diagram is static, it provides a basis for humans to reason about how the whole mechanism generates a phenomenon (in this case, recognizing an object or locating it in space) by performing the operations portrayed in the order shown. In this manner, researchers can mentally simulate the operation of the mechanism.

In thinking about the organization of a mechanism, humans start by thinking sequentially. The edges in Figure 14(b) are thought to carry activity from inputs at the bottom upward to higher processing areas. But the researchers who developed the diagram were very much aware that in the brain there are as many recurrent projections (neural projections from areas viewed as later in a pathway to those viewed as earlier) and that each of these areas sends and receives projections from regions of the thalamus and the basal ganglia (see Section 5.4). One can add additional rectangles and arrows to represent these, but it quickly becomes impossible to simulate the mechanism mentally. Instead, researchers often supplement a verbal and diagrammatic representation of a mechanism with a mathematical one, developing a computational model (Section 3.5). We will illustrate this in Section 6.3, but first we turn to an account of explanation that proposes using computational models to supplant the need for mechanistic accounts.

6.2 Dynamical Systems Explanations

Researchers in the life sciences often compare their sciences to physics. Explanations in many domains of physics appeal to laws that characterize how variables describing a system will change over time (hence, dynamical laws, often taking the form of differential equations). The explanation involves a demonstration that from the law and a specification of conditions at one time, one can derive what will happen at other times (Hempel, 1965). In many cases, the application of laws is far from simple and requires computational simulation to determine the consequences of the laws. Some cognitive and brain researchers apply similar strategies to explain behavior, and some philosophers have embraced these as fully legitimate explanations that do not require characterizing a mechanism.

A common approach of these investigators is to characterize a state space – a multidimensional space in which each dimension corresponds to a variable that describes the system. Consider three dimensions on which a gas can vary: pressure, volume, and temperature. Characterizing such a space would be of little explanatory interest if in fact the system could evolve from any point in the space to any other. What laws do is restrict the trajectory that the system can take through the space. The gas law:

https://doi.org/10.1017/9781108946964 Published online by Cambridge University Press

$$\text{Temperature x Gas constant x Moles of gas} = \text{Volume x Pressure} \tag{6.1}$$

Imposes the restriction that when volume is held constant and temperature increases, so must pressure. If the actual system is shown to be similarly limited in its possible trajectories, then proponents of *nomological explanations* argue that the laws characterizing the possible trajectories through the state space explain why the system behaves as it does.

The gas law example (Eq. 6.1) does not specifically take time into account. But other laws spell out how values of variables will change over time. These give rise to what are termed *dynamical systems explanations*. Some dynamical laws, such as $x_{t+1} = x_t + 1$, are relatively simple: this law simply asserts that the value of the variable x increases by 1 at each timestep. If that is what happens, then the law explains why the variable follows this ascending trajectory. In many cases, the law will involve a more complex equation and produce surprising results. A mathematical function that is often employed to illustrate complex behavior is the logistic map function, $x_{t+1} = Ax_t(1 - x_t)$. The reader is invited to try various values of A between 3 and 4, picking an initial value of x_t between 0 and 1, and calculating the results for several steps. For example, with A = 3.3, values will initially fluctuate (a period referred to as the *transient*) but eventually begin to oscillate between two values (0.47943 and 0.82360). When A is increased to about 3.5, the values, after the transient, will jump sequentially between four values (approximately 0.49, 0.87, 0.38, and 0.83).[11]

These stable values are referred to as *attractors* – the idea is that values in their proximity will move closer to (fall into) the attractor. Figure 16(a) shows a two-dimensional state space in which there is just one fixed point attractor; initial values anywhere in the state space will fall into the attractor at the center. Sometimes attractors have a more complex structure, such as the cyclic attractor shown with a dashed line in Figure 16(b). In this case, no matter where the system starts, it arrives at a circle, around which it will progress indefinitely. Sometimes a space may have multiple attractors so that, starting from different points, the system may settle into different attractors. By representing a state space and identifying attractors in it, researchers can determine how the system will evolve from whatever point it currently occupies.

A much-cited dynamical model developed to explain animal behavior is the Haken–Kelso–Bunz (HKB) model of coordination dynamics. It describes

[11] The logistic map function is of interest because it can also demonstrate what is known as deterministic chaos – for most values above A = 3.6, the function will trace out a continually changing set of values without ever repeating, assuming one calculates the full real value of x. For illustrations, go to www.youtube.com/watch?v=ovJcsL7vyrk.

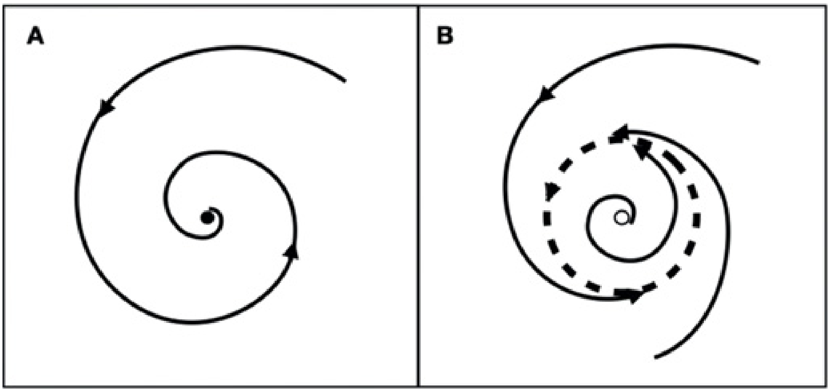

Figure 16 Attractors in a two-dimensional state space. (a) A point attractor. (b) A cyclic attractor.

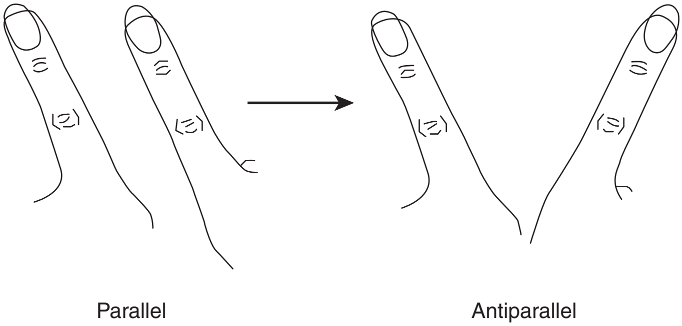

Figure 17 Parallel and antiparallel movement of fingers. Both can be maintained at slow speeds, but at faster speeds, only the antiparallel movement can be maintained. Figure by Hermann Haken released under the Creative Commons Attribution-ShareAlike 3.0 License.

phenomena such as the coordination between one's legs in walking (Haken, Kelso, & Bunz, 1985). To experience what the model describes, place both hands in front of you with the forefinger extended (Figure 17). Pretend your fingers are windshield wipers on a car. In most cars, wipers move in parallel (both tips move left, then both tips move right), but sometimes they move in an antiparallel fashion (the tips come together and then move apart). Try each pattern of movement, first slowly and then gradually faster. At slow-speeds most individuals can maintain both patterns, but when they try to speed up, they can only maintain the antiparallel pattern. The HKB model offers an explanation. It starts by describing the movement with the equation:

https://doi.org/10.1017/9781108946964 Published online by Cambridge University Press

$$V(\phi) = -\mathrm{a}\ \cos\ \varphi - b\ \cos 2\ \phi \tag{6.2}$$

in which ϕ is the phase relation between the fingers (or limbs more generally) and the ratio b/a is inversely related to the rate. In the state space described by Eq. 6.2, when b/a is high, corresponding to a slow speed, there are two attractors. However, when b/a is low, there is just one attractor. The loss of the attractor at faster speeds, on the dynamical systems account, explains your inability to maintain the parallel finger movement.

A notable feature of the HKB model is that the variable employed refers to a feature of the phenomenon (the angle between limbs), not to any proposed mechanism that is decomposed into components so as to account for the phenomenon. Proponents of dynamical systems accounts maintain that one does not need to enter into the nervous system to explain the inability to maintain the asymmetric movement. The phenomenon itself has structure that provides explanation (Chemero, 2000; Chemero & Silberstein, 2008). There are other examples where, merely from the characterization of the structure of phenomena, one can determine specific (and often unexpected) features of it. For example, from knowing the structure of tides around the ocean, one can determine that there must be a point in the ocean in which there is no tide. Likewise, from understanding how the circadian clock, discussed in Section 5.1, responds to light stimuli, one can infer that in some organisms, there is a time at which exposure to light will cause the amplitude of the oscillations to become 0 (which is to say, the clock will stop, as there is no longer an oscillation to represent time). This necessity was in fact demonstrated before the mechanism of circadian oscillation was known and does not depend on any details about the mechanism (Winfree, 1987; for discusssion, see Bechtel, 2021).

6.3 Dynamic Mechanistic Explanations

As we noted in discussing mechanistic explanations in Section 6.1, when mechanisms depart from sequential organization, it becomes challenging to simulate their behavior mentally. Even a simple feedback loop can present a challenge. As many people are aware from examples like thermostat controlling furnaces, feedback loops can generate oscillations (the temperature will rise after the thermostat turns the furnace on and fall after it turns it off). Accordingly, when an intracellular feedback mechanism was proposed to explain circadian rhythms (Section 5.1), it was expected to generate oscillations. But the question was whether the oscillations would be sustained indefinitely or dampen over time. To address that question, Goldbeter (1995) created a computational model that showed that under biologically plausible conditions, the mechanism would instantiate a cyclic attractor (Figure 16(b)). He therefore concluded that the biological mechanism implementing feedback could oscillate indefinitely. This explanation is much like that provided by the HBK model in Section 6.2, but here the variables refer to hypothesized operations of the components of the mechanism. Since in this case the explanation is a hybrid, drawing upon both mechanistic decompositions and the use of computational models to characterize the dynamical behavior of the mechanism, Bechtel and Abrahamsen (2010) refer to them as *dynamic mechanistic explanations*.

Perhaps the best-known dynamical computational model in neuroscience is the Hodgkin–Huxley model of the action potential. Hodgkin and Huxley (1952) decomposed the current across the neuron membrane into components for sodium, potassium, and other ions and developed an equation for how each contributed to the current*I*across the membrane (Figure 18). From this they produced an overall equation that described how the current changes as the electrical potential changes:

$$I_m = C_m \frac{dV_m}{dt} + \overline{g_K} n^4 (V_m - V_K) + \overline{g_{Na}} m^3 h (V_m - V_{Na}) + g_l (V_m - V_l). \tag{6.3}$$

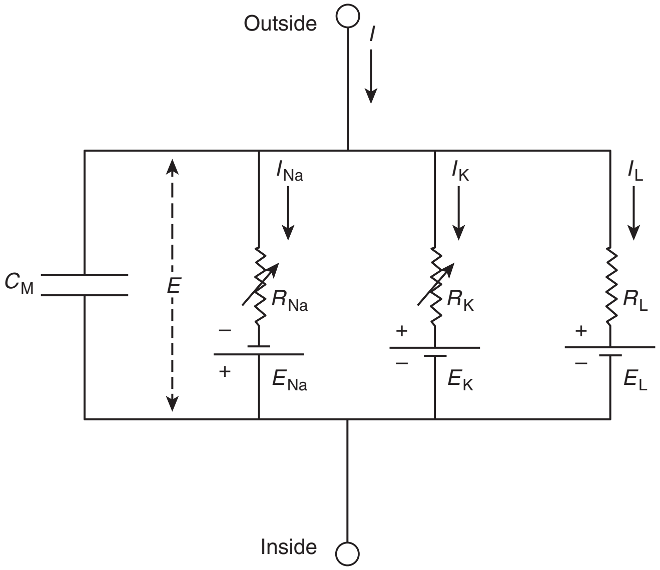

Figure 18 Hodgkin and Huxley's (1952) representation of the current *I* across the membrane in terms of the currents for sodium (Na), potassium (K), and leakage (l) due to other ions. *E* represents the membrane potential and *R* the resistance for each ion.

In Eq. 6.3, I is the current, C_m is the capacitance due to the membrane, V_m the electrical potential across the membrane, V_K, V_{Na}, and V_l represent the potential due to potassium, sodium, and leakage (other ions), and g_K, g_{Na}, and g_l the conductance for the various ions. n, m, and h are parameters used to fit the model to data. From Eq. 6.3, one can generate the pattern of the action potential (Figure 2).

Hodgkin and Huxley's accomplishment, which earned them the Nobel Prize for Physiology and Medicine, has been the focus of considerable philosophical controversy. Weber (2005) treated it as an instance of an explanation that derives a phenomenon from a law. In the nomological tradition, laws are typically distinguished from causal claims, and Weber (2008) subsequently offered a revised account, according to which Hodgkin and Huxley offered a causal explanation in which the ion currents caused the action potential. While granting the usefulness of the model as a description of the action potential, Craver (2006, 2008) has argued that it is not explanatory since it does not include, let alone characterize, what he takes to be the critical parts of the mechanism generating the action potential, the gates on the channels through which ions are allowed to enter or leave the neuron. It turns out that the coefficients of the parameters n, m, and h correspond to features of these gates, but this was only discovered years later. At best, Craver allows, Hodgkin and Huxley offered a sketch of an explanation that was only provided later. More recently, Levy (2013) has argued that the model does in fact provide a mechanistic explanation in so far as it presents the whole current as arising from the aggregate activity of each of the types of ions. The second, third, and fourth summed terms in the equation represent the current generated by each ion as a result of the difference between its current potential and the membrane potential. Levy contends that the Hodgkin–Huxley model captures the crucial activities in the mechanism. As a result, it offers a dynamical mechanistic explanation of how the changing concentrations of the ions give rise to an action potential. This debate illustrates different stances philosophers take on the nature of explanation and what is required to explain a phenomenon.

6.4 Network and Connectomic Explanations

As we have seen in various sections of this Element, the nervous system, and its various subparts, are often characterized as networks. The crucial idea of a network is that it consists of entities (represented as nodes) and connections between them (represented as edges). Networks are ubiquitous – any time entities are connected, they can be represented as a network. But some networks

https://doi.org/10.1017/9781108946964 Published online by Cambridge University Press

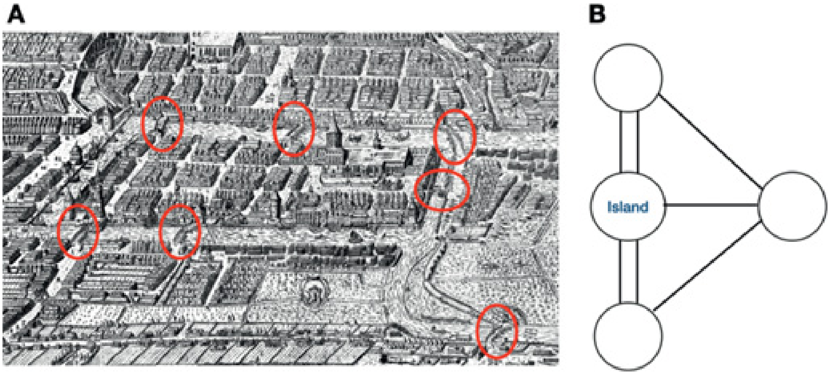

Figure 19 (a) Map of Konigsberg with the river and seven bridges highlighted by Bogdan Giuşcă and distributed under the Creative Commons Attribution-Share Alike 3.0 Unported license. (b) Network graph, in which nodes represent different landmasses and edges the bridges between them.

have distinctive properties that are sometimes viewed as explaining aspects of the behavior of the system instantiating the network.

One of the earliest examples of a network analysis was Leonhard Euler's solution to a problem posed by the bridges crossing the Pregel river in the Prussian town of Konigsberg (Figure 19(a)): Can one cross each bridge just once on a walk? He represented the different landmasses with a node and the bridges with an edge (Figure 19(b)). From this abstract representation, Euler proved no route is possible. For it to be possible to cross each bridge just once, each node other than the ones representing the starting and ending locations must connect to an even number of bridges. In this case, all four nodes connect to an odd number of bridges; accordingly, such a walk is not possible.

Starting in the mid–twentieth century, investigators identified a number of important features of networks that determine the properties of any actual system instantiating the network. Here we introduce just two concepts that turn out to be extremely important for understanding the brain: small worlds and hubs. To introduce these, we need to introduce some of the measures used to describe networks. One is the average of the shortest paths between each two nodes. A second is how clustered a network is: to how many of its neighbors a node is connected. In a randomly connected network, the average shortest path is short but clustering is low. In a regular lattice (a structure in which every node is connected to each of its neighbors), clustering is high (since a node is connected to all its neighbors) but the average shortest path is long. Watts and Strogratz (1998) showed that many networks in the real world are more like the

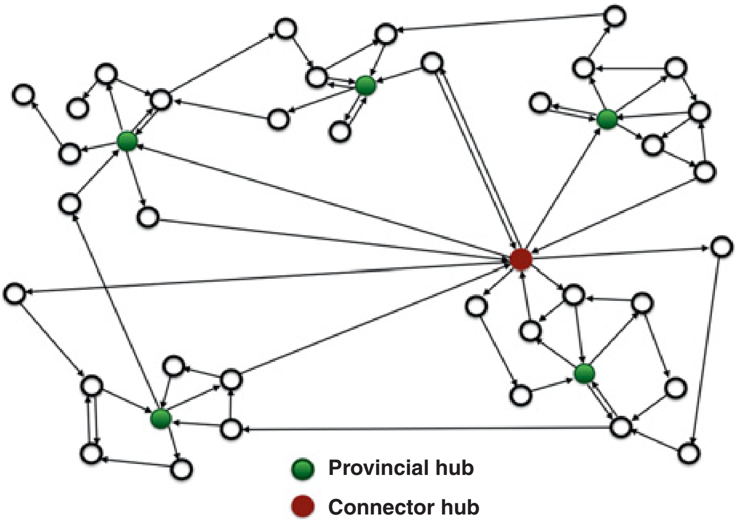

Figure 20 A network with relatively short average path between any two nodes, relatively high clustering, with some nodes having many more connections than others and hence serving as hubs.

one in Figure 20, in which the average shortest path is relatively short but nodes are also highly connected to their neighbors (collections of nodes that are highly connected to each other are often called *modules*). They call these networks *small world networks*.

A third measure is how the *degree*, that is, the number of connections from each node, is distributed. If degree is distributed normally (i.e., if values are equally distributed about the mean and decrease with distance from the mean), no node will be especially highly connected. But in many real world networks, Barabási and Bonabeau (2003) showed the degree is not distributed normally but according to a power law (a mathematical relation of the form $y = \mathrm{a}x^{-k}$). This results in a few nodes being highly connected while most have few connections. As illustrated in Figure 20, those highly connected nodes can be the basis for a local module (provincial hubs) or can serve to integrate modules (connector hubs).

A number of researchers have analyzed nervous systems in network terms. In Section 4.2, we described how researchers produced a complete connectome for the nematode worm *C. elegans*. This network turns out to have small world properties. Developing connectome representations for other species at the level of individual neurons is extraordinarily challenging, although researchers are getting very close to having such a map for the fruit fly (which has about 100,000 neurons). Instead, researchers concerned with connectivity in the neocortex of mammals have focused on connections between brain areas (e.g., BAs) and are analyzing these for their properties (Sporns, 2010, 2012). The principles of short average path length, high clustering, and hubs all appear to apply. Van den Heuvel and Sporns (2011) have further shown that the human brain instantiates a rich club structure – a set of regions, each of which serves as a hub, are more connected than would be expected even given their high degree. Given their network properties, these brain regions are thought to serve as a communication backbone for the whole brain.

Given the potency of concepts such as these to explain activity in networks, Huneman (2010) argues for treating topological explanation as a distinct form of explanation. In particular, he distinguishes it from mechanistic explanation since it does not focus on the contribution of parts but only on how they are connected. In more recent work, Huneman (2018) has addressed how topological and mechanistic explanations can be integrated. The basis for integrating them is that topological principles provide a basis for understanding the consequences of different modes of organization in biological mechanisms. When topological principles such as small world organization suffice to account for the phenomenon, it is the organization, not features of the individual components, that explain the behavior of the mechanism (Levy & Bechtel, 2013).

6.5 Control Mechanistic Explanations

In philosophical discussions, mechanisms are often portrayed as ready to operate whenever their start or setup conditions are realized (Machamer, Darden, & Craver, 2000). To experiment on mechanisms using techniques such as those introduced in Section 3, researchers try to set up conditions in which they do operate in a regular manner. However, in an organism, the continuous operation of a mechanism is often not needed and can in fact be harmful (just consider continually contracting the muscles in your legs). Instead, mechanisms need to be activated and deactivated as needed by the organism. The same is true of the machines human make. We do not desire a furnace to produce heat all the time. Accordingly, we employ thermostats that turn the furnace on when the temperature drops too low and off when it is warm enough. The thermostat represents a second machine that operates on the primary one, changing some of its parts so that it operates in different ways at different times. Biological organisms are replete with mechanisms that operate on other mechanisms. That is, in fact, what neurons and neural mechanisms do: they control the operation of other mechanisms such as muscles and glands.

There is an important difference between biological mechanisms and human-built machines. We design machines to be controlled by us. We turn our car

https://doi.org/10.1017/9781108946964 Published online by Cambridge University Press

engine on or off, and when on, we control the fuel supplied to it through depressing the accelerator. Who controls biological mechanisms? The short answer is the organism itself. Recognizing this, Maturana and Varela (1980) introduced the crucial idea that organisms are autopoietic: they build themselves by procuring matter and energy from their environments and directing it into the synthesis of their own bodies. This requires control over procurement and construction mechanisms. In addition, the tissues that make up organisms are prone to break down, requiring organisms to detect failures and deploy repair mechanisms (Rosen, 1972). In virtue of constructing and repairing themselves, organisms are sometimes referred to as *autonomous systems* (Moreno & Mossio, 2015).

Organisms are not agents over and above the mechanisms that constitute them. Autonomy results from the actions of the mechanisms constituting an organism. More specifically, it results from the deployment of control mechanisms. Like a thermostat, control mechanisms act on and change the configuration of other mechanisms in light of conditions either in the organism or its environment (Winning & Bechtel, 2018). To do this, control mechanisms must make measurements (or utilize measurements made by other control mechanisms upstream of them). The measurement component of the control mechanism results in the state of the control mechanism being determined by the value of the variable being measured. Again, the thermostat provides a model – a component internal to the thermostat is altered by the temperature in the environment. Given the measurement, the control mechanism produces a specific action on the controlled mechanism. This means that control mechanisms must be properly configured so that the changes that they make in other mechanisms are appropriate to the circumstances that the organism faces.

The word *autonomy* includes the Greek words for self (*autos*) and law (*nomos*), and thus signifies that an autonomous system sets laws for itself. Civil laws set norms for behavior. In determining the behavior of other mechanisms, control mechanism likewise impose laws or norms that govern that behavior (Winning, 2020). In the case of the thermostat, these norms ultimately derive from the humans who build and set the thermostat. Biological control mechanisms are not designed by humans. Rather, they are the product of evolution. In the course of evolution, those control mechanisms are retained that apply norms that enable organisms to maintain themselves and reproduce. Those that do not disappear over the course of evolution.

One reason control mechanisms are so crucial to living organisms is that organisms regularly confront different circumstances that require different responses. They need to be able to adapt to these. Some circumstances repeat and, like a thermostat, control mechanisms can direct the same response on each occasion. But organisms often confront novel situations that require tailoring their basic mechanisms in new ways. To deal with these situations, control mechanisms must exhibit a degree of flexibility, directing basic mechanisms to operate in novel ways. In Section 10, we will explore ways in which control mechanisms are organized so as to support creating effective responses to novel situations.

6.6 Summary

We have introduced several different perspectives on explanation: mechanistic, dynamic, and topological. Each appeals to different factors and seems applicable to specific phenomena. This suggests a pluralistic perspective, recognizing different types of explanation. It also suggests that different perspectives might be integrated, and we offered dynamic mechanistic explanations as one integrated perspective. Lastly, we noted the importance of control in biological organisms and described how the mechanistic perspective can be extended to characterize control mechanisms.

7 What Are Levels in Neuroscience and Are They Reducible?

The term *level* is widely invoked in neuroscience, and researchers and commentators often debate whether some levels should be reduced to others. Unfortunately, the term level is used in a wide variety of senses. In this section, we differentiate three notions of level that are prominent in discussions about neuroscience and identify the implications of each for reduction.

7.1 Marr’s Levels (Perspectives)

David Marr, a pioneer in the development of computational modeling in neuroscience (Section 3.5), began his book *Vision* (1982) with a critical assessment of what he saw as the current state of the discipline. Neuroscientists were accumulating many findings about how various parts of the nervous system operate using techniques such as those discussed in Section 3. But they were making little progress in providing an understanding of how the brain works. On his analysis, this was due to focusing on just one level, which he termed the *hardware implementation* level. Accounts at this level focus on parts of the brain and how each operates. To make progress in understanding the brain, he argued for the need for two other levels: those of *representation and algorithm* and of *computational theory*. At the representation and algorithm level, he argued that researchers should treat the parts of the brain as representing content and applying rules to manipulate those representations. Much of Marr’s own work was focused on the representation and

https://doi.org/10.1017/9781108946964 Published online by Cambridge University Press