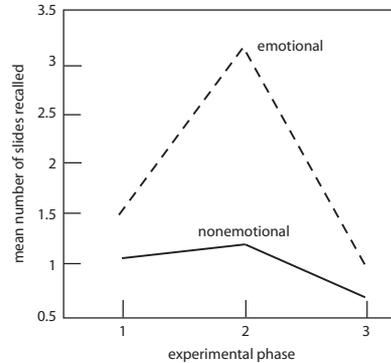
<sup>52</sup>[EICHENBAUM *et al.* 1994]. <sup>53</sup>[COHEN/EICHENBAUM 1993, pp. 94–108].

<sup>54</sup>[MISHKIN 1978] [ZOLA-M./SQUIRE 1984] [ZOLA-M. *et al.* 1989] [BENZING/SQUIRE 1989]. See also [LEDOUX 1998 pp. 184–98].

<sup>55</sup>[ZOLA-M. *et al.* 1991] [ZOLA-M./SQUIRE 1993]. <sup>56</sup>[COHEN/SQUIRE 1980] [EICHENBAUM 2000].

<sup>57</sup>[MISHKIN 1982]. <sup>58</sup>This is at least implicitly supported in [COHEN *et al.* 1997].

## 502 Memory



**Fig. 17.3** Data from [CAHILL/MCGAUGH 1995], showing the incidence of emotional arousal on memory in humans. All participants saw the same set of slides (phase 1 of the experiment) but some heard a story about the contents of the slides that strongly stressed the emotional aspects (phase 2). These were recalled much better later on (until phase 3). Adapted from [WILLINGHAM 2001, p. 178].

- For this reason, the term *declarative memory* also seems to me to be a little bit misleading, since it covers
  - (a) Aspects that are common to all mammals, in particular their (emotionally charged) individual experiences [Subsec. 15.2.3]. Olton has pointed out, against Tulving, that animals also show a so-called episodic memory.<sup>60</sup>
  - (b) Aspects bond to factual (like historical notions) and even semantic (like biological notions) knowledge, which are typical of humans and can only be explained thanks to a symbolic activity.

Indeed, in humans, long-term so-called declarative memory can be divided into<sup>61</sup>

- (i) Memory of events, so-called *episodic memory* (about our personal life and the way we obtain information), which is also common among many animals, at least vertebrates,<sup>62</sup>
- (ii) *Semantic memory* (storage of general notions about facts), which is necessary for the use of language and can eventually be present only in very rudimentary form in other primates as a categorical knowledge [Sec. 16.5]. This aspect is again connected with learning.<sup>63</sup>

#### 17.4.2 Episodic and Semantic Memory

As we have seen, Tulving has introduced a difference between episodic and semantic memory in humans. Let us deal with this problem more carefully. According to Tulving, the main differences between them are<sup>64</sup>:

- In information (source: sensation/comprehension; units: events/concepts; organization: temporal/conceptual; reference: self/universe),

<sup>59</sup>[WARRINGTON/SHALLICE 1984] [WARRINGTON/MCCARTHY 1987]. <sup>60</sup>[OLTON 1984].

<sup>61</sup>[TULVING 1984]. <sup>62</sup>[REZNIKOVA 2007, p. 77].

<sup>63</sup>A study shows a clear shift from so-called episodic memory to semantic memory as a form of learning [CONWAY *et al.* 1997].

<sup>64</sup>[TULVING 1984] [TULVING/THOMSON 1973].

- In operations (registration: experiential/symbolic; temporal coding: present/absent; context dependence: more/less pronounced),
- In applications (education: irrelevant/relevant).

The two types of memory are functionally distinct: One can operate independently of the other, although not necessarily as efficiently as it could do with the support of the other system. In semantic memory, the durability of the mnemonic trace is a positive function of the depth of processing, i.e. of the degree of semantic involvement.<sup>65</sup> Deeper encodings take a longer time and are better recalled in subsequent tests. Memory performance is enhanced when the context is congruous with the processed item. Episodic memory is characterized by encoding specificity, i.e. the engram and the retrieval cue must match and complement each other for remembering to occur. Tulving introduced the notion of *ecphory* in order to denote the synergetic process that combines the episodic information from the engram and the semantic information from the cue. There can be recall without recognition because recall is referential while recognition is associative. They are different in two respects: (1) in recognition a copy of the to-be-remembered item is given, in recall it is not; (2) in recognition the subject has to make a familiarity judgment, while in recall the item has to be named.

A basic tenet of Tulving's theory<sup>66</sup> is that episodic recollection, infused with the *autonoetic awareness* of one's existence in subjective time, is closely related to other mental capabilities and achievements that are uniquely human. In two important respects human episodic memory resembles those classes of behavior, such as complex problem-solving, that are often classified as supervisory functions: It requires a higher level of control that can be adapted to situational demands, and it depends on widely distributed cortical and subcortical networks, the prefrontal cortex playing a central role. *Human* episodic memory, then, can be thought to bear a close family resemblance to some high-order mental achievements and capabilities (the frontal functions): The prefrontal cortex, in conjunction with its reciprocal connections with other cortical and subcortical structures, empowers healthy human adults with the capacity to consider the self's extended existence throughout time. The most complete expression of this autonoetic awareness occurs whenever one consciously recollects or reexperiences a happening from a specific time in the past, attends directly to one's present or online experience, or contemplates one's existence, projecting it to a time in the future. By autonoetic consciousness, adults are empowered with the ability to mentally represent and become aware of their experiences in subjective time. Like other frontal phenomena, autonoetic consciousness itself is contentless; contents are represented in the posterior cortex and can be permeated by autonoetic awareness in *ecphory*.

### 17.4.3 Again on Learning and Memory

From the previous considerations it seems to follow that *semantic memory* constitutes to a certain extent a form of synthesis between memory and learning,<sup>67</sup> since it presents a form of pragmatics, given the previously discussed fact that symbols always have a procedural component.<sup>68</sup>

Therefore, it seems to me to be more appropriate to speak of *experiential* memory for fulfilling three different goals:

- (a) Covering the proper memory processes as *common to all mammals*,

<sup>65</sup>[CRAIK/TULVING 1975].

<sup>66</sup>[WHEELER *et al.* 1997].

<sup>67</sup>[MORRIS *et al.* 1977].

<sup>68</sup>This could be the explanation for the results of a study showing that both semantic and episodic memory can be impaired by the same kind of lesions [GABRIELI *et al.* 1988]. The authors suggest that this can be due to the circumstance that these two "memories" share some common component.

**504** Memory

- (b) But also as *distinct from learning*, and  
 (c) Keeping apart a symbolic (semantic) memory as *typically human*.

Evidence for these distinctions comes from the fact that there is an important difference between memory activation (for instance, to complete an incomplete statement) and procedural learning<sup>69</sup>: Even if both operate in an unconscious way on preexisting representations, activation [Sec. 17.1] influences the performance for only 1 or 2 hours, whereas procedural learning can persist for months. Moreover, activation works on a single structure, whereas procedural learning involves the combination of a sequence of preexisting memory structures for a specific task [Sec. 16.1]. Another even more striking evidence is that all those neural areas are in competition. In particular, using the techniques of Functional Magnetic Resonance Imaging (fMRI), the authors of a relative recent study<sup>70</sup> show that, at least in humans, there is a competition between the medial temporal lobe-based and striatum-based memory systems: Experiential memory strongly relies upon the MTL [Sec. 17.3] whereas so-called procedural memory (i.e. learning) relies more on the striatum. These systems are continuously activated and deactivated, whenever two fundamentally incompatible requirements occur: The need for flexibly accessible knowledge (supported by MTL) and the need to learn fast, automatic responses in specific situations (supported by the striatum).

As anticipated, the conclusion of this investigation could be that the hippocampus is not involved in learning. Recent studies<sup>71</sup> show that only the so-called episodic (experiential) but not the so-called semantic component of memory is dependent on the hippocampus. For LeDoux<sup>72</sup> this is evidence that the hippocampus is involved in remembering personal experiences but not objective (i.e. semantically loaded) facts.

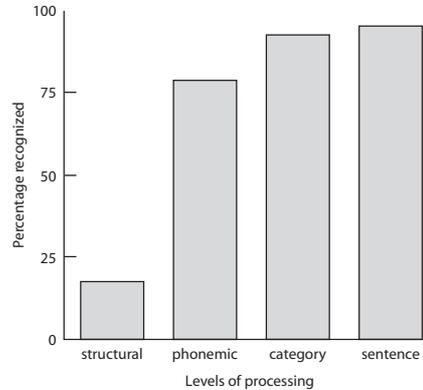
**17.5 Stages of Memory Elaboration**

As we have seen, in the brain there are both specialized and general-purpose systems.<sup>73</sup> Specialized memory subsystems have the ability to retain what they have just processed for a few seconds. The general-purpose systems consist of a workspace and a set of mental operations (called executive functions) that are carried on by information held in the workspace. The “where” processes are active during delay periods in spatial tasks and therefore are implied in the indexical functions. Executive functions are not partitioned in the prefrontal cortex on the basis of stimulus domain, which shows that they seem to be spread across multiple regions in the frontal cortex.

**17.5.1 Iconic Memory**

The first mnemonic recipient of external inputs is the so-called iconic memory, a sensory blackboard (for the distinct sensory modalities) from which the contents decay rapidly in a spontaneous way (even if the agent does nothing). This decay begins as soon as the stimulus first appears. However, decay can also happen because a second stimulus is perceived that *masks* the previous one.<sup>74</sup> The distinct forms of iconic memory are connected with sensory modalities and are therefore called visual memory, echoic memory, and somato-sensory memory [Sec. 4.1]. The capacity of these specialized subsystems is an indexical ability: Through it we refer to evanescent experiences by attaching new items to them and therefore by performing given operations.<sup>75</sup> This is through

<sup>69</sup>[GRAF *et al.* 1984].<sup>70</sup>[POLDRACK *et al.* 2001].<sup>71</sup>[VARGHA-KHADEM *et al.* 1997b].<sup>72</sup>[LEDOUX 2002, pp. 108 and 114–16] [COHEN/EICHENBAUM 1993].<sup>73</sup>[LEDOUX 2002, pp. 176–88].<sup>74</sup>[WILLINGHAM 2001, pp. 145–51].<sup>75</sup>[PEIRCE *CP*, 1.169].



**Fig. 17.4** The basic levels of processing effects in humans, showing that words that are processed more deeply are better remembered than words that are processed more shallowly. Adapted from [WILLINGHAM 2001, p. 176]. However, this situation can be reversed when there is matching or mismatching between memory encoding and retrieval [WILLINGHAM 2001, pp. 187–90].

the ability to cross top-down experience anticipation and bottom-up sensory updating through assimilated experience. Evidence of this indexicality is that for humans, who have a symbolic form of expression, it is about as easy to remember seven letters as seven words or concepts (information is organized in stacks or information chunks<sup>76</sup>).

The iconic memory is the mnemonic reservoir in which the iconic aspect of representation is made available for further processes and decisions [Subsec. 3.7.2] and can especially drive further attentional processes by representing the context out of which some specific details can emerge or change.<sup>77</sup> The iconic memory has a large capacity but this information is mostly and very soon deleted through a filter interfacing the iconic and the primary working memory.

### 17.5.2 Short- and Long-Term Memory

Hebb was the first to propose a distinction between long-term memory (LTM) and short-term memory (STM) [Subsec. 3.8.2]. The idea was that an item acquired by the STM and kept active for a certain time (some minutes) subsequently can pass to the LTM (where it can last years). It was believed that the more an item is meaningful (and therefore processed), the more it is consolidated and stored in the LTM [Fig. 17.4]. This proposal fitted with certain data; particularly in damaged human brains it was shown that there is a double dissociation between LTM and STM. Atkinson and Shiffrin refined the model by advancing a distinction between sensory stores (iconic memory), short-term storage and long-term storage<sup>78</sup> [Fig. 17.5]. Short term memory is characterized here by a very limited capacity and fragility of storage. It was assumed that in STM there is a spontaneous decay process whereas for LTM interference is important only in forgetting.<sup>79</sup> Subsequent works have shown that there is actually a mix of spontaneous decay and interference.<sup>80</sup> We do indeed distinguish between proactive interference when new learning is

<sup>76</sup>[MILLER 1956]. <sup>77</sup>[CHUN/NAKAYAMA 2000].

<sup>78</sup>[ATKINSON/SHIFFRIN 1968]. See also [EYSENCK/KEANE 2000, pp. 151–82].

<sup>79</sup>[BROWN 1958] [PETERSON/PETERSON 1959]. <sup>80</sup>[BADDELEY 1990, pp. 31–7].

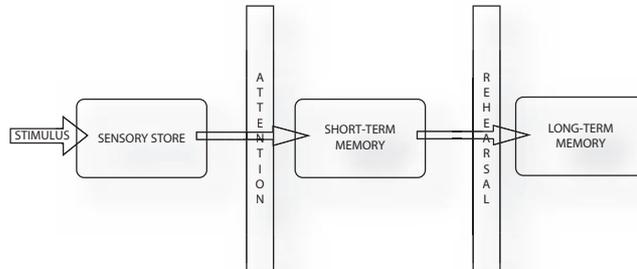


Fig. 17.5 Atkinson and Shiffrin's model of memory stores.

disrupted by old habits, and retroactive interference when new learning disrupts old habits. In my opinion, in both cases we are dealing with interference between learning and memory [Sec. 17.4].

In several studies of Bailey, Chen, and Kandel<sup>81</sup> it was shown that the LTM is not simply a temporal extension of the STM, since the actual number of involved synapses, as time flows, grows with long-term sensitization and decreases with long-term habituation [Sec. 16.3]; that again confirms learning and memory to be alternative modalities. The fact that the same cerebral site can give rise to both STM and LTM and is consistent with a one-process theory. However, the fact that the mechanisms ruling these two forms of memory are different is consistent with the two-process theory, since STM produces *changes in the function* of the synapse (releasing less or more neurotransmitters), strengthening or weakening preexisting connexions, whereas LTM requires *anatomical (structural) changes* [Subsec. 3.4.4]. The chemical basis of the difference between STM and LTM is that in the former case there is a single pulse of serotonin [Fig. 17.6], in the latter there is at least a series of five separated pulses.<sup>82</sup> The reason is that repeated pulses of serotonin produce higher concentrations of cyclic AMP, causing protein kinase A to move into the nucleus, where it activates genes, starting an expression process (a pure informational step): Protein kinase A is able to activate a regulatory protein called CREB (another cyclic AMP) which binds to a promoter.<sup>83</sup> Actually, on the line of Jacob and Monod's contribution [Subsec. 11.1.2], two different types of CREB were found, an activator (CREB-1) and a suppressor (CREB-2). Protein kinase A activates the former and deactivates the latter [Fig. 17.7]. It is important to stress here that without local protein synthesis there is no growth of the new synapse. It is, as we have seen, a form of marking [Subsec. 8.2.3].

### 17.5.3 Working Memory

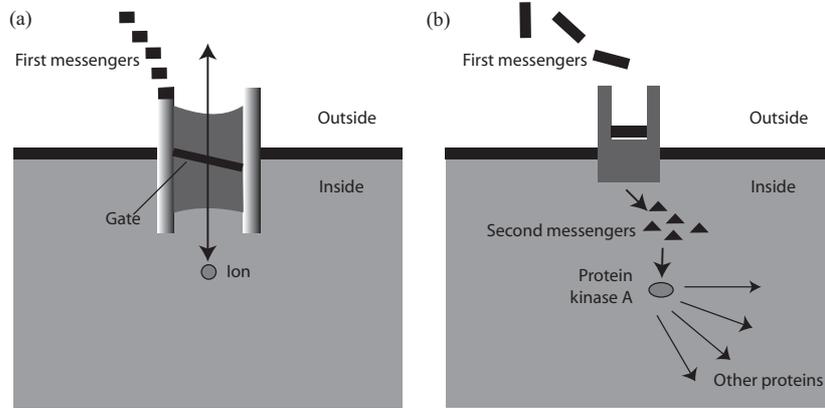
With the accumulation of further evidence, however, it seemed increasingly difficult to admit sharp boundaries between two (or three) different memory stores. There was mounting evidence that memory was rather an issue of gradual temporal processing and sedimentation.<sup>84</sup> Therefore, Baddeley and Hitch proposed using the concept of working memory instead of short-term storage.<sup>85</sup>

<sup>81</sup>[BAILEY/CHEN 1988, BAILEY/CHEN 1988, BAILEY/KANDEL 1993]. See also [KANDEL 1999, pp. 212–15].

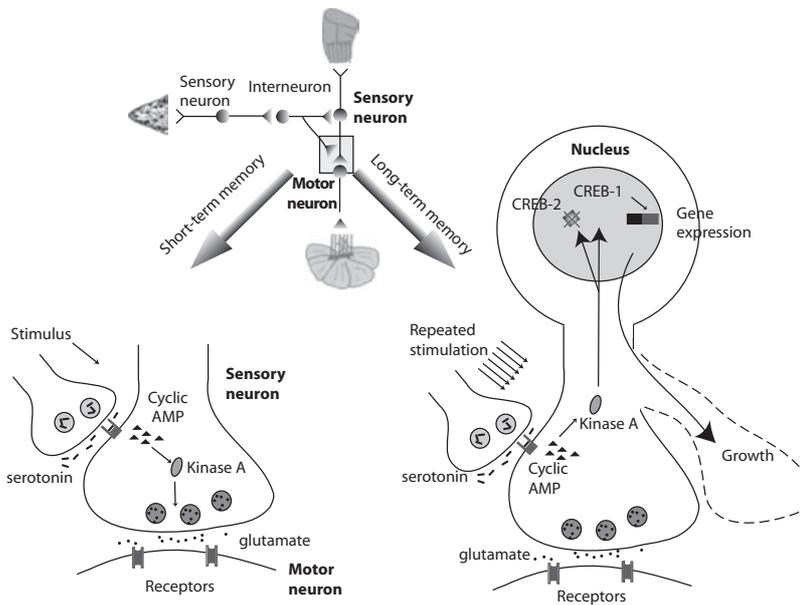
<sup>82</sup>[KANDEL 1999, pp. 212–15 and 263–7]. <sup>83</sup>[GOELET *et al.* 1986] [DASH *et al.* 1990].

<sup>84</sup>[FUSTER 2003, p. 120].

<sup>85</sup>[BADDELEY/HITCH 1974]. The original scheme of Baddeley strongly relies on a proposal of L. Lichtheim [CHANGEUX 2002, pp. 116–17].



**Fig. 17.6** (a) Fast action of first messengers (neurotransmitters) through an ionotropic receptor, opening a gate of an ion channel (STM). (b) Slow action of first messengers (neurotransmitters like serotonin) through a metabotropic receptor (LTM). This type of receptor has no ion channel. Instead, it activates a second messenger (cyclic AMP) that in turn frees the catalytic unit of protein kinase A. This protein induces a process in which neurotransmitter glutamate is finally released [Subsecs. 7.6.2 and 11.3.1]. Adapted from [KANDEL 2006, p. 228].



**Fig. 17.7** Chemical mechanisms underlying short-term (bottom left) and long-term (bottom right) memory. Here, neurons are represented by circles, synapses by triangles (compare with Fig. 16.2). These processes are magnifications of the box shown above as a gray insert. Adapted from [KANDEL 2006, p. 265].

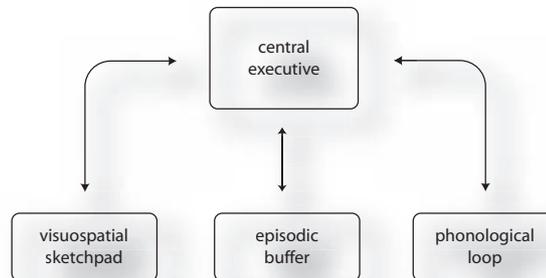


Fig. 17.8 Structure of the working memory according to Baddeley.

Working memory is an *active* memory [Sec. 17.2] lasting between a few seconds and one minute, and it has a limited capacity (it selects among the information of the very short iconic memory).<sup>86</sup> The necessity for a working memory arose from a functional requirement: Many brain areas bring together elements out of which behavior is composed [Sec. 15.1]. A memory is needed to hold these elements and keep them assembled together for the time of the action.<sup>87</sup> Therefore, the working memory connects sensory or even motor features with memory.<sup>88</sup> In humans, the working memory consists in a modality-free central executive that is responsible for attention focusing, a phonological loop (holding information in a phonological form), a visuo-spatial sketchpad, specialized for spatial and/or visual coding. To this initial model proposed by Baddeley and Hitch, the first author later added an episodic buffer (related to the personal story of the subject [Subsec. 17.4.1]) in order to account for the episodic LTM, at least in humans<sup>89</sup> [Fig. 17.8]. Every component is relatively independent from the others. Later on, the phonological store was distinguished from a specific articulatory process for speech production.

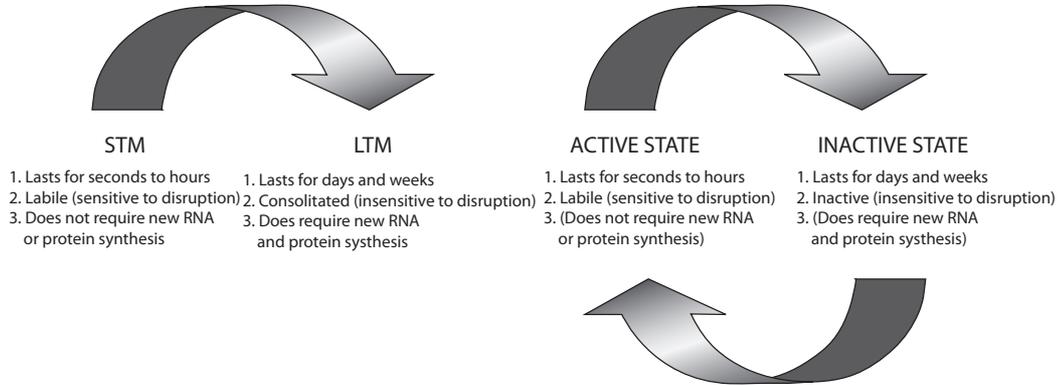
An examination of the working memory shows that neuronal firing is location-specific so that it provides a good example of a compartmentalized and constrained architecture of cognitive systems that is reminiscent of that observed in sensory systems. The organization of the prefrontal cortex into domains is consonant with its anatomical connections.<sup>90</sup> However, there are also some flaws with this concept of working memory. Recently, it has been pointed out that the phonological loop has little practical significance. With regard to the visual sketchpad, there is also evidence of separate visual and spatial systems. Finally, the central executive does not seem so unitary. Despite these problems, it seems that this is a good general model with a certain explanatory power. A similar model should also be true for other animals, and certainly for nonhuman primates. It is very probable that here we also have at least a spatial-visual part of the working memory as well as some equivalent of the episodic buffer to account for a higher social interaction, at least of primates.

The previous distinctions also suggest that the working memory is rather connected with attentional processes [Subsec. 14.1.1], and therefore with behavior and learning rather than

<sup>86</sup>[BADDELEY 1988]. <sup>87</sup>[NEWELL 1990, p. 309].

<sup>88</sup>[SUPÈR *et al.* 2001]. See also [JEANNEROD 2009, pp. 135–6].

<sup>89</sup>[BADDELEY 1992, BADDELEY 2000]. <sup>90</sup>[GOLDMAN-RAKIC *et al.* 2000].



**Fig. 17.9** On the left the traditional view of memory, on the right the new understanding. Adapted from [NADER 2003a].

with memory,<sup>91</sup> while the true memory system is the LTM, which requires a certain (smooth) consolidation time. This is confirmed in part by the fact that the traditional idea of a serial structure of the three systems is ultimately not supported by experimental evidence<sup>92</sup>: For instance, sensory (iconic) memory may be *directly* stored in the LTM.

As we have seen, memory is a continuous process of recreation and reenacting.<sup>93</sup> The crucial point is that any reactivation of previous memories will return them to the labile state they were in before consolidation. In other words, instead of distinguishing between an STM and LTM or even between a working memory and a LTM, it is much more suitable to distinguish between *active* *inactive* states of a *single* memory circuit [Fig. 17.9]. Evidence of this dynamic aspect of memory is the case of a person with an amputated arm who can feel the fingers of this arm if it is touched to the face, which implies that there is a sensory map of the hand on the face skin.<sup>94</sup> This means that, after amputation, there was a cortical *rearrangement* such that the hand area was innervated by the surrounding cortex (the face region is near to the hand region). Also Kaas’s work shows how the brain reconstructs areas damaged where sensory information is mapped. If the nerve of a finger is deactivated by cutting it, its representation in the hand map is covered by the representation of the other fingers.<sup>95</sup>

## 17.6 Concluding Remarks

Memory is not a collection of static entries but rather a dynamic activation and reconstruction of memories in the context of the operation at hand:

- An important emotional component takes part in the mnemonic activity.
- So-called procedural memory turns out to be rather a part of the learning activity, while an experiential character is typical of memory.

<sup>91</sup>[BADDELEY 1990, pp. 49–88].

<sup>92</sup>[GAZZANIGA *et al.* 1998, pp. 301–50] [MARKOWITSCH 2000].

<sup>93</sup>[NADER 2003a, NADER 2003b].

<sup>94</sup>[GAZZANIGA *et al.* 1998, pp. 611–53]. On this point I also suggest that the reader have a look at the lively summary in [FRITH 2007, pp. 70–4] and literature therein.

<sup>95</sup>[KAAS 1995, KAAS 2000]. See also [ANTONINI/STRYKER 1993].

**510** Memory

- Tulving introduced a distinction between semantic and episodic memory. My own suggestion is that the proper memory is the episodic one while the so-called semantic memory is concerned with the human system of learning.
- The hippocampus is involved in the experiential memory.
- Traditionally there is a step-processing view of memory, from short-term memory to long-term. However, today the distinction between the active state and the inactive state of a single memory circuit is preferred.

Let me also summarize the results of the last three chapters. The functions of behavior, learning, and memory (as well as the decisional, emotional, and cognitive systems) will be the basis of the higher symbolic functions (and systems) that will be treated in the next part of the book. *Behavior* is mainly involved in old (acquired) responses to new stimuli and always has an instinctual basis; *learning* is involved with new responses to old (already encountered) stimuli and is always accompanied by some form of awareness; *memory* concerns old (stored) responses to old (past) stimuli and its ground is habit formation [Subsecs. 8.2.1 and 12.5.1]. This shows that, ultimately, even if selection by consequences also enters into behavior [Sec. 16.2], it is Pavlov's (and Watson's) variety of behaviorism that captures the essence of (basic) behavior, while Skinner's (and Thorndike's) is essentially concerned with (basic) learning [Subsec. 15.1.2], resolving in this way a certain ambiguity that beset this school from the start [Subsecs. 15.1.3–15.1.4]. Some of the worries derive from a sort of overextension of behaviorist methodology to learning and even memory [Subsec. 15.2.3].

Summing up, *behavior* is essentially present-directed. It concerns consuetudinary actions without connections with problematic aspects or traumatic experiences. It is the function and process in which learning and memory are mixed. *Learning* is mainly addressed, in a top-down fashion, to the future, since it consists in solutions presenting general (iconic) characteristics [Secs. 8.1, and 12.1] and therefore applicable to a wide range of possible situations. It is therefore aimed at problem-solving and error correction. *Memory* is associative (indexical) in character and, in a bottom-up way, deals with traumatic events and external impacts on the subject. Therefore, if two events have been experienced in the past and one of them has a strong traumatic valence, the other one will also maintain such a valence thereafter, even if it is fully unrelated with the traumatic event and only contiguous in space or time [Subsec. 14.2.2].

Therefore, even if the memory and learning functions obviously intersect, they respond to:

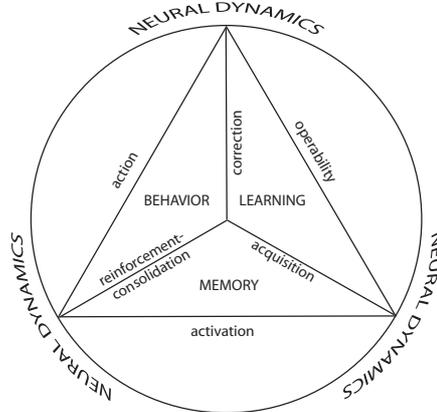
- Different goals: On the one hand what the relevant facts are to be *remembered*, and on the other hand what fundamental competencies need to be *acquired*, in order to be operative in the right context,
- Different mechanisms: On the one hand *habituation* through associative acquisition, on the other hand *sensitization* through error correction.

Also behavior and learning respond to:

- Different goals: On the one hand *to act* properly in a given environmental niche or in a certain ontogenetic context, on the other hand to assimilate operative specific *competencies*.
- Different mechanisms: On the one hand *automatization* of behavior through reinforcement, on the other hand *sensitization* through error correction.

Finally, memory and behavior also respond to:

- Different goals: On the one hand which facts should be *remembered* as especially significant for the organism, on the other hand *which acts are necessary*.



**Fig. 17.10** Behavior and learning share error correction (determined by the decisional system), learning and memory share acquisition (determined by the sensory-cognitive system), memory and behavior share consolidation (reinforcement, determined by the emotional system). All the interactions between systems and crossing processes are due to the dynamics of neuron population, as explained in the previous chapters. The whole is a cybernetic or functional circle in the sense explained in Sec. 8.4 [see also Fig. 8.4].

- Different mechanisms: On the one hand *habituation* through associative acquisition, on the other hand *automatization* through reinforcement.

This helps us to understand how high functionalities have phylogenetically arisen [Secs. 9.10 and 11.7]. The brain arises as a specialized organ for monitoring the effects of motor actions on the environment [Sec. 12.3]. Across a long evolutionary history, it shows increasing specialization in both structures and functions [Sec. 11.5 and Ch. 13]. The differentiation between behavior, learning, and memory is a true circuit-, or function-modularization process. However, I have already stressed that there is no modularization without integration and centralization [Subsecs. 2.4.4 and 3.4.5, Sec. 3.6]. It is indeed important to emphasize the fact that behavior and learning share error correction [Fig. 17.10]. The reason is that they share the motor-attentional-decisional system, which is the brain subsystem enabling higher organisms to correct errors [Fig. 13.11]. On the contrary, memory and learning share acquisition (in the forms of memory acquisition and skill acquisition, respectively): Indeed, they share the sensory-cognitive system which is the source of any information acquisition by the brain. Finally, memory and behavior share consolidation (memory consolidation and reinforcement, respectively); the reason is that they share the emotional system, the only subsystem of the brain concurring to a durable but plastic consolidation of any act or percept.

Summarizing, the triad acquisition-correction-consolidation is an expression of the mammal brain, articulated in sensory, motor, and emotional brain subsystems with the higher functionality of learning, behavior, and memory. Nevertheless, that triad expresses an elementary functionality of life that already exists at a bacterial level. We have indeed remarked [Sec. 12.2] that life is characterized by the ability to acquire information about the environment, to correct responses, and to consolidate those patterns that (to a certain extent) fit with external conditions. In conclusion, I would say that behavior and learning *emerge* from memory,<sup>96</sup> memory and learning *merge* in behavior, and memory and behavior *converge* in learning.

<sup>96</sup>[PEIRCE CP, 1.167].

**OUP UNCORRECTED PROOF – REVISES, 1/6/2011, SPi**