Networks of Gene Regulation, Neural Development and the Evolution of General Capabilities, Such as Human Empathy*

Alfred Gierer

Max-Planck-Institut für Entwicklungsbiologie, Spemannstrasse 35/IV,
D-72076 Tübingen, Germany

Z. Naturforsch. **53c,** 716–722 (1998); received April 14, 1998

Gene Regulation, Brain Evolution, Neural Network, Empathy, Evolution of Altruism

A network of gene regulation organized in a hierarchical and combinatorial manner is crucially involved in the development of the neural network, and has to be considered one of the main substrates of genetic change in its evolution. Though qualitative features may emerge by way of the accumulation of rather unspecific quantitative changes, it is reasonable to assume that at least in some cases specific combinations of regulatory parts of the genome initiated new directions of evolution, leading to novel capabilities of the brain. These notions are applied, in this paper, to the evolution of the capability of cognition-based human empathy. It is suggested that it has evolved as a secondary effect of the evolution of strategic thought. Development of strategies depends on abstract representations of one's own possible future states in one's own brain to allow assessment of their emotional desirability, but also on the representation and emotional evaluation of possible states of others, allowing anticipation of their behaviour. This is best achieved if representations of others are connected to one's own emotional centres in a manner similar to self-representations. For this reason, the evolution of the human brain is assumed to have established representations with such linkages. No group selection is involved, because the quality of strategic thought affects the fitness of the individual. A secondary effect of this linkage is that both the actual states and the future perspectives of others elicit vicarious emotions, which may contribute to the motivations of altruistic behaviour.

Introduction

Most general capabilities of the human brain, such as language, mental self-representation, long-term strategic thought and cognition-based empathy may have evolved quite recently in the evolutionary time scale, perhaps only some hundred thousand years ago. Mainstream explanations are mostly gradualistic. For instance, the evolution of brain capabilities is often discussed in relation to increases in brain size. Indeed, the theories of bifurcation and dynamic instabilities reveal how distinct features can gradually develop in many small steps without specific initiations, and it is conceivable that quantitative extensions of the neural network allow for the emergence of additional functions. However, there is also the possibility that new capabilities originate from rare specific initiating genetic changes that open up an innovative direction for further development. In the history of technology, both types of innovation are documented. For instance, in 19th century ocean traffic, the fast ocean clippers were mainly the result of quantitative changes in the construction of sailing vessels – slim hulls, huge sails – whereas steamship development was dominated by qualitative innovations: the combination of Watt's steam engine with sea vessels; the replacement of the wooden by the iron hull, and of the paddle wheel by the screw propeller. Analogous distinctions may be applicable to biology: though it is not logically necessary to invoke qualitative initiating events to explain the evolution of novel features, they may be more efficient in some cases, and evolution prefers efficient pathways. This may apply, in particular, to the evolution of specifically human brain capabilities, despite the intellectual appeal of theories emphasizing processes of 'self-or-

* This communication is a contribution to the workshop on "Natural Organisms, Artificial Organisms, and Their Brains" at the Zentrum für interdisziplinäre Forschung (ZiF) in Bielefeld (Germany) on March 8–12, 1998.

Reprint requests to Prof. A. Gierer.
Fax: (07071) 601448.

0939–5075/98/0700–0716 $ 06.00 © 1998 Verlag der Zeitschrift für Naturforschung, Tübingen · www.znaturforsch.com. D

ganization' as indirectly related to, and as remote from specific genetic determinants as possible.

In this article I would like to discuss the 'initiation' hypothesis in relation to networks of gene regulation, neural networks and their function, with an emphasis on cognition-based empathy as one of the specifically human brain capabilities.

A Main Substrate of Brain Evolution Is the Network of Gene Regulation Involved in Neural Development

The evolution of capabilities of the neural network is to be related to the main mechanisms of evolutionary changes at the DNA level as revealed by molecular genetics (see Alberts *et al.,* 1997). The genome consists of nucleotide sequences coding for proteins, as well as non-coding sections. Within the latter, specific regulatory sequences of the DNA occur which bind specific regulatory proteins. In turn, these bound proteins contribute, in a combinatorial fashion, to the regulation of transcription of nearby coding sequences, thereby controlling the synthesis of the corresponding protein. Certain regulatory proteins affect the production of other regulatory proteins; this hierarchical network of gene regulation is capable of assuming different stable states characterized by different combinations of regulatory proteins, allowing for cell differentiation in the course of the spatiotemporal development of the organism. The spatial order is organized, in part, by the response of the cells to position-dependent morphogenetic signals, such as graded distributions of morphogens. Many processes not discussed here are involved in developmental regulation, pattern formation and morphogenesis, including further translational as well as transcriptional control mechanisms, induction, signal transduction, effects of cofactors, and cytoskeleton dynamics. However, a central mechanism of the genetic control of development is the control of expression of a large number of proteins by the binding of regulatory proteins to the regulatory parts of the genes (see Arnone and Davidson, 1997). The latter can be regarded as microprocessors of information, responding in a combinatorial fashion to the set of regulatory proteins that is specific for a particular cell type, developmental stage, and position in the organism (see Alberts *et al.,* 1997; Gierer, 1973).

This is relevant, in particular, for spatiotemporal regulation of the expression of proteins involved in axonal guidance and targeting: though the relation between genes and brain functions is an indirect one, it is sets of regulatory genomic sequences that are instrumental in implementing appropriate rules of connectivity within the developing neural network. The network laid down in this way is then refined and altered by internally generated, as well as externally driven, activity-dependent processes (for review, see Gierer and Müller, 1995). The resulting network properties, in turn, determine functional capabilities of information processing and set the stage for further modifications elicited by experience, including learning, in different phases of the individual organism's lifespan.

In the course of evolution, segments of the genome are duplicated, recombined and transposed to new positions at widely different rates, to be modified and fine-tuned by subsequent mutations. Duplications of coding sections of genes give rise to families of related proteins. Duplications and translocations of non-coding sections may affect the expression of given proteins in new contexts of development. In this way, new directions of evolution may be initiated; though immediate phenotypic effects on fitness are most probably small, they may gradually become larger by way of further mutations in the direction of the innovative evolutionary pathway.

In line with these concepts, it is proposed that the *evolution* of *new* mental capabilities related to pre-existing ones is initiated, at least in some cases, by duplication of sections of DNA containing distinct sets of regulatory DNA sequences involved in brain development, in particular of those belonging to the upper strata of hierarchies of developmental regulation, and their introduction into a new context within the genome, followed by further modification. In populations of millions, in the course of many thousands of generations each regulatory section has a chance to be introduced into many other contexts of the genome. In such a way, rare combinations with positive effects on genetic fitness may have a chance to initiate a new direction of further evolution. Widely distributed features of the neural network could be affected in this way, and novel combinations of subroutines of gene regulation may initiate the evolution of new features of the neural network, leading to novel algorithmic capabilities on the basis of existing ones.

Evolution of Cognition-Based Empathy: Some Major Issues

These notions will now be applied to the evolution of human cognition-based empathy. Little is known about the neurobiological basis of this capability, and the concepts and hypotheses discussed below cannot do justice to the subtle psychology of empathic emotions; they will be limited to the following aspects:

1. Human empathy is based on specifically human cognitive capabilities, and is a source of altruistic behaviour.
2. The evolution of empathy depends on effects giving rise to an increase in fitness; this is postulated to occur by upgrading the quality of strategic thought.
3. Strategic thought depends on the representation of selves and of others in the brain, capabilities that are encoded in the human genome.
4. The evolution of the capability of empathy implies the linkage of representations of others with one's own emotional centres.
5. This linkage may have been initiated, in evolution, by specific combinations of subroutines of gene regulation.

Cognitive Capabilities as a Basis of Human Empathy

Empathic, 'vicarious' emotions of a person sharing, in an attenuated and modified way, the emotional states of another person, are not confined to immediate responses to present states of others, such as overt pain; they include cognition-based participation in the future perspectives of others, and their emotional correlates, such as anxiety and hope. The development of empathy in childhood and its relation to cognitive development is widely covered by psychological research (e.g. Hoffman, 1981; Eisenberg, 1986; Miller *et al.,* 1991). In the earliest stages there are immediate responses to certain distress signals. Recent research demonstrated pro-social behaviour, presumably involving perspective taking, by very young children from 14 months of age. Studies by Bischof-Köhler (1989) suggest correlations between self-representations, expressed as the child's capability of recognizing him- or herself in a mirror, and the capability of empathy. Later, the child is able to realize that the internal states of others, including their knowledge and emotions, differ from his or her own (Wimmer and Perner, 1983). This enables him or her to infer false beliefs in others. Eventually, the capability is developed to allow assumption of the roles of others, the child then imagining, playing and acting in accordance with their assumed internal states.

Whereas the mature capacity for empathy encompassing subtle information on the mental states of others is expressed only after a sequence of stages of cognitive and emotional child development, and the expression of empathy is dependent on individual character, education, personal experience and culture, the ability for cognition-based empathy as such is most probably encoded, in an abstract manner, in genetic determinants of the human brain. Whether and to what extent "theory of mind" capabilities exist in apes is difficult to detect. Chimpanzees' behaviours include, for example, food sharing and gestures of reconciliation. However, it is not easy to decide, on the basis of behavioural studies, whether an animal is able to infer mental states of others, such as intentions, desires, beliefs and knowledge. This is borne out, for instance, by a collection and controversial discussion of episodes suggesting deception in animals (Whiten and Byrne, 1988). Cheney and Seyfarth (1990) state that "chimpanzees, if not other apes recognize that other individuals have beliefs, but there is little evidence that chimpanzees recognize discrepancies between their own state of mind and the state of mind of others. They show little empathy for each other and they do not teach each other." In a recent review, Povinelli and Preuss (1995) conclude that "humans might have evolved a cognitive specialization in theory of mind forever altering their view of the social universe." It appears likely that cognition-based empathy is a distinct feature that has arisen in the course of human evolution, presumably in its later stages. Empathy is capable of inducing altruistic and cooperative human behaviour, though induction does not necessarily occur and is itself subject to motivation and learning (Miller *et al.,* 1991). This raises the question as to why and how the capability of human empathy could have evolved.

Evolution generally favours egoistic behavioural dispositions directed toward higher reproduction rates for the corresponding genetic trait; cooperativity and other forms of altruistic behaviour call for specific explanations. One of them is "inclusive fitness" (Hamilton, 1964; Maynard Smith, 1964) accounting for cooperation between relatives sharing many of their genes by common descent: a gene encoding behavioural dispositions for cooperation may spread within the population even if the encoded behavioural traits reduce the fitness of some of its carriers for the sake of others, that is of close relatives. Since kinship is often statistically correlated with common upbringing, cooperation based on "familiarity" may include non-relatives if they share features of socialization. Then there is the theory of "reciprocal altruism" (Trivers, 1971) which is sustained by analysis of evolutionary stability of behavioural dispositions toward cooperation on the basis of game theory (Axelrod and Hamilton, 1981). Cooperation at the expense of one's own fitness at a given time may be motivated by the expectation of cooperation from the partner (with oneself or one's close relatives) in the future. In human populations, "indirect reciprocity" may be based on probabilistic expectations summarizing information on a member of the group in the form of his or her reputation (Alexander, 1987).

Altruistic behaviour induced by empathy, however, cannot be generally explained in terms of kin selection or reciprocity. It is not restricted to kin, although it may be stronger among relatives than among others; and it can occur in many situations in which reciprocity would not be expected. Therefore, empathy, especially its human, cognition-based form, appears to be a source of human altruism that calls for an independent explanation in terms of evolutionary theory. I would like to suggest that the capability of human cognition-based empathy may have evolved in close association with the evolution of strategic thought that, in turn, contributes to individual fitness and may compensate for reductions in fitness by dispositions of altruism.

Mental Representations of Selves and Others, Strategic Thought, and Cognition-Based Empathy

Efficient strategic thinking depends on adequate representations of spatio-temporal features of the outside world in the human brain; in addition, it requires good "self-representations", of actual as well as possible "selves". The latter are to be assessed for emotional quality in case they should become real in the future, so that behavioural strategies can be developed for optimizing one's own emotional wellbeing. Self-representations are a basic feature of human consciousness. They are systems' properties of the physical state of the corresponding brain; and yet, it may be impossible to develop a complete algorithmic theory of the mind-brain relation (Gierer, 1983). Dynamic concepts on "multiple selves" including "possible selves" are sustained by an impressive body of psychological studies (Markus and Wurf, 1987; Markus and Nurius, 1986).

Representation of selves requires specific capabilities of the neural network for analysing its own content and abstracting, processing and storing past, present and possible future feature combinations of oneself, the last to be evaluated with respect to emotional desirability. These capabilities may have evolved by generalizing modes of analysis originally concerned with forecasting external events, towards anticipating possible future mental and emotional states of the brain itself.

However, optimizing strategic thinking requires not only representations of possible states of one's own features; it also depends on a good assessment of the possible behaviour of others. Learning from past experience alone could lead to predictions concerning their future behaviour, but such predictions would be relatively time-consuming and liable to error compared with an approach that makes additional use of the essential biological similarities of human brains: the relationship between personal situations and emotions is expected to be generally similar for others and oneself. Therefore, representations of *others* that include information on *their* mental features allow us to predict and assess *their* possible actions. Adequate representations of others include aspects of how *they* see themselves, and others (including oneself) and of how they emotionally evaluate actual and possible situations.

For these reasons, an important contribution to the capability of strategic thought appears to be the linkage of representations of others including their present as well as their potential future mental states to one's own emotional subsystems. Emotional states of others and their behavioural consequences can then be assessed and integrated into one's own strategic thought for the generation of behavioural dispositions. Such assessment, based in the first approximation on the similarity of human brains, can be modulated and improved by the incorporation of knowledge, acquired by learning, of differences in individual character, personal experience and socio-cultural background.

The postulated linkages of representations of others with one's own emotional centre not only improve the quality of strategic thought with respect to predicting the behavioural responses of others; they also support the capability of learning from the experience of others. These effects increase individual fitness. Therefore, evolution tends to establish such linkages. Referring to the hypothesis of a key role of relatively rare genetic changes in initiating new directions of evolution, it is suggested that the evolution of the capability for representing the features of others in the brain might have originated from duplications and subsequent variations of genomic subroutines encoding the neural connectivity underlying the capability of self-representation. In order to give rise to cognition-based empathy, the representations of others must maintain or evolve linkages with one's own emotional centres by neuronal connections that resemble those between self-representations and one's own emotional centres; the capability of cognition-based empathy combining perspective taking and vicarious emotions could thus be encoded.

Most probably, the representations of selves, as well as of others, are widely dispersed rather than localized in the neural network. Prominent roles may be played by the prefrontal areas of the cerebral cortex, which are concerned with planning, goal-directed behaviour and other highly integrating functions involving novelty, complexity and temporal organization (Fuster, 1985; Goldman-Rakic, 1988), and by limbic structures, which are known to be central gateways for emotions and are extensively interconnected with the prefrontal areas of the cortex; the specific role of the latter in emotional assessment is strongly supported by recent evidence (Davidson and Sutton, 1995). The actual neurobiological correlates of representations of selves and others in this context are not yet known.

Irrespective of the hypothetical details of such mechanisms, a secondary effect resulting from the connection of representations of others with one's own emotional centres in the brain is that others' joy, suffering, moods and, further, more subtle states of mind, both in present, actual and in future, possible situations, are reflected in one's own emotions. This, in turn, may lead to behaviour aimed at relieving pain and achieving wellbeing for others, ultimately for the sake of one's own positive emotions; empathy induces trans-kin cooperativity and other forms of pro-social behaviour.

The Evolutionary Stability of Empathy and the Gene-Culture Relationship

In terms of evolutionary theory cognition-based human empathy increases individual fitness by upgrading the quality of strategic thought while the secondary effects of empathic altruistic behaviour reduce fitness. Evidently, the latter effect would prevent the evolution of empathy if it were to lead to a net decrease of individual fitness. Therefore, it cannot have been too pronounced from the outset, and it may have been reduced in the further course of evolution by attenuation of the linkage between representation of others in one's brain and one's own emotional centres. However, there may be limits to attenuation if the positive effects for strategic thinking are to be maintained, and it may be that there are no evolutionary pathways of genetic changes in the neural network that would reduce empathy to zero within time spans consistent with human evolution.

In addition, cultural mechanisms may contribute to the stabilization of behaviour expressing empathy. The generation and transfer of information by cultural means sometimes imposes different, and often less stringent, constraints than the conditions of evolutionary stability of purely genetic evolution, especially with regard to features related to group selection (Boyd and Richerson, 1990; Richerson and Boyd, 1992). This, in turn, suggests that moderate reductions of individual fitness resulting from cooperative behaviour induced by empathy might be compensated by cultural motives and indirect benefits, such as the conformity reward posited by Richerson and Boyd. Such possible co-evolutionary aspects of empathy need further studies.

The Case for Distinct Genetic Changes Initiating the Evolution of Human Capabilities, Such as Cognition-Based Empathy

The proposed concept of self-representation and representation of others in the human brain as a basis for strategic thought as well as for empathy is consistent with recent (off-mainstream) ideas (Povinelli and Preuss, 1995) 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." It is emphasized that only capabilities for empathy are genetically encoded, whereas culture then determines in which contexts and for what purposes these capabilities are activated and expressed.

The hierarchical and combinatorial features of the network of gene regulation involved in the development of the neural network suggest that a limited number of distinct genetic changes affecting a few genomic subroutines may initiate an innovative evolution of algorithmic capabilities of the brain, such as cognition-based empathy. Such mechanisms are consistent with the notion that general human capabilities may have evolved rather recently within a short interval on the evolutionary time scale. It is not claimed that the assumption of rare initiating events is logically required to explain the evolution of new capabilities of the brain, but it is suggested that this is a most efficient way of generating novel algorithmic capabilities in a stage of evolution where highly developed capabilities are already available for new combinations and modifications. Our hypothesis on the role of rare initiating events is in full accord with *phenotypic* gradualism: the initial effects on fitness of the primary genetic changes forming a new direction of evolution were presumably small, to be enhanced only in many subsequent accumulating mutational steps.

A more profound understanding will require advances in developmental neurobiology, in combination with explicit theoretical models. 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.

Acknowledgements

I am much indebted to Prof. Holk Cruse and Prof. Gerhard Vollmer for their critical reading of the manuscript and their most valuable suggestions.

Alberts B., Bray D., Lewis J. , Raff M., Roberts K. and Watson J. D. (1994), The Molecular Biology of Cell, 3rd edition. Garland Publishing Inc., New York, pp. 385–395, 417–432, 1119–1130.

Alexander R. D. (1987), The Biology of Moral Systems. Aldine de Gruyter, New York.

Arnone M. J. and Davidson E. H. (1997), The hardwiring of development: organization and function of genomic regulating systems. Development **124,** 1851–1864.

Axelrod R. and Hamilton W.D (1981), The evolution of cooperation. Science **211,** 1390–1396.

Bischof-Köhler D. (1989), Spiegelbild und Empathie. Huber, Bern, Stuttgart, Toronto.

Boyd R. and Richerson P. J. (1990), Group selection among alternative evolutionary stable strategies. J. Theoret. Biol. **145,** 331–342.

Cheney D. L. and Seyfarth R. M. (1990), How Monkeys See the World. University of Chicago Press, Chicago, p. 254.

Davidson R. J. and Sutton S. K. (1995), Affective neuroscience: the emergence of a discipline. Curr. Opin. Neurobiol. **5,** 217–224.

Eisenberg N. (1986), Altruistic Emotion, Cognition and Behaviour. Lawrence Erlbaum, Hillsdale, New Jersey.

Fuster J. M. (1985), The prefrontal cortex and temporal integration. In: Cerebral Cortex Vol. **4** (Peters A., Jones E. G., eds.). Plenum Press, New York, 151–177.

Gierer A. (1973), Molecular models and combinatorial principles in cell differentiation and morphogenesis. Cold Spring Harbor Sympos. Quant. Biol. **XXXVIII,** 951–961.

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

Gierer A. and Müller C. M. (1995), Development of layers, maps and modules. Curr. Opin. Neurobiol. **5,** 91–97.

Goldman-Rakic P. S. (1988), Changing concepts of cortical connectivity: Parallel distributed cortical networks.

In: Neurobiology of Neocortex (Rakic P., Singer W., eds.). Wiley, Chichester, pp 177–202.

Hamilton W. D. (1964), The genetic evolution of social behaviour. J. Theoret. Biol. **7,** 1–52.

Hoffman M. L. (1981), Is altruism part of human nature? J. Pers. Soc. **40,** 121–137.

Markus H. and Nurius P. (1986), Possible selves. Amer. Psychologist **41,** 954–969.

Markus H. and Wurf E. (1987), The dynamic self-concept: A social psychological perspective. Ann. Rev.-Psychol. **38.** 229–337.

Maynard-Smith J. (1964), Group selection and kin selection. Nature **201,** 1145–1147.

Miller P. A., Bernzweig J., Eisenberg N. and Fabes R. A. (1991), The development and socialisation of prosocial behaviour. In: Cooperation and Prosocial Behaviour (Hinde R. A., Groebel J., eds.). Cambridge University Press, Cambridge, 54–77.

Povinelli D. J. and Preuss T. M. (1995), Theory of mind: Evolutionary history of a cognitive specialization. Trends Neurosci. **18,** 418–424.

Richerson P. and Boyd R. (1992), Cultural inheritance and evolutionary ecology. In: Ecology, Evolution and Human Behaviour (Schmitt E. A., Winterhalder B., eds). De Gruyter, pp. 61–92.

Trivers R. L. (1971), The evolution of reciprocal altruism. Quart. Rev. Biol. **46,** 35–57.

Whiten A. and Byrne R. W. (1988), Tactical deception in primates. Behav. Brain Sci. **11,** 233–273.

Wimmer H. and Perner J. (1983), Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. Cognition **13,** 103–128.