Perspectives

Theoretical approaches to holistic biological features: Pattern formation, neural networks and the brain-mind relation

ALFRED GIERER

Max-Planck-Institute for Developmental Biology, Spemannstraße 35, D-72076 Tübingen, Germany

(Fax, 49-7071-601 448; Email, alfred.gierer@tuebingen.mpg.de)

The topic of this article is the relation between bottom-up and top-down, reductionist and "holistic" approaches to the solution of basic biological problems. While there is no doubt that the laws of physics apply to all events in space and time, including the domains of life, understanding biology depends not only on elucidating the role of the molecules involved, but, to an increasing extent, on systems theoretical approaches in diverse fields of the life sciences. Examples discussed in this article are the generation of spatial patterns in development by the interplay of autocatalysis and lateral inhibition; the evolution of integrating capabilities of the human brain, such as cognition-based empathy; and both neurobiological and epistemological aspects of scientific theories of consciousness and the mind.

[Gierer A 2002 Theoretical approaches to holistic biological features: Pattern formation, neural networks and the brain-mind relation; *J. Biosci*. **27** 195–205]

1. The whole and its parts

One of the most interesting facets of modern biology is the explanation of basic features of living systems in terms of the physical properties and interactions of its components. In this context, relating the whole to its parts is a major challenge for theoretical biology. The philosophical intricacies of this relation have a long history, and they were accentuated recently because attention peaked following the sequencing of the human genome. The Genome Project was originally defined in a more or less reductionistic mood: determine all genes; find out their respective functions; draw biological conclusions; apply them to medical problems. In some cases this is indeed possible, but by and large we realize that while the sequence of nucleotides in the human genome is now known, we do not yet understand it. One protein-coding gene can have different functions, one function is often due to a larger set of such genes; and non-coding sequences are involved in the regulation of gene activation in a most complex manner. . . . New programs of 'transcriptomics' and 'proteomics' are expected to help us to understand the function of the genome and its parts by revealing the interactions of proteins and the regulation of their synthesis in the cell. While this requires collection of even more data than the Human Genome Project, it also reflects a far-reaching change in attitudes: it is systems as a whole that scientists now want to understand.

In fact, in the history of biology, reductionistic and holistic approaches have alternated in different fields. Is holistic biology "back on stage"? Actually, the term "holistic" is rarely used. This is probably because in the past holistic ideas were often connected with the claim that physics as such is not enough to explain features of life. Nowadays, there is little doubt that the laws of physics apply to all events in space and time, including biological processes – and yet, systems have features that their components do not have. This is relevant not only to cell science and developmental biology, but also to behavioural and neural sciences.

In this essay, I would like to discuss aspects of three fields of biology that are of particular interest to this wishing to understand the physical basis of life processes: the generation of spatial patterns in development, the evolution of general capabilities of the neural network, and the brain-mind relation. The first topic is the one that has involved controversies between mechanistic

Keywords. Consciousness; Goedel; Heisenberg; human genome; lateral inhibition; systems theory; Trembley; Turing; Wolff

and organistic, reductionistic and holistic lines of thought right from the start in the 18th century. Only in recent times has it become clear that the *de novo* generation of spatial order is consistent with known laws and processes of physics, thus reconciling holism and physicalism. Understanding brain evolution requires that networks of gene regulation be related to neural networks, which is a real challenge for systems theory. The third topic – the brain-mind relationship – is entangled with epistemological questions about the range and limits of science in general, which are crucial for human self-understanding.

Encouraged by the editor's invitation to write from a personal angle, let me include a few remarks on the motives for science and on persisting and changing scientific attitudes. They may perhaps be traced back to my years as a student of physics in post-war Göttingen where I found, in Werner Heisenberg's new institute, an intellectually most stimulating environment, with three particularly remarkable facets: discussions on the philosophy of nature and human cognition right from the start; an ideal of dramatic, romantic science, the model being the invention of quantum mechanics in the twenties of the 20th century; and the enthusiastic prognosis put forward by my Ph D thesis advisor, Karl Wirtz, that molecular physics would eventually provide the basis of our understanding of living systems. I soon shifted to biology and worked in the new Max Planck Institute for Virus Research in Tübingen – just in time to experience a most romantic phase, the rise of molecular biology between the discovery of the double helix and the resolution of the genetic code. Our model was Tobacco Mosaic Virus, in which we showed and studied the role of virus nucleic acids as genetic material, followed by work on the role of polyribosomes in protein synthesis. But then, in the mid-1960s, I thought the real challenge would now be to understand multicellular organisms and their development. This field still had an old-fashioned flavour, being described as frustrating owing to lack of specificity of effects, such as those of Spemann's organizer. To be sure, at first we had the reductionistic attitude that it might now be up to us molecular biologists to tell the embryologists, with all their fuzzy notions of morphogenetic fields, polarity, competence and gradients, what their field was all about. But this attitude changed: soon our wonder at the marvelous holistic and phenomenological world of embryology reached the highest levels, and I became particularly intrigued by what is perhaps the most holistic of the problems, the generation of spatial patterns. We chose *Hydra* regeneration as our experimental system because it is a most puristic model for the *de novo* generation of spatial order within originally almost uniform tissue. Thereafter, we took up developmental neurobiology. As a theoretical project, my colleague Hans Meinhardt and I searched for physical principles underlying pattern formation, the topic of the next section. And then, what about philosophical facets? Yes, I am fascinated by what biology might tell us about the roots, scope and limits of human cognition and the intricacies of the brain-mind relation, the topic of the last section.

2. Organisms, mechanisms and the origins of developmental biology

How can physical laws and processes account for the *de novo* production of spatial patterns in cells and tissues starting from near-uniform conditions? Obviously knowledge of the molecules involved would be necessary for a full explanation, but it would not be sufficient. Even a complete list of all these molecules would not in itself explain the resulting spatial structure of, say, a mouse. In general, pattern formation is a systems feature. A cloud is condensed water, a snowflake is frozen water, H_2O; there are no mysteries about the molecules involved, and yet this fact is not enough to allow any of us to understand the form of clouds or the beauty of snowflakes. Ultimately, it is a combination of material knowledge and systems theory, of both molecular and mathematical facts, that is required. In the same way, investigating biological morphogenesis is a two-way process, bottom up starting from interactions of molecules and cells, and top down starting from phenomena, such as patterns and proportions. And since the two approaches are often correlated with the different outlooks of the scientists involved, not excluding their philosophical and metaphysical ideas, it is not surprising that their coming together may be retarded by psychological obstacles. This can be traced a long way back into the history of science.

In a sense, Aristotle can be considered the founder of biology as a science, because it was he who first postulated that reproduction and metabolism – and not features such as breathing – define life. And he took life processes, with their holistic and goal-directed features, as a model for physics as a whole – here was harmony, not conflict, between physics and biology. Only in modern times, when Galileo, Kepler and Newton laid the foundations of modern physics with mathematical laws governing the movements of bodies, did the relation between mechanisms and organisms, between the living and the non-living world, become such very challenging and puzzling problems. Only then did the question arise: How can a physics developed exclusively from studies in the inorganic domain claim validity for all events in space and time, which do, after all, also contain living organisms? Is the living body just a machine, as Descartes postulated, with perhaps some vague allowance for effects of the soul mediated in man by a small part of the brain,

the pituitary gland? Is the seemingly new formation of the organism in each generation just an illusion, and was it actually present even in the egg? Does this mean that all future generations of an organism are contained in the body, like Russian dolls within dolls? Is there nothing but mechanical unfolding of preexisting, invisibly small structures? If this were so, it would imply that there is no real development at all and therefore, of course, no developmental biology either, and no developmental biologists after all.

It appears rather strange to us that this doll-within-doll concept was a dominant theory in the 18th century, when it was propagated especially by Bonnet and Haller. The alternative is that there is real epigenesis, as Aristotle had postulated 2000 years earlier. In 1759, Caspar Friedrich Wolff, in his Ph D thesis entitled "Theoria generationis", provided strong support for the concept of *de novo* generation by his experimental work on the chick embryo (see Roe 1982), thus becoming one of the early pioneers of modern developmental biology. Fifteen years before, in 1744, Abraham Trembley had published his spectacular discoveries on the regeneration of Hydra: each dissected piece of a polyp will develop into a complete animal.

The impact of Trembley's discovery sounds almost unbelievable in our times. Regenerating polyps were talked about in literary, philosophical and theological circles. Embassies considered it their duty to keep their governments informed about progress in Hydra research, and an observer declared that the discoveries of polyp regeneration and those on electricity were the two outstanding achievements of the 18th century. But even with less enthusiasm, giving modern scientific skepticism its due, it is hard to overlook the importance of Trembley's work. He can be considered the first experimental developmental biologist in the history of science. And his attitude towards science was also most remarkable. In the last pages of his book he writes that "*what we actually know is still very little in relation to the innumerable wonders of nature. The best method to understand known facts is to discover new facts. Nature is to be understood with the help of nature, not by our preconceptions which are too limited for grasping such great objects of research as a whole.*"

3. Some physical, mathematical and logical aspects of biological pattern formation

Thus, Caspar Friedrich Wolff, and Abraham Trembley even more, left us with the question of how the generation of patterns in biological systems, the development of striking spatial order by internal processes within cells and tissues, can be explained on a physical basis. Many scientists, up to the time of Spemann in the 1930s, had thought or guessed that this might not be possible at all. A new physics, or some extraphysical principles, might be required. However, we now know that this is not the case; rather conventional molecular interactions and movements, even passive movements of diffusion, are good at pattern generation. This was the fundamental discovery made by Turing in 1952. He designed equations for reaction-diffusion systems that generate spatial concentration patterns starting from near-uniform initial distributions. His deduction was based on Fourier methods, that is on the analysis of destabilization of uniform distributions by concentration waves of certain wavelengths. Thus, normal chemical reactions in liquid media are able to generate concentration patterns.

Do such processes have biological significance for morphogenesis? To answer this question it is necessary to explore the conditions for pattern formation in molecular terms and, most importantly, to explain the fascinating and challenging self-regulatory features of developing biological systems, such as proportion regulation – the adaptation of sizes of a part to sizes of the whole. With these aims in mind, in the 1970s Hans Meinhardt and I proposed a theory of pattern formation based on two concepts: autocatalytic activation and lateral inhibition (Gierer and Meinhardt 1972). Our starting point was a line of thought originally introduced into the field of pattern recognition (Hartline *et al* 1956; Kirschfeld and Reichardt 1964). Let us draw an egg (figure 1a). Is our drawing a true representation of an egg? No, it is the contour of an egg. The image of a real egg on the retina looks different (figure 1b). To obtain the contours, the local intensities projected onto the retina are processed there by local activation in conjunction with an inhibitory effect extending into the environment of activation. Inside

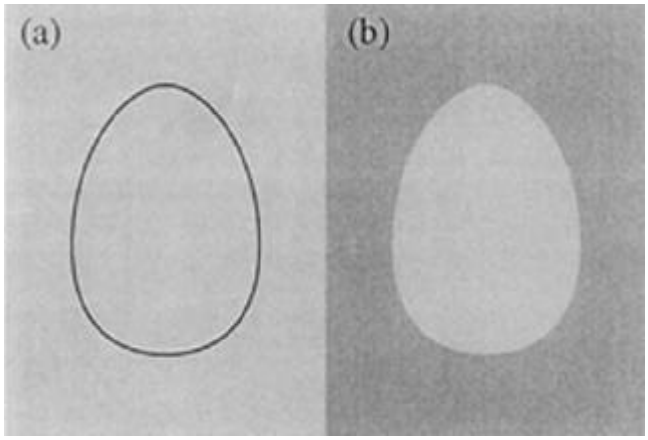

Figure 1. Lateral inhibition in pattern recognition. Drawing an egg means drawing the contour of an egg (**a**), which is abstracted from the primary image of the egg on the retina (**b**) by mechanisms involving lateral inhibition.

the bright area, high inhibition cancels high activation; inside the dark area, low inhibition cancels low activation. But at the edge, high activation is not cancelled because the inhibition extending from the neighbouring dark area is low; therefore the edge is enhanced and the contour of the egg outlined. This mechanism is well established in both psychophysical and neurobiological terms, and it shows an intriguing general feature: it allows for the generation of striking patterns, starting with very shallow ones. For instance, our eye can recognize edges of areas that are very slightly less grey than the areas surrounding them.

We adapted and applied the inspiring principle of lateral inhibition in modified form to pattern formation in the course of development, now with molecules interacting and moving instead of neurons firing: our theory says that patterns are formed by local self-enhancing reactions, controlled, spatially limited, and disciplined by a wider ranging inhibitory effect, ranges being given, in the simplest case, by the mean distance between production and decay or removal of molecules. Then, within an initially near-uniform distribution, local activation is self-enhancing, but activation at some location can proceed only at the expense of deactivation elsewhere, so that a striking spatial pattern is formed. Non-linear interactions are required for the generation of reproducible, stable patterns. Power laws for the order of reactions can be introduced to analyse general conditions determining which type of systems would generate patterns and which would not; we have used one of the simplest ones as a model of the models, but the general conclusions would apply to other versions as well.

The regeneration of *Hydra* – this polyp was our favourite system for experimental work in the 1970s – is a particularly puristic example of *de novo* pattern formation for getting the logic straight (figure 2). Any section regenerates an animal with head and foot. The pattern formed is *oriented* by previous polar cues, but aside from that one single bit of information – deciding on orientation to the left or to the right – the pattern itself, namely a head-activating morphogenetic gradient, is newly formed. The basic features of this biological process are rather straightforward consequences of mechanisms based on autocatalysis and lateral inhibition (figure 3). The resulting pattern is self-regulating, and is a product of molecular interactions and movements within the initially near-uniform tissue – requiring no dolls within dolls, however well hidden. And this type of mechanism gives rise to the striking regulatory features that are so characteristic of biological development, ensuring reliability despite complexity: in particular, details of the initial conditions do not matter. Regeneration is possible, as are induction, inhibition and, in certain conditions, proportion regulation. Not only gradients, but also symmetrical and periodic distributions, whether stable or pulsing in time, can be generated in this manner. And the mechanisms are not restricted to molecular diffusion; any range-dependent signalling across cells and tissues will do. The autocatalysis-plus-lateral inhibition theory has been applied to modelling pattern-forming systems ranging from individual cells to vertebrate embryology (Meinhardt 1982; see Meinhardt and Gierer 2000).

How about pattern formation in systems with more than two variables? This, after all, is the biologically most likely case: feedback loops, for instance, consisting of a chain of reactions. Consider schemes of, say, seven or ten reactions. What is the criterion for pattern formation in these conditions? We might think of collecting those compounds that have autocatalytic effects and analysing them individually, but this leads nowhere. For instance, activation can result just from inhibition of inhibition, allowing for pattern formation even if there is not a single directly activating reaction. An approach that will prove more adequate is different: apply the lateral inhibition concept from the outset; begin the analysis by sorting the molecules involved into those with a short range on the one hand and those subject to a wider distribution (more generally, those exerting long-range effects) in the tissue on the other. Then, check whether

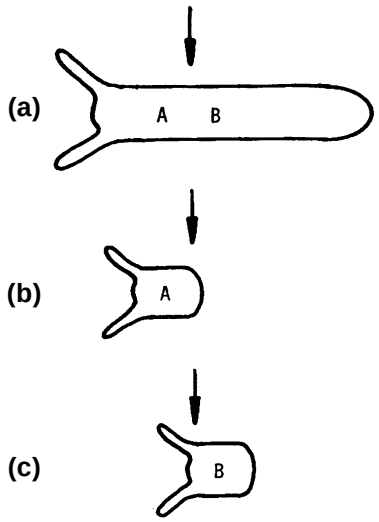

Figure 2. *De novo* pattern formation in *Hydra* regeneration. Any section cut from the gastric column of *Hydra* regenerates to form an animal with head and foot. Thus, the same part of the body column (arrow) may produce nothing (**a**), a foot (**b**), or a head (**c**), depending on whether and where the section is cut. It follows that no pre-existing local property of the tissue (such as the local concentration of a graded distribution determining the orientation of regenerates) can *per se* decide whether and where a head is formed; this can be decided only by the formation of a new morphogenetic gradient after the onset of regeneration.

the short-range subset, taken together, is in itself autocatalytic as a system, and whether the long-range subset prevents an overall autocatalytic explosion. In this way, the concept "pattern formation by the conjunction of activation and inhibition" can be generalized into multicomponent systems with activation and inhibition as features of subsystems rather than of individual substances (Meinhardt and Gierer 1980; Gierer 1981), while the basic regulatory capabilities that characterize biological development are maintained nonetheless. In the course of time a very substantial body of mathematical investigations, results and literature on pattern-forming reactions has evolved (Murray 1989; Harrison 1993).

What can we learn from such approaches of theoretical biology to understanding pattern formation? Certainly not what the molecular mechanisms are. These can be revealed only by developmental molecular genetics as advances in this field reveal more and more of the often rather complex reactions that may be involved. Further, it must be admitted that the principles of *de novo* pattern formation and its regulation cover only a small proportion of the experimental and theoretical studies on development. This is because once concentration patterns are there, they may function as morphogenetic fields and specify "positional information", directing subpatterns and the formation of secondary structures in a wide variety of ways. Two aspects, however, are worthy of note in this context: first, the involvement of *de novo* pattern formation is a logical *must* in the generation cycle, no matter how difficult it is to uncover it. The organism cannot contain its spatially organized progeny in disguise, by analogy with the Russian dolls within dolls. To provide even for only a few generations ahead, this would require that spatially organized rudiments be confined in volumes smaller than that of an atom, which would be inconsistent with the basic laws of quantum physics. Moreover, mechanisms of pattern generation by self-enhancement and lateral inhibition are such that distortions are corrected rather than enhanced. The inclusion of such error-correcting self-regulation in the generation cycle is an indispensable condition of high-fidelity reproduction of complex structures.

How does the autocatalysis-plus-lateral inhibition approach, which is primarily linked to biological deve-

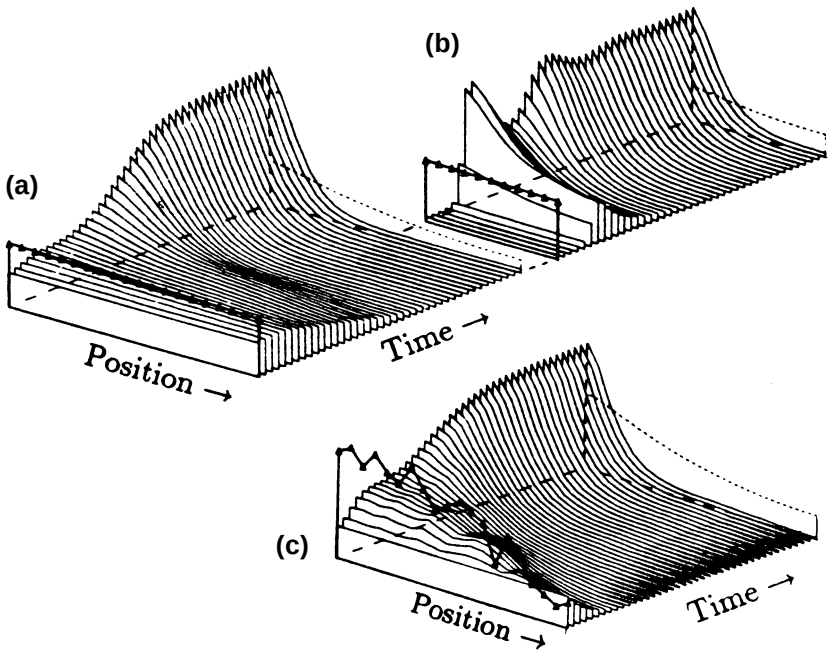

Figure 3. Pattern formation by autocatalysis and lateral inhibition explains characteristics of developing biological systems, in particular self-generation of patterns and their robustness against distortions, as shown by these computer simulations. (**a**) A striking pattern (that is a morphogenetic gradient specifying positional information, for instance for head formation in *Hydra*) is formed, starting from near-uniform initial distributions. (**b**) A section cut from (**a**) regenerates a new pattern. (**c**) Different, rather bizarre, initial conditions lead to essentially the same graded distribution as in (**a**). Thus, only the orientation of the pattern is determined by initial conditions, whereas its form is self-regulating, and the pattern-forming mechanism is capable of correcting distortions to a considerable extent.

lopmental regulation, relate to Turing's mathematical Fourier-type stability analysis, i.e. the detection of spatial wavelengths of distributions towards which the uniform distribution is unstable? The relation between the two approaches is by no means obvious. Only a close analysis has shown that the mathematical content of both lines of thought is very similar (Granero *et al* 1977; Gierer 1981). Depending on the perspective, we may then consider the principle of lateral inhibition as a specification of conditions for pattern formation by reaction-diffusion mechanisms. From a different perspective, however, reaction-diffusion mechanisms may be taken as a special case of the more general principles of structure formation by autocatalysis and inhibitory or depletion effects, within but also outside biology. Self-enhancement plays a major part in the formation of stars and galaxies, waves and clouds, crystals and dunes, and even in the formation of towns and in psychic processes: success generates success, and frustration generates even more frustration. What this reflects is not that all these processes are similar in mechanistic terms, but rather that mathematics is universal: theories of non-linear dynamics reveal the general principles involved and provide mental tools to help us to understand processes in very different fields.

4. Gene regulation, neural networks and the evolution of such general capabilities as cognition-based empathy

To a considerable extent, the neural network is formed under the control of the genes. There are far more neurons (10^{10} to 10^{11} in humans) and neural connections (more than 10^{13}) than there are genetic determinants (made up of fewer than 3×10^9 nucleotides). Therefore, genetic specifications of networks must involve some rules, and finding them is a real challenge for theoretical biology. Indirectly, genes determine the formation of spatio-temporal patterns, as discussed in the previous section, in both neural and non-neural tissues; such patterns, in turn, give rise to positional and directional cues to growing axons. Genes also determine and modulate features of growing axons themselves, resulting in growth cone navigation to their appropriate target and the formation of organized neural networks, to be refined subsequently by self-organization and learning. In simple cases, there is evidence for the involvement of algorithmic principles: topographic projections formed by guidance of retinal axons to their appropriate position in the brain appear to involve graded distributions of spatial cues analogous to longitudes and latitudes and their role in ocean navigation. Complex network features are more difficult to reveal, and the greatest challenge is the evolution and development of network properties giving rise to higher capabilities, such as human language, cognition and empathy – human capabilities that may have evolved relatively rapidly some 100,000 years ago.

It is the formal network of gene regulation that exerts control over the formation of the real network of neurons. The control of axonal guidance and targeting is mediated by the regulation of protein synthesis. This involves the role of non-coding sequences in the genome, acting as microprocessors, of a sort, in development. Selection operates on the neural network and its function, but variation occurs mainly at the level of networks of gene regulation. Will there ever be an algorithmic theory of the interrelation of these networks? Can we expect this to contribute to an understanding even of higher brain functions? We are still at the stage of highly speculative hypothesis, a long way from more sustained theories, but I think it is adequate to consider unbiased concepts rather than restricting ourselves to changing mainstream notions. Mainstream wisdom still is and certainly was, that genetic changes occur in many rather unspecific steps, and new brain capabilities may "emerge" in this way, perhaps with a strong contribution of self-organization in development. Logically, however, there may well be genetic changes that are highly improbable on an individual basis but are nonetheless relevant for evolutionary processes, involving large populations and thousands of generations. One may think of accidental, but highly specific duplications, transpositions and combinations of larger genomic sequences. Indeed, genetic changes of this type appear to have been "an important force in the evolution of the Human Genome" (Eichler 2001) though their functional implications are not yet known. Such changes may have small effects initially, but some of them may open up a truly novel direction for further evolution.

An – admittedly highly hypothetical – example we can use in discussing this issue is the evolution of empathy, the ability to imagine oneself in another's place and understand the other's feelings, desires, ideas and actions. Empathy depends on the integration of widely dispersed brain processes in which the "mirror neurons" of the primate frontal cortex appear to play an important part: they fire if an action is either performed or observed. Presumably, they subserve the ability to imitate others, but also to attribute intentions to others, in such a way that mental features of others are represented in one's own brain (Gallese and Goldman 1998; Rizzolati *et al* 2001). Beyond this, an important aspect of human empathy is the inclusion of plans, fears and hopes concerning different scenarios in an open future. I suggest that such cognition-based capabilities of human empathy have evolved in close conjunction with that of strategic thought.

Strategic thinking by comparative assessment of different scenarios appears to be characteristic for humans. It

depends on abstract representations of one's own possible future states in one's own brain to allow assessment of their emotional desirability; but it also benefits from the representation and emotional evaluation of possible states of others, facilitating anticipation of their behaviour. In neurobiological terms, this is best achieved if there are multiple representations of possible future states of others that are connected to one's own emotional centres in a manner similar to self-representations. For this reason, the evolution of human brains is assumed to have established capabilities of adequate representations with such linkages. Possibly, these have been initiated by duplications and transpositions of DNA segments that gave rise to one or a few accidental, but highly specific, novel combinations of pre-existing subroutines of gene regulation of the upper strata of the regulatory hierarchy, affecting large parts of the neural network (Gierer 1998). As a secondary effect of such evolving linkages, both the actual states and the future perspectives of others can elicit vicarious emotions, which may, in turn, contribute to altruistic behaviour.

These lines of thought are still rudimentary and highly speculative, but so are more mainstream ideas, especially if they draw on vaguely defined notions of "emergence". And, by now, there are also challenging new concepts in *favour* of roles of phenotypic and genetic *novelty* in human evolution. Thus, Povinelli and Preuss (1995) wrote on "important differences in how humans, great apes and other animals interpret other organisms", suggesting that "at some point in human evolution, elements of a new psychology were incorporated into existing neural systems"; and Pääbo (2001) strongly supports the idea that, perhaps some hundred thousand years ago, there were "one or few genetic accidents that made human history possible – a realization that provides us with a whole new set of philosophical challenges." A profound understanding of human evolution will require advances in developmental neural biology, in combination with explicit theoretical models. As discussed above, a crucial issue is the indirect relationship between the order of the network of gene regulation involved in neural development and the order of the corresponding neural network that underlies its functional capabilities.

There is one more aspect that I would like to mention in the context of general brain capabilities. According to the principles of biological evolution, such capabilities evolved because they led to increased fitness. And yet, general capabilities usually have potentials not involved in the causes of their origin. This applies to technical development: the invention of the first wheels was not motivated by anticipating the full range of further developments, such as that of the windmill, the screw propeller, and the turbo jet, and this may also apply to cognitive capabilities of the human brain. According to the concepts of evolutionary epistemology pioneered by Lorenz (1973), they evolved because they increased fitness immediately, and this then allowed for developments of their cognitive potentials beyond their immediate applications. Nevertheless, the range of generalizable cognitive capabilities expressed in the course of cultural history in human societies, encompassing even fields as far removed from the world of hunters and gatherers as quantum physics and theoretical biology, gives rise to further scientific and philosophical problems about the relation between human cognition and the order of nature, between mind and matter. . . . And perhaps this aspect of generalizable capabilities is worth keeping in mind in the discussion of my next and last topic: human consciousness as a feature of the human brain.

5. On brains, mind and consciousness: possible limits of decodability

Systems approaches contribute to the increasing scientific interest in what is perhaps the most challenging feature for human self-understanding, consciousness and the brain-mind relationship. Brain research is revealing more and more correlates between conscious experience and processes on the one hand and brain activities on the other – and often widely distributed neural activities. Functional nuclear magnetic resonance and other highly sophisticated techniques play their part in the search for 'NCC' – the neural correlates of consciousness.

The most thorough theoretical investigations in this field, such as those of Crick and Koch (1998), place emphasis on such correlates. The selection and integration of processes involved in a given conscious activity constitutes the 'binding problem', referring to the mechanisms by which activities in different parts of the brain that belong together in a given context are selectively linked in such a way that the most adequate interpretation of complex situations can be found, allowing for appropriate actions. In a specific model following similarly holistic lines of thought, it is proposed that a special neural network extends across the entire cortex, serving as a global workspace integrating perceptual, motor, memory, evaluative and attentional processing and allowing for appropriate solutions of complex non-routine tasks (Dehaene *et al* 1998). Only the results of integration are assumed to be conscious, and not the way in which they are reached.

Can we really expect to understand mental processes fully in neurobiological terms? In some sense the brain is a system for storing and processing information, analogous to a computer. For computers a general rule holds: every function that one can model in formal, mathematical terms can, in principle, be executed by a computer. The analogy between the brain and the computer

has its limits, but the nerve cell's capacity as a building block for information processing is greater, not smaller, than that of the digital yes/no switch of computers. This is why all the formally representable functions of the human brain are expected to be based on physical chemical processes in the neural network. This argument supports the idea that a scientific explanation is possible in principle; but it is not the explanation itself, which can only be achieved by neurobiology. And then there is the important question of whether *all* features and functions of the brain can be described formally in scientific terms. What can be formally represented can be seen in the research into artificial intelligence, and the list is impressive: object recognition, conceptional abstraction, memory, planning and the comparison of different strategies for future behaviour all feature on this list – in other words, most of what are considered to be the higher capacities of the brain.

How far will such investigations take us? Does it depend solely on our scientific efforts, or are there fundamental limits – limits not just concerning complex details, but also limits to intrinsic, interesting, and central features of consciousness? Many practising neurobiologists are adherents of the 'asymptotic' position on the resolution of the mind-brain problem, holding that progress in their field will be able to explain more and more about the relationship between brain and mind and that there are no limits in principle, even if some questions are too complicated and some calculations too exhausting for concrete solutions. Consciousness, for them, is a property of systems of nerve cells in the brain, much like supraconductivity is the property of systems of certain metal atoms at low temperatures; after all, we have learned how to understand supraconduction in terms of physics – why should this not be possible for consciousness as a system's property of neural networks in the brain? But such comparisons are not completely on track. Supraconduction is objectively defined – the electrical resistance is zero – while consciousness is not. Consciousness is primarily accessible through self-awareness and through communication of the awareness of others; it is doubtful whether in principle a complete formal or objective definition is possible. The difficulty, if not impossibility, of a complete definition of human consciousness in objective terms does not, however, justify regarding the mind-body problem as an illusionary issue that disappears upon careful conceptual deliberations. Undoubtedly, there *is* a relation between the mental and the physical, and the question of how and how far it can be decoded is a genuine scientific problem.

Let us select, as an example for states of consciousness, general behavioural dispositions for the future, that is, intentions of an individual for various patterns of behaviour depending on various scenarios for the future. Such dispositions are stored in our brain and are at least partially accessible to consciousness. Let us perform a thought experiment: let us suppose that we can simulate states and processes of the brain by a correspondingly constructed and programmed computer. Theoretically, we could calculate what would happen to a given initial brain state over time when it was exposed to certain exterior conditions, and what behavioural responses would result. One could now argue that in this way we could test all possible exterior conditions of the future, one after the other, with the final goal of determining the general behavioural dispositions that correspond to the initial state of the brain but are valid for different scenarios in an open future – and in this way of decoding the present brain state, objectively and exhaustively with respect to conscious states, at least insofar as they are related to behavioural dispositions.

But on second thoughts, we realize that this would not work; a procedure of this sort is seen to be impossible when we consider the finite nature of the world and take it seriously in epistemological terms (Gierer 1983). The world's intrinsically finite nature also limits the decidability of problems. Even a computer made up of the mass of the entire universe, encompassing some 10^{80} nucleons and running for 15 billion years – the age of the universe –, would still only be able to execute a finite number of operations (some 10^{40} per nucleon in 15 billion years) – a very liberal upper limit would be 10^{120}. But numbers of this huge magnitude do occur even in everyday problems as the number of *possibilities*. The number of *possible* letters with various contents, even if each were only a few pages long, is much larger. The same holds true for the number of possible future physical states that a particular behavioural disposition may apply to; and the number of different *possible* behavioural dispositions is also so large that they certainly could not be checked one after the other in a finite decision-making process to find out which dispositions correspond to a given physical state of the brain. What are the implications of this thought experiment for the potential range and limits of psychophysics in general? Of course, in any field of science it is possible to discover many relations, rules and laws applying to widest domains. This also holds true for research on consciousness; but there will presumably be no *general* procedure for decoding brain states with respect to all mental states. It is more likely that some essential aspects of the brain-mind relationship are not fully resolvable by finite analysis.

As for the aspects of consciousness that a scientific theory may not be able to fully encompass, only more or less educated guesses are possible. We can find hints in certain results of mathematical decision theory, which show a self-referential characteristic: the internal consis-

-logical systems cannot be decided by their own means. By analogy, the characteristic properties of consciousness, like the generation of behavioural dispositions, are also self-referential. We appear in our own memories, fears and hopes, desires and plans – as we are, or as we believe ourselves to be, or as we wish to be seen by others, as we want or do not want ourselves to become, and as we see our past and our future possibilities. Behavioural dispositions are influenced by these 'self-images', which of course do not represent concrete spatial conceptions but are rather abstract representations of features of the individual in his or her own brain. Perhaps these multiple 'self-images' belong to the aspects of consciousness that cannot be fully determined by analysis of the physical state of the brain.

To sum these considerations up, the applicability of physics to the brain together with the unique correspondence of mental states to physical states of the brain does not mean that all behavioural dispositions can be deducted from the physical state of the brain in a finitistic process. Rather, we have reason to believe (though as yet no proof) that there are limits to the decodability of brain states with respect to mental states. According to everything that we know, the brain follows the same physical laws as do machines; but a machine that we were capable of understanding could not do everything a human can do, and if a machine that could do everything a human can do existed it would be impossible for us to fully understand it. If we know the mental state of a fellow human as expressed in language and gestures, we may then know far more than it would be possible to know through a purely physical analysis of her or his brain, however far reaching.

6. Range and limits of science

Let me compare the present epistemological and psychological state of the art in neuroscience with that in physics and mathematics some hundred years ago (figure 4). Up to the 1920s, all events in space and time were assumed

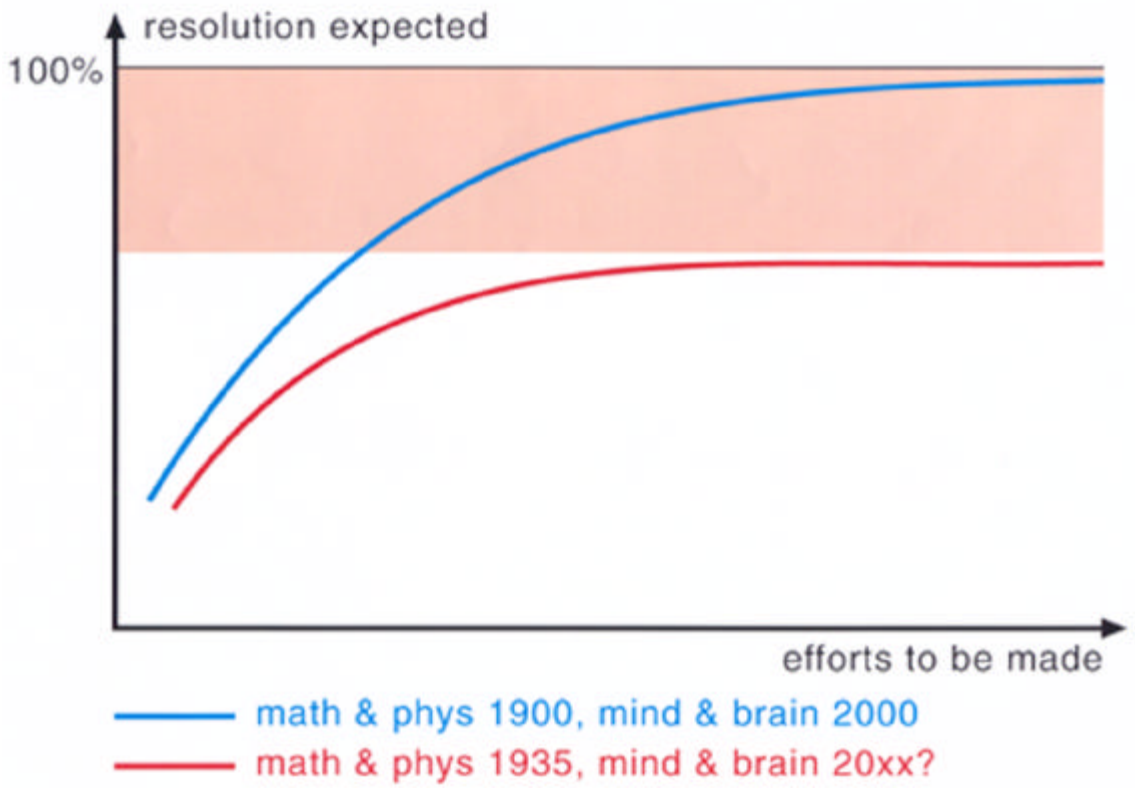

Figure 4. Scope and limits of science. Early in the twentieth century, physics and mathematics were still expected to resolve all well-stated problems in their fields asymptotically (blue line). By 1935, Heisenberg's uncertainty and Gödel's undecidability theorems had demonstrated that there were fundamental, unsurpassable limits to scientific determination and decidability (red line). Around 2000, many neurobiologists now embrace blue-line optimism, assuming that all major problems of brain research will eventually yield, in contrast to the view, shared by myself, that such problems as the brain-mind relationship might not be fully amenable to resolution by scientific means (red line).

to be fully calculable and predictable by the laws of physics, and it was expected that all interesting physical problems would be asymptotically resolved in the course of time: there were ultimately no limits. Then, the development of quantum mechanics demonstrated that some well-stated problems were amenable to scientific resolution while others were not. Energy states of stable material systems can be determined very precisely, depending on the effort involved; in this way, for instance, we have come to understand chemical bonds very well. But there are also questions for which there are no definitive answers no matter how much effort is expended on them. Predicting individual events at the atomic level is not precisely possible, regardless of the effort involved in measurement and calculation, and this impossibility is itself a law of physics! Measurements or observations inevitably interfere with the states that are to be measured. This is the essence of the famous "uncertainty principle" of quantum physics that Heisenberg discovered in 1927.

How about mathematics? Up to around 1930, a formal system of mathematics and logic, allowing for the proof of its internal consistency – the proof that contradicting statements cannot possibly arise within the system – was still seen as one of the goals of mathematics, and Hilbert claimed that there is no such thing as an unsolvable problem (see Reid 1970). Then in the 1930s Gödel, and also Turing, discovered the theorems of undecidability already alluded to above: the consistency of logical systems, except very primitive ones, cannot be proven by using their own means. Within every such system of formal thought and calculation there are questions that are undecidable for fundamental mathematical reasons.

It is remarkable that both physical uncertainty and mathematical undecidability are limitations related to self-referential operations: measurements affecting measurements in physics; logic applied to logic in mathematics. This would also hold for limits of a scientific theory of consciousness; it would involve consciousness of consciousness. Because of the dynamic development of current neuroscience, it is not surprising that a mainstream view of brain scientists is that all genuine scientific problems can be solved asymptotically in the long run. This view is analogous to that of physicists and mathematicians early in the 20th century. Will it persist with respect to brain science, say through 2030 or 2100? Like others, I subscribe to the educated guess that there will also be fundamental limitations to the scientific resolution of the brain-mind relation, particularly with regard to its self-referential aspects.

It is precisely our knowledge about the limits of knowledge as it evolves in different fields of science that shows us why scientific knowledge, despite its unambiguous content with respect to spatio-temporal processes and laws, is and remains ambiguous at the meta-theoretical level. We are, in our physical selves and our thoughts, an inseparable part of the world that we would like to get knowledge of. The basic limitations of knowledge are concerned with the relationship between the order of nature and human cognition, and they are linked in this way with fundamental questions of man's image of himself and the universe. It is for this reason that, in contrast to many ideas that existed in the nineteenth century, modern science is capable of being, and needs to be, interpreted on philosophical, cultural and religious levels and is consistent with more than one such interpretation, though of course not all. The ancient Greek philosophers have bequeathed us building blocks for possible interpretations – logos, number, spirit and matter. Non-European cultures provide elements of their own for metatheoretical interpretation. The interpretation itself, then, is the task of the present, a task for science, art, and the humanities, even if the languages spoken in different cultural sections are often quite different from one another.

References

Crick F and Koch C 1998 Consciousness and neuroscience; *Cereb. Cortex* **8** 97–107

Dehaene S, Kerszberg M and Changeux J P 1998 A neuronal model of a global workspace in effortful cognitive tasks; *Proc. Natl. Acad. Sci. USA* **95** 14529–14534

Eichler E E 2001 Recent duplication, domain accretion and the dynamic mutation of the human genome; *Trends Genet.* **17** 661–669

Gallese V and Goldman A 1998 Mirror neurons and the simulation theory of mind-reading; *Trends Cognitive Sci.* **2** 493–501

Gierer A and Meinhardt H 1972 A theory of biological pattern formation; *Kybernetik* **12** 30–39

Gierer A 1981 Generation of biological patterns and form: Some physical, mathematical and logical aspects; *Prog. Biophys. Mol. Biol.* **37** 1–47

Gierer A 1983 Relation between neurophysiological and mental states: Possible limits of decodability; *Naturwissenschaften* **70** 282–287

Gierer A 1998 Networks of gene regulation, neural development and the evolution of general capabilities, such as human empathy; *Z. Naturforsch.* **53C** 716–722

Granero M I, Porati A and Zanacca D 1977 A bifurcation analysis of pattern formation in a diffusion governed morphogenetic field; *J. Math. Biol.* **4** 21–27

Harrison L G 1993 *Kinetic theory of living pattern* (Cambridge: Cambridge University Press)

Hartline H K, Wagner H G and Ratliff F 1956 Inhibition in the eye of limulus; *J. Gen. Physiol.* **39** 651–673

Kirschfeld B and Reichardt W 1964 Die Verarbeitung stationärer optischer Nachrichten im Komplexauge von Limulus; *Kybernetik* **2** 43–61

Lorenz K 1973 *Die Rückseite des Spiegels. Versuch einer Naturgeschichte menschlicher Erkenntnis* (München: Piper)

Meinhardt H and Gierer A 1980 Generation and regeneration of sequences of structures during morphogenesis; *J. Theor. Biol.* **85** 429–450
Meinhardt H 1982 *Models of biological pattern formation* (London: Academic Press)
Meinhardt H and Gierer A 2000 Pattern formation by local self-activation and lateral inhibition; *BioEssays* **33** 753–760
Murray J D 1989 *Mathematical biology* (Heidelberg: Springer)
Pääbo S 2001 The human genome and our view of ourselves; *Science* **291** 1219–1220
Povinelli D J and Preuss T M 1995 Theory of mind. Evolutionary history of a cognitive specialization; *Trends Neurosci.* **18** 418–424
Reid C 1970 *Hilbert* (Heidelberg: Springer)
Rizzolati G, Fogassi L and Gallese V 2001 Neurophysiological mechanisms underlying the understanding and imitation of action; *Nature Rev. Neurosci.* **2** 661–670
Roe S A 1982 *Matter life and generation. 18th century embryology and the Haller–Wolff debate* (Cambridge: Cambridge University Press)
Trembley A 1744 *Mémoires pour servir à l'histoire d'un genre de polypes d'eau douce à bras en forme de cornes* (Leiden: Jean and Herman Verbeek)
Turing A 1952 The chemical basis of morphogenesis; *Philos. Trans. R. Soc. London Ser. B* **237** 37–72
Wolff C F 1759 *Theoria generationis* (Halle)