Contents

Impressum

Evolution and Cognition: ISSN: 0938-2623 **Published by:** Konrad Lorenz Institut für Evolutions- und Kognitionsforschung, Adolf-Lorenz-Gasse 2, A-3422 Altenberg/Donau. Tel.: 0043-2242-32390; Fax: 0043-2242-323904; e-mail: sec@kla.univie.ac.at; World Wide Web: http://www.univie.ac.at/evolution/kli/ **Editors:** Rupert Riedl, Manfred Wimmer **Layout:** Alexander Riegler **Aim and Scope:** "Evolution and Cognition" is an interdisciplinary forum devoted to all aspects of research on cognition in animals and humans. The major emphasis of the journal is on evolutionary approaches to cognition, reflecting the fact that the cognitive capacities of organisms result from biological evolution. Empirical and theoretical work from both fields, evolutionary and cognitive science, is accepted, but particular attention is paid to interdisciplinary perspectives on the mutual relationship between evolutionary and cognitive processes. Submissions dealing with the significance of cognitive research for the theories of biological and sociocultural evolution are also welcome. "Evolution and Cognition" publishes both original papers and review articles. **Period of Publication:** Semi-annual **Price:** Annuals subscription rate (2 issues): ATS 500; DEM 70, US$ 50; SFr 60; GBP 25. Annual subscriptions are assumed to be continued automatically unless subscription orders are cancelled by written information. **Single issue price:** ATS 300; DEM 43; US$ 30; SFr 36; GBP 15 **Publishing House:** WUV-Universitätsverlag/Vienna University Press, Berggasse 5, A-1090 Wien, Tel.: 0043/1/3105356-0, Fax: 0043/1/3197050 **Bank:** Erste österreichische Spar-Casse, Acct.No. 073-08191 (Bank Code 20111) **Advertising:** Vienna University Press, Berggasse 5, A-1090 Wien. **Supported by Cultural Office of the City of Vienna, Austrian Federal Ministry of Science/ Transportation and the Section Culture and Science of the Lower Austrian State Government.**

Hubert Hendrichs

Accessing Animal Minds: Epistemological and Empirical Problems

Abstract

The paper outlines the origin and aims of an interdisciplinary project on the problem of animal minds, involving the cooperation of ethology, neurobiology, robotics, and philosophy. The contributions, although heterogeneous and reflecting different positions in these disciplines, nevertheless indicate that there is a common ground to be considered in future theoretical and empirical studies. The behavior of organisms is oriented and organized by a compound of highly integrated components of different dimension and rationality. In human beings some of these components are named cognition, emotion, and consciousness. Their evolutionary development can only be understood as a multilevel phenomenon, and by considering the heterogeneities in dimension and rationality, in structure and processing, and their increasing integration.

Key words

Animal minds, evolution of human mentality, cognition, emotion, consciousness.

SOME 20 YEARS AGO, AT the 15th International Ethological Conference held in Bielefeld, a senior ethologist, Don GRIFFIN, had proposed a round table on cognitive ethology, on awareness and intentionality in animals.[1] The faculty of biology of the recently founded university had only started work one year before with a very small number of scientists. I was given the task to organize the workshops and round tables at the conference. I remember that this proposal caused irritation among the organizers. No sober biologist was—at that time—expected to commit himself to such unscientific topics, if he wanted to keep his scientific reputation. That attitude has changed considerably—not completely—in the meantime. At an annual meeting of the Deutsche Zoologische Gesellschaft in Oldenburg a round table was held, entitled "Toward a biological concept of consciousness" (cf. PFANNENSTIEL 1996).

Without doubt interest has grown and great progress has been achieved in the last years, especially in neurobiology, investigating and documenting correlations between brain activity and behavioral states[2], and in primatology, investigating communication and cheating in apes and other primates (cf. PREMACK/PREMACK 1983; WHITTEN/BYRNE 1988; CHENEY/SEYFARTH 1990). Important contributions could in addition be recruited from cognitive psychology (cf. BIRBAUMER 1996; MARKOWITSCH 1992; BAARS 1997) and philosophy (AGAZZI 1981; PYLYSHYN 1984; SEARLE 1983, 1992; FODOR 1987; DENNETT 1987; DRETSKE 1988; DREYFUS/DREYFUS 1988; PENROSE 1989; RADNER/RADNER 1989; KIM 1996), providing valuable conceptual and epistemological tools. But in spite of this great interdisciplinary enthusiasm and effort[3] some appearingly simple (ontological) questions asked by Don GRIFFIN in the handout for his round table 20 years ago have remained unanswered:

- Do animals know what they are doing?
- Can communicative behavior be used to ask questions?
- What hypotheses can be formulated concerning animal awareness and intention?
- What observations or experiments could show, whether animals have internal images?

This indicates that something is seriously wrong or missing with the categories and conceptualizations used in the various disciplines concerned with systems—living and nonliving, human and nonhuman—which generate functionally adequate actions in situationally changing contexts, responding to various forms of stimulation. To the unanswered questions one may add:

- To what extent is observable behavior influenced by processes involving feelings, cognitive representations, beliefs, and
- what types of intentional states are required,
- do they have to be individually acquired or produced?
- What are the established empirical findings?

■ How can these processes be conceptualized, and
■ what are the criteria that allow us to ascribe representations and intentional states,
■ what are the necessary differentiations?

These questions concern a wide range of different types of scientific approaches, from the evaluation of empirical data to the epistemological analysis of argumentations, but they center around an area in between—and outside—the established scientific and philosophical disciplines: an area in which conceptualizations have to be developed that can be connected with specific concepts of several disciplines and can thereby form a bridge between these disciplines. Such bridges are required and have to be developed to find answers to the above questions. These conceptualizations have to be built with scientifically and philosophically acceptable tools and building blocks (although outside the discipline specific tools; this is a problem: what may be acceptable to one discipline may not to an other, and vice versa). They are concerned with the performance of structures and processes which are functional and goal-directed, while their steering and their intentionality and cognitive orientation may be very different. These kinds of differences pose unsolved problems, for example the difference between artificial and living intelligence, between machines and living systems, between the rights of animals and humans, and between animals at different levels of complexity.

The ultimate aim of such efforts will be something like an ontological evaluation of states, processes, and structures in systems, which may have a potential for producing states of consciousness and suffering, and which are relevant when considering the status of these systems—morally, legally, and otherwise. But this is not the foremost topic of our present volume. Questions concerning ontological—and ethical—states, such as:

■ what does it mean to be a living system, able to intend and to suffer,
■ do these qualities require a specific kind of treatment of the system,
■ how do specific requirements of such systems have to be considered in case of conflicting interests?

are not immediately relevant at this time. They are obviously very important, and it is a serious problem that there are no generally accepted answers to these questions, and that (just or fair) solutions have to be found without knowing the ontological status of the systems involved. But this fundamental knowledge is a goal that obviously lies ahead, albeit in the future, and the first steps on the road to this goal are impeded by a number of unsolved problems which have to be coped with first. They include conceptual and methodological, that is, epistemic and empirical, problems as addressed in the contributed papers. Although this faraway goal will not yet be a central issue of this volume, it is nevertheless an important background, providing an orienting horizon for its main task: discussing, evaluating, and connecting the different concepts proposed by different disciplines and possibly reflecting different positions within one discipline. Thus, the aim here is not yet to find answers to unanswered questions or solutions for unsolved problems but, much more modestly, a contribution to the following three interrelated tasks:

■ to gain a better transparency of the basic concepts underlying the various conceptualizations of the different disciplines, and
■ to look for possibilities of adequately associating these different conceptualizations, finding and describing possible links or bridges, and
■ thereby mapping out a field in which new questions may be asked and possibly answered.

Although relatively modest, this still is a daring enterprise.

Thus, the distant goal is to establish a solid base for a careful ontological characterization of living systems. This will be approached by describing and evaluating epistemological and empirical requirements and problems. One central aim is to draw attention to differences between animals of lesser and greater mental complexity, of fewer or more levels in the organization of orientation and intentionality. These differences are considered to be related to those between organisms and machines, between cognitive processes in real and in virtual spaces, and those in the degrees of freedom, individuality, personality, and dignity of objects. The careful semantic analysis of concepts and argumentations by cognitive philosophy is essential, but not sufficient. The neurobiological analysis of physiological and structural foundations is very important, but not sufficient either. The careful and highly differentiating ethological analysis of a wide spectrum of animal behaviors is essential, but likewise not sufficient with respect to an assessment of the intentional states involved. It is necessary to combine these different efforts, and for such a combination between the disciplines, modelings are required that (a) are compatible with the conceptualizations in the different disciplines, and (b) can be related to empirical research. With careful empirical analysis, it may then be possible to distinguish different types and levels of organization in

such a way that it becomes possible to distinguish ontologically between objects of these different types and levels. Then it may become possible to connect with these objects a specific ontological status, to be taken into account by moral and legal regulations.

Finding an orientation in this difficult field requires the combination of two distinctive and heterogeneous approaches: (a) a careful epistemological analysis of argumentations, and (b) the modeling of properties that can be ascribed empirically to objects of observation. The combination of these two approaches is so difficult that one of them is usually abandoned to improve effectiveness of the other. In an attempt to draw attention to this problem, and find a way to overcome it, a cooperation or at least discussion between philosophers, ethologists and neurobiologists has been initiated. The various papers were discussed at a workshop held in September 1997 at the newly founded Hanse-Wissenschaftskolleg in Delmenhorst near Bremen. The contributions have not been integrated, and do not use the same terminology. Even within one discipline different theoretical positions are presented: in ethology, neurobiology, and philosophy. They have primarily been chosen for their suitability as potential building blocks for a bridge to be built between these disciplines. They may or may not represent the latest developments in their disciplines, but the authors, who where chosen for their competence in their respective fields, have developed aspects of their work that might be suitable building blocks. With this future goal in mind, behaviors of plants, animals, and machines are described and analyzed with respect to the levels of intentionality involved. Phylogenetic levels of brain organization are described by neurobiologists, and phylogenetic levels of complex behavior by ethologists. In a number of contributions, the relevance of such investigations for ontological and ethical questions is pointed out. Each of the papers is an interdisciplinary attempt addressing several of these tasks and including results of several disciplines.

In view of a growing number of contributions and the wide range of topics they cover, we decided to divide the papers up into two blocks, each with a different focus, which will appear as special issues in two scientific journals: the more epistemologically oriented papers in "Erkenntnis" (STEPHAN/HENDRICHS/DRECKMANN 1999) and the more general papers, including ontological considerations, in "Evolution and Cognition", respectively. The "Erkenntnis" volume combines papers (in alphabetical order) by: Colin ALLEN, Marc BEKOFF, Gordon BRITTAN, Donald DAVIDSON, Fred DRETSKE, Frank DRECKMANN, Hubert HENDRICHS, Joëlle PROUST, Daisie RADNER, and Achim STEPHAN, while the present volume combines papers of Colin ALLEN (the same paper as in the Erkenntnis volume), Marc BEKOFF (the same paper as in the Erkenntnis volume), Gordon BRITTAN, Gordon BURGHARDT, Holk CRUSE, Hans-Johann GLOCK, Hubert HENDRICHS, Rolf LACHMANN, Hans LENK, Gerhard ROTH, Ted SCHATZKI, Franz SEITELBERGER, and Gunther STENT. Several of the contributions of this volume address possibilities of modeling for empirical investigations: from the side of biology those of BEKOFF, BURGHARDT and CRUSE, from the side of philosophy those of ALLEN, BRITTAN, GLOCK, LACHMANN, and SCHATZKI. Some papers emphasize important ontological distinctions: SEITELBERGER and STENT as neurobiologists, LENK and SCHATZKI in philosophy; some mainly focus on continuities: ROTH in neurobiology, CRUSE in biological cybernetics, GLOCK in analytical, and LACHMANN in phenomenological philosophy.

BRITTAN draws attention to some of the central questions and specifies some of the problems. While behavior of machines and animals and that of animals with a different complexity of mental organization are often indistinguishable for an observer, performances can be distinguished on the basis of the type of programming and the kind of learning involved, that is, referring to differences in the life history of the performing objects, the different ways in which they acquired and use their capacities. He stresses that using representations is different from being aware of these. In addition he states reasons for his belief that the attention placed by some authors on language is misleading as other expressions of mind are important, and that animals are continuously evaluating and making judgments.

BURGHARDT investigates the evolution of play, being an important type of behavior in animals with higher levels of mental organization. Its generation combines cognitive and emotional components. Although not necessarily linked to intelligence, play constitutes a fertile ground for the joint development of perception, emotion, cognition, and consciousness. Like cognition and consciousness play provides an example (a) of fundamental difficulties of defining mental properties in relation to behavioral performances, and (b) of an emergence of new qualities in ontogeny and phylogeny.

ALLEN distinguishes between social and individual concepts and is particularly interested in the relationship between concepts and perception. He gives reasons not to abandon the notion of concepts and suggestions for empirical investigations of concepts in

nonhuman animals. He proposes clauses to be used, illustrating his arguments with examples from own studies with domestic pigs.

BEKOFF is interested in the use of the notion of intentionality in biology. He addresses the question how individuals negotiate cooperative agreements and exchange information concerning their beliefs and goals in different systematic taxa. Drawing on his own empirical research he uses examples from social play in mammals and antipredator behavior of birds. He defends the use of the notion of intentionality when investigating the behavior of animals and the use of classical ethological techniques for this investigations.

ROTH gives a general overview over brains in the animal kingdom, especially characterizing the vertebrate brain, and then concentrates on the brain of mammals, differentiating between primates and nonprimates, humans and apes. He considers different states of awareness and relates them to different brain areas, which contribute in different ways to the different types of awareness. He distinguishes several parameters of brain developments and investigates their potential for possible distinctions. Clear results in the field relating cognitive abilities to brain activity are still outstanding. Human consciousness includes various diverse phenomena produced by different brain areas. Most of these are present in all tetrapods (amphibians, reptiles, birds, mammals). The relatively large human isocortex is shared with most primates, the large prefrontal cortex with apes.

SEITELBERGER also states that, despite remarkable progress in neurosciences and bioinformatics, our knowledge of (higher) brain performances is still scant and peripheral. After distinguishing between states and contents of consciousness, he points out difficulties in describing and investigating brain performances. Based on his experiences as a neurobiologist he describes the construction of mental experiences in man. Information processing does not produce images, but compatible abstract reconstructions that are not conscious. Only some selected transcoded results, with autonomous ways of self-construction, do become conscious in humans. He notes new levels of complexity not reached in other primates and sees the human brain as an essential innovation with new structural completions and a total cortalization of inputs.

STENT, as a neurobiologist with extensive interdisciplinary experience, enters a very difficult philosophical domain, addressing fundamental philosophical and legal questions from a dualistic position. It is not the intention of this volume to oppose materialistic and idealistic, monistic and dualistic positions, but to approach their integration by distinguishing ontologically and epistemologically between several levels of (mental) organization which differ in number, form, and connectedness in different living and non living "behaving" objects (cf. HENDRICHS 1994).

CRUSE investigates possible similarities and differences in cognition, awareness and subjective feelings between humans and highly developed future robots. He points out transitions between external and internal perspectives and, in internal experience, distinguishes different degrees of connectedness with symbols, finally leading to manipulable internal models. For artificial neural networks, he describes possible states of relaxation, with a high "harmony value" when the positions of variables match, and of tension, when the harmony value is low; when two neural layers both represent the body, their representations may (a) coincide, resulting in a relaxed state, or (b) differ, resulting in an activated state, possibly inducing the network to change into a new state. It is proposed that this second state (b), when experienced, constitutes a first elementary condition for subjective internal perspectives.

LACHMANN, from a phenomenological perspective, addresses the question of animal awareness starting from conceptions developed by Alfred North WHITEHEAD, Arnold GEHLEN, and Maurice MERLAU-PONTY. Referring also to the studies of Susanne K. LANGER he points to a pragmatic value perception in animals and discusses the possibility of gaining access to components of animal awareness by analyzing basic components of human consciousness.

LENK, starting from anthropological propositions, especially those of Ernst CASSIRER and Susanne LANGER, specifies the "second nature" of human beings, characterizing human culture as constituted by several levels of symbolization. Symbolic meta-representations of symbolic representations are constructed in a "symbolic ascent". The "animal metasymbolicum" has to integrate several levels of interpretation and can thereby be distinguished from animals with highest mental capacities, as apes for example.

GLOCK, from an analytical perspective, addresses the question to what extent thoughts and concepts are dependent on verbal language. Avoiding extreme positions he develops an intermediate position allowing for simpler kinds of thoughts and concepts in living beings without verbal language. He proposes, that (a) the possession of concepts does not require verbal language, and (b) animals can use thoughts of a simple kind, that do not require concepts.

SCHATZKI analyzes important recent contributions proposed for modeling agency, and combines them into a promising new type of actor-network theory. He distinguishes two types of networks maintaining different functional relations to the activity of an actor: composing networks and embedding networks, both effecting agency in a different way. He sees an actor as an association of heterogeneous types of phenomena in relation to a multilevel and interconnected array of networks. He applies discourse multiplicity to this network multiplicity and describes three ways in which agency is effected by embedding networks. He points out significant differences between these compositional and embedding networks, and thereby develops a tool to distinguish between different types of actors: not only between animals, machines, and tornados, but also between scallops, dogs, chimpanzees, and humans.

In the following some brief remarks refer to my own position.[4] Human behavior can be influenced by beliefs, but it is not necessarily caused by them. There can be a "gap" (SEARLE 1999) in the cognitive causation of acts, leaving space for other components of orientation and organization. Already in higher mammals open areas in the programming of their behavior can be relevant for the integration of competing influences of different dimensions (cf. HENDRICHS 1997). However, most of these openings, which exist at birth in the programmed mind, become closed during ontogeny by a development of habits and customs, that is learned patterns of behavior. In higher mammals, these patterns cease to be fixed, but remain open for an individual orientation in fields of learned "rules". These rules, and the processes of orientation that make use of them, mostly are not consciously known by the acting individual.

In the domain of orientation and of organizing decisions, the mammalian brain has undergone drastic changes in comparison to the earlier reptile and bird-like brains. It could be described as a brain working with "gaps": sensual inputs are not transported in a linear fashion to central areas, but become transformed and connected to each other in highly complex associations. Thereby the possibilities of reactive behavior become enlarged, the field of possible reactions is opened up, and requires specific orientations and decisions in each situation. This appears to have occurred at different levels of mental processing (and similarly in the human analysis of these processes). In between the programmed structures, "hard-" and "soft-wired" instructions for organization, orientation, and action, there remain "gaps" which are constitutive for efficient action: uninstructed open areas that have to be bridged by the behaving or acting organism with specific decisions, combining conscious and subconscious, cognitive and emotional components. (The transformation of information and the generation of decision in future computing robots on one side and in mammals on the other will thus possibly continue to rely on different processes, with different causalities and different rationalities.)

Motivations, behavior, social relations, and knowledge are domains of different dimensions and involved to different extents in the production of decisions in mental and intentional acts. An experienced observer, knowing thoroughly the observed animals and their behavior, can in suitable situations on the basis of a careful, that is attentive, critical, and patient registration of behavioral details, recognize with some degree of reliability the different processes combining in the production of specific "decisions" in the observed animals. He or she can distinguish between different types of awareness in the behaving individual and between programmed and intentional behavior. He or she does not ascribe intentionality when the behavior only "looks as if" it were intentionally brought about, not even when the behavior in "shorthand" is described in intentional terms. The experienced observer knows that behavior produced without any intention on the part of the behaving individual may strike an inexperienced observer as "obviously" intentional. That of course is easier to realize when observing a deer rather than an ape. But this aspect is also very important when observing apes, even here many actions that appear "obviously" intentional can be produced without a specific awareness on part of the behaving animal. In particular, the experienced observer realizes that object-related behavior of an animal does not require a recognition of this object by the behaving animal, not to speak of a concept of this class of objects. Likewise the correcting of a behavior after "recognizing" a mistake in a specific situation (e.g., mistaking a higher-ranking group member for a lower-ranking one) does not imply a concept of situation or mistake (and is not dependent on verbal language).

Propositional contents of beliefs and desires in human orientation in most cases remain closely connected to subconscious processes and thus related to non-mental organismic and environmental processes. They directly and indirectly are in contact with structures and processes of an observer-independent reality. The contacted outside structures and processes are of different dimensions—physical, chemical, social, mental—and of different organizations. An orientation in these

different domains has to combine different forms of logic and to integrate them into one rationality. The rationality of decisions, moral and otherwise, therefore also depends on the consistency of the required integrations, and not only on calculations in "formal logic" (i.e., this rationality is possibly not of a digital/binary type as commonly used).

Looking back upon the heterogeneous papers from various disciplines it is apparent that the different contributions, though they did not arise from a common ground, can be related to a common center. This can be characterized by the following, relatively simple, but nevertheless important finding: There is no simple dichotomy between phenomena and underlying reality, but a multilayered structure of different types of observable phenomena in one common reality. The layers may have different dimensions, and their analysis may require different rationalities. The access to a multilayered complex of different components—like "consciousness"—depends on specific properties of the approach. It has to be able to integrate different dimensions and rationalities in its analysis. In our case: for an analysis and an understanding of human consciousness and conscience, and their evolution, it is not sufficient to consider cognitive propensities and performances. Other components of different dimension and rationality may have to be included, at least that of emotionality.

The evolution of cognition is closely connected to the evolution of emotionality (WIMMER/CIOMPI 1996; cf. LEWIS 1997; LABAN HINTON 1999). And the evolution of both these faculties is required for the phylogenetic and ontogenetic development of morality, the ability and the disposition to take responsibility for individual behavior and its consequences. The highest form of this conscientiousness is only found in humans, but there are many dispositions and tendencies in higher animals that can be considered roots of this highest human quality (HENDRICHS 1999). They develop in socially transmitted fields of orientation and consist of rules and taboos in networks of relations and contexts. They are required for example for the organization of life in a social context without excessive aggressions, that is, for creating a social situation in which the young can develop without excessive strain and thereby develop their social and mental competence to the full species specific level (HENDRICHS 1997). In a world of increasing virtuality, it is becoming increasingly difficult not to loose sight of those differences between dogs, guinea pigs, and tamagotchis that are essential for their adequate treatment. To consider the cohesive evolution of emotion, cognition, and conscientiousness will, in this situation, be essential for any attempt to contribute to a task indispensable to a future global society: finding acceptable and sustainable solutions for agreements between humans of different cultures and for the fair treatment of living beings with different levels of mental complexity.

Author's address

Hubert Hendrichs, University of Bielefeld, Department of Ethology, Morgenbreede 45, D-33615 Bielefeld, Germany.

Acknowledgments

I thank Gordon BURGHARDT, Frank DRECKMANN, Ursula HENDRICHS, Rolf LACHMANN, and Achim STEPHAN for their critical comments on earlier versions of this paper, Gerhard ROTH for providing, at the Hanse-Wissenschaftskolleg, possibilities for the discussion of various papers, and Adelheid E. BAKER for language assistance.

Notes

1 Don GRIFFIN later documented his efforts: GRIFFIN (1981; 1992), cf. RISTAU (1991).

2 Cf. CHANGEUX/KONISHI (1987), OESER/SEITELBERGER (1988), EDELMAN (1989), ROTH (1995); for cognitive ethology cf. RIEDL (1980), GOULD/GOULD (1994), ALLEN/BEKOFF (1997).

3 Some of the older contributions to this field contain valuable proposals for interdisciplinary conceptualizations of the problems addressed in our volume, although they have only partly been taken up again in the more recent discussions. Cf. BROAD (1925), UEXKÜLL/KRISZAT (1932), HEBB (1949), GREY WALTER (1953), BOULDING (1956), ZIFF (1959), LORENZ (1963/65), (1981), POLANYI (1966), LANGER (1967; 1972), PIAGET (1971), ECCLES (1973), LURIJA (1973), CAMPBELL (1974), PUTNAM (1975).

4 For details cf. HENDRICHS (1999).

References

Agazzi, E. (1981) Intentionality and artificial intelligence. Epistemologia 4: 195–228.

Allen, C./Bekoff, M. (1997) Species of mind. The philosophy and biology of cognitive ethology. MIT Press: Cambridge MA.

Baars, B. J. (1997) In the theater of consciousness. The workspace of mind. Oxford University Press: Oxford.

Birbaumer, N. (1996) Selbstregulation des Gehirns und Verhalten. Verhandlungen der Gesellschaft Deutscher Naturforscher und Ärzte 119: 297–311.

Boulding, K. E. (1956) The image. Michigan University Press: Ann Arbor.

Broad, C. D. (1925) The mind and its place in nature. Routledge and Kegan, London.

Campbell, D. T. (1974) Evolutionary epistemology. In: Schilpp, P. (ed) The philosophy of Karl Popper, vol. 1. Open Court: La Salle, pp. 413–463.

Changeux, J. P./Konishi, M. (1987) The neural and molecular basis of learning. Wiley: Chichester.

Cheney, D. L./Seyfarth, R. M. (1990) How monkeys see the world. Inside the mind of another species. University of Chicago Press: Chicago.

Dennett, D. C. (1987) The intentional stance. MIT Press: Cambridge MA.

Dretske, F. (1988) Reasons in a world of causes. MIT Press: Cambridge MA.

Dreyfus, H. L./Dreyfus S. E. (1988) Mind over machine. The power of human intuition and expertise in the era of the computer. Free Press: New York.

Eccles, J. C. (1973) The understanding of the brain. McGraw-Hill: New York.

Edelman, G. M. (1989) The remembered past. A biological theory of consciousness. Basic Books: New York.

Fodor, J. A. (1987) Psychosemantics. The problem of meaning in the philosophy of mind. MIT Press: Cambridge MA.

Gould, J. L./Gould, C. G. (1994) The animal mind. Freeman: New York.

Grey-Walter, W. (1953) The living brain. Duckworth: London.

Griffin, D. R. (1981) The qestion of animal awareness. Rockefeller Univeristy Press: New York.

Griffin, D. R. (1992) Animal minds. University of Chicago Press: Chicago.

Hebb, D. O. (1949) The organization of behavior. A neurophysiological theory. Wiley: New York.

Hendrichs, H. (1994) Tier—Mensch—Maschine: Übereinstimmungen und Verschiedenartigkeiten in der Regulierung ihres Verhaltens. In: Letzgus, K. et al. (eds) Für Recht und Staat. Festschrift für Herbert Helmrich. Beck: München, pp. 1107–1124.

Hendrichs, H. (1997) On the development of psychosocial individualities in group living mammals. In: Weingart, P. et al. (eds) Human by nature. Between biology and the social sciences. Erlbaum: Mahwah NJ, pp. 400–407.

Hendrichs, H. (1999) Different roots of human intentionality in mammalian mentality. In: Stephan, A. et al. (eds) Animal mind. Erkenntnis 51: 145–164.

Kim, J. (1996) Philosophy of mind. Westview Press: Boulder Col.

Laban Hinton, A. (ed) (1999) Biocultural approaches to the emotions. Cambridge Univeristy Press: Cambridge.

Langer, S. K. (1967) Mind: An essay on human feeling, vol. I. The Johns Hopkins University Press: Baltimore, London.

Langer, S. K. (1972) Mind: An essay on human feeling, vol. II. The Johns Hopkins University Press: Baltimore, London.

Lewis, M. (1997) The self in self-conscious emotions. Annals of the New York Academy of Sciences 818: 119–142.

Lorenz, K. (1963/65) Haben Tiere ein subjektives Erleben? In: Lorenz, K. (1965) Über tierisches und menschliches Verhalten, vol. II. Piper: München, pp. 359–374.

Lorenz, K. (1981) The foundations of ethology. Springer: New York.

Lurija, A. R. (1973) The working brain. Harmondsworth. (Deutsch: Das Gehirn in Aktion. Einführung in die Neuropsychologie. Rowohlt: Hamburg 1992).

Markowitsch, H. J. (1992) Neuropsychologie des Gedächtnisses. Hogrefe: Göttingen.

Oeser, E./Seitelberger, F. (1988) Gehirn, Bewußtsein und Erkenntnis. Wissenschaftliche Buchgesellschaft: Darmstadt.

Penrose, R. (1989) The emperor's new mind. Concerning computers, minds, and the laws of physics. Oxford University Press: Oxford.

Pfannenstiel, H.-D. (ed) (1996) Verhandlungen der Deutschen Zoologischen Gesellschaft. 89. Jahresversammlung, G. Fischer: Stuttgart, New York.

Piaget, J. (1971) Biology and knowledge. An essay on the relations between organic regulations and cognitive processes. University of Chicago Press: Chicago.

Polanyi, M. (1966) The tacit dimension. Doubleday: New York.

Premack, D./Premack, A. (1983) The mind of an ape. W. W. Norton: New York.

Putnam, H. (1975) Mind, language, and reality. Philosophical papers, vol. II. Cambridge University Press: Cambridge.

Pylyshyn, Z. (1984) Computation and cognition. Toward a foundation for cognitive science. MIT Press: Cambridge MA.

Radner, D./Radner, M. (1989) Animal consciousness. Prometheus: Buffalo.

Riedl, R. (1980) Biologie der Erkenntnis. Die stammesgeschichtlichen Grundlagen der Vernunft. Parey: Berlin, Hamburg.

Ristau, C. A. (1991) Cognitive ethology. The minds of other animals. Essays in honor of Donald R. Griffin. Erlbaum: Hillsdale, London.

Roth, G. (1995) Das Gehirn und seine Wirklichkeit. Suhrkamp: Frankfurt.

Searle, J. R. (1983) Intentionality. An essay in the philosophy of mind. Cambridge University Press: Cambridge.

Searle, J. R. (1992) The rediscovery of the mind. MIT Press: Cambridge MA.

Searle, J. R. (1999) Rationality in action. Opening lecture of a symposium "Speech acts, minds, and social reality. Discussions with John R. Searle" held at the Zentrum für interdisziplinäre Forschung, Bielefeld, June 21–23.

Stephan, A./Hendrichs, H./Dreckmann, F. (eds) (1999) Animal Mind. Erkenntnis 51.

Uexküll, J. v./Kriszat, G. (1932) Streifzüge durch die Umwelten von Tieren und Menschen. Fischer: Frankurt.

Wimmer, M./Ciompi, L. (1996) Evolutionary aspects of affective–cognitive interactions in the light of Ciompi's concept of "affect logic". Evolution and Cognition 2: 37–58.

Whitten, A./Byrne, R. (1988) Tactical deception in primates. Behavioral and Brain Sciences 11: 233–273.

Ziff, P. (1959) The feelings of robots. Analysis 19: 64–68.

Gordon G. Brittan, Jr.

The Mental Life of Other Animals

Abstract

DESCARTES claims that it is in principle possible to design and construct machines whose behavior is indistinguishable from that of any nonhuman animal. If the behavior is indistinguishable, then it follows that we have to attribute no more capacity to the animal than to the machine. Since the machine has no more than a "mechanical capacity," we need to assume no more than a "mechanical capacity" (what DESCARTES calls a "disposition of the organs") on the part of the animal. But this argument turns on a non-sequitur. Machines are programmed, animals are conditioned, and each acquires its capacity in a very different way. The fact that the behavior is indistinguishable does not imply that the capacity is indistinguishable. Although disagreeing with him in a variety of ways, many philosophers (including KANT and DAVIDSON) have followed DESCARTES in stressing the important of linguistic competence in making a distinction between humans, on the one hand, and animals and machines, on the other. But a second non-sequitur is involved. Even supposing that other animals are not competent linguistically does not imply that they do not enjoy a mental life, in particular that they do not have beliefs or desires, or are not self-conscious.

Key words

DESCARTES, behaviorism, linguistic competence, animal mental life.

1. Most major twentieth century philosophers reject DESCARTES' metaphysical position, in particular his dualism of mind and body. So it is perhaps surprising that they also tend to accept the rather narrow distinction that he makes between human beings and other animals, especially since he uses that distinction in turn as an important support of his dualism. I won't try to unravel this apparent inconsistency here. In my own view, an adequately naturalistic account of human beings must embrace their continuity with other animals.

2. DESCARTES' argument for the distinction between human beings and other animals, set out in detail in the fifth part of the *Discourse on Method* and the letter to the Marquess of Newcastle of November 23, 1646, turns for the most part on questions of linguistic competence. The passage from the *Discourse* is already very well known. Here is what is said in the letter to the Marquess of Newcastle:

"…none of our external actions can show anyone who examines them that our body is not a self-moving machine but contains a soul with thoughts, with the exception of words, or other signs that are relevant to particular topics without expressing any passion. I say words or other signs, because deaf-mutes use signs as we use spoken words; and I say that these signs must be relevant, to exclude the speech of parrots… I add also that these words or signs must not express any passion, to rule out not only cries of joy or sadness and the like, but also whatever can be taught by training to animals. If you teach a magpie to say good-day to its mistress, when it sees her approach, this can only be by making the utterance of this word the expression of one of its passions. For instance, it will be an expression of the hope of eating, if it has always been given a tidbit when it says it… it seems to me that the use of words, so defined, is something peculiar to human beings… there has never been an animal so perfect as to use a sign to make other animals understand something which expressed no passions and there is no human being so imperfect as not to do so, since even deaf-mutes invent special signs to express their thoughts". (KENNY 1970, pp206–207)

3. The main line of the argument can be put as follows:

(i) Any (relatively complex and uncoerced) behavior which can be explained in terms of the "passions" does not require thought, any behavior which cannot be so explained does.

(ii) All animal behavior, including apparently linguistic behavior (the purring of cats, the barking of dogs, the chattering of magpies) can be explained in terms of the "passions".

(iii) Therefore, no animal behavior requires thought.

(iv) But at least some human behavior, in particular human linguistic behavior, cannot be explained in terms of the "passions".

(v) Therefore, at least some human behavior requires thought.

(vi) Therefore, human beings are to be sharply distinguished from other animals; they alone have minds.

4. The argument is evidently valid. We need to look more closely at its premises. Let me begin with the second, that all animal behavior, including apparently linguistic behavior, can be explained in terms of the "passions". Whatever the precise meaning of the word "passion" in its seventeenth century context, DESCARTES' point is clear. All apparently linguistic behavior on the part of animals can be explained either as a relatively spontaneous outpouring (the purring of a contented cat) or an instinctive reflex (a dog barking at the approach of strangers) or as a conditioned response (the magpie taught to say goodday to its mistress). In none of these cases, then, does the behavior require any thought. Spontaneity, instinct, and conditioning suffice to explain[1] the production of what, in the *Discourse*, are called "natural sounds".

We will not worry for the moment whether or not this claim is true, although it is worth remarking that in making it DESCARTES downplays the importance of the traditional distinction between what is "learned" (DESCARTES says that the magpie has been "taught") and what is merely "instinctive". What concerns me now are the implications he takes the claim to have. One has already been mentioned; it is that if spontaneity, instinct, or conditioning suffice to explain behavior, then its further explanation in terms of "thought" is precluded. The other implication is that animals cannot be distinguished from machines. The first implication is problematic. The second is clearly false.

For the purposes of argument, we need to consider machines whose behavior is (or can be imagined to be) of the same order of complexity as, for example, that of cats, dogs, and magpies, not only the hydraulic automata in the gardens of the king whose movements so fascinated DESCARTES, but also so-called "servo-mechanisms" like homing rockets. The behavior of these latter, unlike the animals DESCARTES mentions, is not to be explained in terms of spontaneity, instinct, or (especially) conditioning, but rather in terms of programming.

There are at least two crucial differences between conditioning and programming. One is that conditioning is in a certain sense after the fact. An animal is rewarded or punished for behavior that has already occurred. Such reinforcement accordingly raises or lowers the probability that the animal will engage in the same behavior on the next occasion when it is presented with approximately the same conditions. A machine that has not already been programmed (in some wide sense of the word that includes hydraulic automata as well as homing rockets) will simply not do anything. In this sense, programming is "before the fact". A second difference between conditioning and programming is aligned with the first. Systems of reward and punishment work only if the "reward" is accompanied by feeling of pleasure and the "punishment" by feelings of pain. but the kinds of machines we are considering do not have pleasure and pain receptors, hence they cannot be conditioned. On the other hand, something can be programmed only if it is capable (consciously or not) of following a set of instructions.

It follows, as against DESCARTES, that animals are not machines (at least not the kinds of machines DESCARTES had in mind). Animals, but not machines, can be conditioned; it would be inappropriate, to say the least, to raise a chess-playing computer's voltage after a successful match. If animals can be programmed, by way of a particular genetic endowment for example, then the behavior which such programming explains (much of what we otherwise term "instinctive") is not the same sort of ("learned") behavior which conditioning explains. Is a machine that could be conditioned, and in this sense that learns, conceivable? I don't see why not. The point is that it would be very much more like human beings than DESCARTES was willing to allow.

Now the distinction between animals and machines has important corollaries. One of these corollaries has to do with the way in which DESCARTES sets up his argument. He claims that it is in principle possible to design and construct machines whose behavior is indistinguishable from that of any animal in particular. If the behavior is indistinguishable, then it follows that we have to attribute no more capacity to the animal than to the machine. Since the machine has no more than what might be called, globally, a "mechanical capacity," to be explained in terms of the motion and interaction of its parts, we need to assume no more than a "mechanical capacity" on the part of the animal, to be explained at least in large part in terms of what DESCARTES calls the "disposition of its organs"..

But we granted that the behavior of animals and machines was indistinguishable and yet argued that they were to be distinguished nevertheless, in terms of the fact that the former but not the latter was to be accounted for in terms of conditioning. In addition to behavior, we must also take into consideration the way in which the capacity to behave was acquired, that is, we must also take into account the histories of the animals and machines involved. However indistinguishable their behavior might be, if they have different histories of how their behavioral capacities came to be acquired, then they are to be distinguished *tout court*. Human beings have the same sort of natural, if not also cultural, histories that many other animals do. From this perspective, human beings are in a crucial, although not decisive, respect like animals, both of them to be distinguished sharply from machines.

The point needs to be stressed. DESCARTES assumes that behavior provides the only evidence we have for hypotheses concerning the minds of others. The behavior of animals and machines, it might be granted, is indistinguishable. Therefore, one has no less and no more "mind" than the other. But machines don't have "minds". Therefore, animals don't either. Human verbal behavior, on the other hand, differs in kind from both animal and machine behavior; it is creative, adaptable, and all the rest. Therefore, human beings do have minds. These conclusions follow, however, only if we assume that behavior provides the only evidence we have for hypotheses concerning the minds of others. It doesn't. Among other things, we also have the histories of those involved. Behaviorism, even in the weak way in which DESCARTES understands it, as limiting the evidence on the basis of which our inferences to the minds of others are made, is false.

The other corollary of the distinction between animals and machines is this. In saying that animals are conditioned, we presupposed that they are "motivated". DESCARTES insists on this point. We teach the magpie to say good-day to its mistress on her approach by giving it a tidbit only because it wants the tidbit.[2] If the prospect were not motivating, then offering it after an appropriate "good-day" would not raise the probability of more "good-days" in the same circumstances. That is, in explaining how conditioning works, it is inevitable that we attribute certain desires, among them the "hope of eating," to the animal. In the case of machine behavior, on the other hand, the attribution of desire would be inappropriate.

The point needs to be elaborated. First, in saying that an animal has desires, we are, at least initially, saying no more than that its behavior is (consciously or unconsciously) goal-directed. That such goal-directedness is not simply a matter of responding in a particular way to a given stimulus is clear when we consider cases in which the animal is frustrated in achieving its goal. In such cases, there is no option to saying that the animal is *trying* to reach its goal, that is, desires to reach its goal but is not able to. Second, in attributing desires to animals we eventually go beyond describing their behavior to positing the existence of something "going on inside their heads," internal representations of what is not now present, the goal-state to be achieved. Third, in explaining the behavior of machines we need posit no such internal representations. It is enough, at least in principle, to refer to their programs and to their design. If there is a sense in which a machine such as a homing rocket can be said to have goals, the destruction of a plane or missile target for example, and in this same sense to have desires, then these goals and desires are derivative of the goals and desires of the person who designed and built the machine in the first place. Animals, of course, are not in the same way designed and built with a certain end in view. They evolved.

5. I have so far said nothing about the crux of DESCARTES' argument, the issue of linguistic competence. To this point, I have only wanted to maintain that animals are to be distinguished from machines and that the grounds of this distinction do not have to do with behavioral differences (for the sake of argument we can assume that there are none), but with the different ways in which behavioral capacities are acquired. Animals acquire at least some of their behavioral capacities through conditioning, machines do not. Moreover, it is in virtue of their respective histories that it is appropriate to attribute desires to animals, but not to machines. To this extent, that animals have desires (that they are "motivated" to behave in certain ways), they have a mental life. To this same extent, DESCARTES would (perhaps not quite consistently, given his animal-machine hypothesis) agree. He too attributes to them such otherwise "human" motivations as fear, hope, and joy. But, he claims (more in my words than in his), we cannot attribute beliefs to them. For the attribution of belief is somehow bound up with linguistic competence, and animals are not competent linguistically.

6. The apparent importance of language to the mental life of other animals can be brought out by considering the views of Immanuel KANT. Although he puts the point in very different ways and with different emphases than I have, KANT too argues for attributing desires (and in the process, representations) to other animals, distinguishing them from machines and assimilating their behavior to ours. Perhaps the most important of the several passages in which he does so is in the *Critique of Judgment* (in a footnote to AK.5: 464).[3]

What we can quite correctly infer by analogy, from the similarity between animal behavior (whose basis we cannot perceive directly) and men's behavior (of whose basis we are conscious directly), is that animals too act according to representations [*Vorstellungen*] (rather than being machines, as DESCARTES would have it), and that regardless of the difference in specific kind between them and men, they are still of the same general kind (namely, as living beings).

As his examples (beaver building lodges, and so on) make clear, KANT takes at least some animal behavior to be intentional or purposive, and thus to require the attribution of desires (a type of internal representation in his framework) in its explanation. In this respect, it is very much unlike machine behavior.

At the same time, KANT thinks that there is an important difference to be made between human beings and other animals. The former, but not the latter, are self-conscious. In this respect, of course, he sides with DESCARTES, for whom the essential thing about animals (and also machines) is that they don't know what they are doing. If their behavior is goal-directed, it is not consciously so, and in this fact lies the principal difference between us and them.

KANT has a number of reasons for wanting to deny that animals are self-conscious. Many of them have to do with the fact that humans, unlike other animals, are moral agents and can be held responsible for their actions. The conceptual tie between moral agency and self-consciousness is supplied by the fact that although animals have desires, they are not aware that they do, and thus cannot consciously choose to act on them or not, as we do. Moral agency presupposes freedom, and animals, lacking self-consciousness, are not free.

But these reasons for wanting to deny that animals are self-conscious do not amount to a proof that they are not, short of begging important questions. In fact, so far as I can tell, KANT has no argument for the claim that animals are not self-conscious, although he repeats this claim on a number of occasions. Apparently he does not think that analogies between our own behavior and that of other animals establish that they, too, are self-conscious in the way that analogies establish that they are conscious. I take it, moreover, that for KANT no analogies between their behavior and ours *could* establish that other animals were self-conscious, since *all* we can infer by analogy on the basis of behavior is that the behaver has "representations," and not that it is further aware of these representations. But that we cannot establish self-consciousness by analogy on the basis of behavior does not prove that other animals do not possess it, especially since, as we have already seen, there are other factors which must be taken into account in attributing a mental life to others. This is as true in the case of other human beings as it is in the case of other animals.

KANT also says that animals do not possess self-consciousness since they are incapable of making judgments, which among other things involves comparing representations. This time the link between the making of judgments and the possession of self-consciousness is by way of the fact that the possession of self-consciousness involves, at a minimum, distinguishing between how things stand with the subject and how things stand with the world, and this distinguishing involves the making of judgments. But again, no reason has been given why other animals cannot make judgments (or possess the appropriate concepts) and thus why they cannot distinguish between how things stand with them, and with the world.

Of course, KANT would have an effective argument in both cases if he added another premise. In the first case, the premise is that another animal is self-conscious, e.g., possesses an "inner life" of its own, only if it could tell us in some detail about this life. Since it cannot tell us in any detail about this life, it is not self-conscious. In the second case, the premise is that one can make a judgment only if one possesses a language in which to make it. That is, a judgment is a propositional attitude and propositions are linguistic in character. Since other animals lack language, they cannot make judgments and are hence not self-conscious.

7. The sort of argument that KANT, and ultimately DESCARTES, hint at has recently been supplied in some detail by Donald DAVIDSON in his essay, "Rational Animals" (DAVIDSON 1982). It has two main premises:

(i) In order to have a belief, it is necessary to have the concept of belief.

(ii) In order to have the concept of belief, one must have language.

In order to reach the conclusion that other animals do not have beliefs, and thus do not in any rich sense have a mental life, we must, of course, add the premise that they do not have a language, a premise which DAVIDSON accepts but does not argue for. The argument in general, then, is that

(i) animals have a mental life only if they are competent linguistically (this follows from DAVIDSON's two premises);

(ii) animals are not linguistically competent;

(iii) therefore, animals do not have a mental life.

In fact, most people who want to deny the conclusion of this argument follow MONTAIGNE in denying its second premise. I think it is dangerous to do so, although not for the reason that DESCARTES gives, that if other animals had a language they would long since have succeeded in communicating with us and with each other. Failure to understand another does not entail that he or she is not linguistically competent, witness our failure to understand the language of the ancient Mayans. Four other sorts of consideration must be kept in mind.

First, there is preliminary evidence that linguistic competence is very closely tied to evolutionary developments in the cerebral cortex unique to human beings. There are other aspects in which our species is unique biologically, and it might very well be the case that a series of mutations occurred after we left the savannah which made speech possible. Some venturesome linguists have already begun to suggest dates when speech emerged.[4]

Second, it is at the very least controversial whether the apparent competence of nonhuman primates in manipulating conventional signs (American Sign Language, for example) is not simply to be accounted for, as DESCARTES insisted, in terms of subtle cuing and conditioning. Since we do not think that human language can be accounted for in terms of cuing and conditioning, no matter how subtle, there is reason to deny that these animals are "speaking" in the same way that we do. This is not to deny that other primates, and for that matter dolphins and whales as well, manipulate signs or produce a rich variety of sounds. It is to suggest that at this point we cannot make a case for animal language on the basis of what we know about them. It might be added, again at this point in time, that there is scarcely more reason to attribute a mental life to those species that manipulate conventional signs than to those species that do not, for the behavior of these latter is in many cases just as complex and apparently "intelligent" as the behavior of the former.

Third, to attribute language to other species, given what we now know, would be to beg the most important questions, for to attribute language to them would be to endow their gestures, signs, and sounds with meaning, and the only way in which they could be said to have meaning would be as the expression of beliefs. But whether or not they can be said to have beliefs is precisely what is in question.

Fourth, and most important, we do not need to attribute linguistic competence to other animals in order to defend the claim that they have a mental life. That is, the problem lies with the first premise, that animals have a mental life only if they are competent linguistically.

DAVIDSON claims, to begin with, that in order to have a belief, it is necessary to have the concept of belief. Why should this be? We need to focus on a particular case.

Grizzlies, apparently unique among bears in this respect, routinely cache food. We explain an individual's returning to its cache and eating what was stored there by saying that it was hungry, desired food, and believed, on the basis of what it remembered, that it would find it at that spot. There is, in fact, no currently available option to anticipating its behavior (or the behavior of a great variety of other animals) than to provide it with this sort of intentional explanation. The explanation is appropriate when, among other things, the behavior is not coerced or prompted, not accidental (the bear did not stumble across the cache), not automatic (the bear does not return to this particular place at regular intervals, whether or not food is there). Moreover, we can tell, in a perfectly objective way, whether or not the behavior is coerced or prompted, that is, we can tell whether or not the bear was pushed or cued to that spot, just as we can tell whether it returns merely coincidentally or as a function of habit. Occasionally determining whether these conditions are satisfied requires that we set up experiments. If they are satisfied, then we are justified in taking what Daniel DENNETT calls the "intentional stance" with respect to the animal's behavior (DENNETT 1988).

Now it should be clear that taking the "intentional stance" presupposes that the bear's behavior is adaptable to changed circumstances, that is, that it re-adjusts its behavior in the light of new information which it receives. On the assumption that its desire remains the same, to get something to eat, the re-adjustment of behavior can only result from a re-adjustment of beliefs. But to say that the bear re-

adjusts its beliefs is to say that, in the light of the new information, it gives up some beliefs for others. This entails that it has a belief about other beliefs, namely, that they are false. But to have a belief about others of its beliefs, what we might call a second-order belief, it must have the concept of belief.

Talk here of "concepts" is tendentious, for it already tips the scales in favor of the importance of language to the having of beliefs. How, after all, could one have *concepts* without also having a language in which those concepts were inter-related and applied? But I see no problem in attributing a belief about beliefs, a second-order belief, to the grizzly, even in the absence of linguistic "concepts".

Suppose that the bear has cached its food. When the bear has left the site, we crawl in and remove it. Later the bear returns to the cache site, looking for food. We know that it expects to find food there[5], not simply on the basis of its "searching" behavior, but also because we are familiar with its history, among other things with the facts that it hid food there in the first place and that when it hides food it customarily finds it. How else are we to describe the situation than to say that the grizzly is surprised to find that the food is gone? But if this is the correct description (and it is certainly the natural, pre-theoretical description), then, if it eventually leaves the site without coercion or prompting and is still hungry, it must also be the case that it has given up the belief that food was there. That is, it must have a belief about one of its beliefs, namely, that it is false.

DAVIDSON's claim that "surprise about some things is a necessary and sufficient condition of thought in general" (DAVIDSON 1982, p326) seems much too strong. For surely surprise is not a necessary condition of thought in general. But it is plausibly a sufficient condition. And the fact is that some animals are from time to time clearly surprised.

But we have not yet come to the crux of DAVIDSON's position, that in order to have the concept of belief one must have language. Two related lines of thought are developed to reach this conclusion. Both depend on granting that having a belief about a belief presupposes having the "concept" of belief.

The first line of thought has to do with the individuation of beliefs. Behaviorists sometimes say that to have a concept of something, say of a cow, it is enough to be able to discriminate that thing from other things, to distinguish between cows and non-cows. Clearly many animals have this power of discrimination with respect to a variety of objects; their survival depends on it. But, it might be argued, although the ability to discriminate objects is a necessary condition of having a concept, it is not also sufficient. For to have a concept, once again of a cow say, one must also be able to identify and individuate cows, and this requires something like a language.

We can illustrate the point with reference to our grizzly bear case. Although the bear can distinguish between food and non-food, at least up to a point, it does not really "have the concept" of *food*. "Having the concept" involves having a number of beliefs about food, that it is nutritious, that the body requires it on a regular basis, and so on, in terms of which we can identify food as *food*, in the process distinguishing it from other objects which are not nutritious, which the body does not require on a regular basis, and so on. The only way in which these various beliefs may themselves be distinguished, their particular content identified, is with respect to the sentences which express them. But I have admitted that the bear is not capable of sentences. In which case it does not really have the concept of *food*. In which case it does not really have any *beliefs* about food, in particular that some awaits it at the cache site. Although, *faute de mieux*, we might attribute "beliefs" to it in order to explain and anticipate its behavior, this is no more than a *facon de parler* and should not be taken literally.

This line of argument is not persuasive. For one thing, it does not recognize that the identification of objects as such is a matter of degree, and often a matter of context. The child, the chef, the glutton all have a number of beliefs about food. These beliefs are presumably not the same. Does it follow that they have different *concepts* of food, to the point where we can say confidently that their beliefs about food are not *really* about food? But if it is a matter of degree, and there is no principled way in which to list a body of beliefs as necessary and sufficient for one to *have* the concept of food (Wittgenstein taught us to say that with respect to concepts like this, different usages are tied together by no more than family resemblances), then there would seem to be no reason why the *bear* couldn't have some concept of food (although we would have to identify that concept in our language), and to that extent beliefs about it. That the word "belief" is correctly used here is underlined by the fact that a bear could believe that food was at the cache site when we had, in fact, removed it; that is, the belief in this sort of case clearly has the "aboutness" characteristic of intentionality.

For another thing, although *we* identify beliefs by way of sentences, there is something highly problematic about using this fact to argue that insofar as

an animal is incapable of sentences, it is incapable of beliefs.

DESCARTES seems to assume that we understand another when he or she speaks to us, and thus that in uttering a sentence he or she at the same time identifies the content of a particular belief. But assume that we do not, at least initially, understand what he is saying. How do we come to understand him? To put a very complicated story very briefly, we come to understand him only by interpreting his behavior, including what might be called his "verbal" (written, spoken, and "body language") behavior. But this interpretation requires, on DAVIDSON's argument, that we attribute any number of other beliefs to him before we can attribute the possession of any concepts to him and thus attach meanings to his behavior. That is, we must ascribe any number of beliefs (inevitably, our own) to him before we can ascribe to him a language in terms of which these beliefs are to be identified. And what justifies such ascription in the case of human beings, for example an imagined primitive people whose beliefs are few in number and relatively simple ("food" for them might indicate anything in the environment that was green, that is, edible or digestible) and not in the case of at least some other animals if it is not a prejudice in favor of human beings from the outset, that they have a mental life simply in virtue of sharing a common physiology with us? DESCARTES says that deaf-mutes will always make themselves understood to us. Perhaps (in fact, I am not so sure). But if they do, it is only because we have knowledge both of the context in which they find themselves and of what typically motivates their behavior, the same sort of knowledge we bring to bear when we say that the dog, scratching at the door, is "trying to tell us something". Difficulties about the minds of animals inevitably become difficulties about the minds of other human beings.

The other line of thought behind the conclusion that in order to have the concept of belief (and thus to have beliefs) one must have language has to do with self-consciousness. For in order to have the concept of belief, I must (as indicated earlier) have the concept of a *true* belief, and this in turn requires that I be able to distinguish between how things stand with me (my subjective perception of the situation) and how things stand with the world (what is objectively the case). But to distinguish in this way between what is objective and what is merely subjective requires concepts and eventually a language. KANT put it this way: only beings capable of making judgments (about what is the case) are capable of self-consciousness, and only beings in possession of a language are capable of making judgments (about what is the case).

I agree that if an animal is not self-conscious, then it is inappropriate to attribute *beliefs* to it on anything more than an "intentional stance". The question is: does an animal that lacks language necessarily lack self-consciousness? Let me begin my answer to this question by making a very general point, and then turn more directly to the DESCARTES–KANT–DAVIDSON line of argument.

The very general point is that all of the attention placed on *language* in connection with the minds of others, animal or human, is misleading in at least two closely related ways. One, this attention suggests that the use of language simply "reports" what is going on inside the mind of others, as though the others were "phoning out" their thoughts and feelings, with no way of determining whether what is reported is true. On the other, attention to language suggests that the mental states ostensibly reported are in every respect as determinate as the sentences reporting them. But both suggestions are misleading: the first, if taken seriously, leads to scepticism about other minds, the second conflicts with the fact that our mental states are often unclear and indistinct (until, after the fact, we reconstruct them in precise verbal form).

These problems surface in the "argument by analogy" which KANT takes to be our way "into" other minds. It is not, largely because it presupposes a mistaken view of the way in which behavior stands to mind. On this "argument," to put it briefly, analogies or similarities between our behavior, as effect, and that of other animals allows us to infer an analogy or similarity between the causes of the behavior, i.e., an inference to a mind more or less like ours. The closer the analogies between the behavioral effects, the more tenable the inference to their mental causes.

There are a variety of problems with this sort of "inference". Unlike other arguments by analogy, to the existence and structure of atoms for example, it is impossible in principle to verify its soundness; that is, there is no way in which we can ever verify whether our inference to the mind of another animal is correct. Moreover, the notion of an "analogy" or similarity is of little help, as philosophers of science have long since discovered; there will always be an innumerable number of both similarities and dissimilarities between any two pieces of behavior, and only some prior agreement about what constitutes the *relevant* similarities or dissimilarities will provide

the foundations for an inference. But to assume such prior agreement is already to have "interpreted" the behavior in terms of what it expresses. Most important, the argument from analogy suggests, once again, that the mind is hidden from view, somehow "trapped inside" the body, and only an inference from the "outside" can reach it, an inference which, we have just noted, can never be justified.

But we can justify the attribution of mind, even with respect to species which seem dissimilar to us in many ways. It is a question of "reading" their behavior, in the same sort of way that we read a text. For just as a text does not stand to its author as effect to cause or conclusion to premise, so too the behavior of a grizzly bear is an *expression* of its mind. It is all a question of interpretation. The text is not a "report" on the author's mind whose accuracy, in the nature of the case, can never be verified. In the same way, when we properly interpret some animal's behavior in its spatial and temporal contexts, there is little room for asking, yes, but does this *really* signify a mind? The behavior itself, contextually understood, answers that question.

One might contend at this point that we still do not "really know" what it is like to be a grizzly bear, cannot get "inside" its head. There are two points to be made in this connection. One is that there is no difference in this respect between human beings and other animals, and this is all that is necessary for the purposes of my argument. In the sense indicated, we cannot "get inside" the head of other human beings either, even when they tell us what is going on "inside". Language is no more conclusive than other forms of behavior. The other point is that we can, and often do, "get inside" the heads of both human beings and other animals. We "read" the behavior of those around us all the time, sometimes rightly, sometimes wrongly, through the behavioral texts they supply us with. Reading texts, whether "verbal" or simply behavioral, and entering into the minds are two aspects of one and the same activity. Experienced trainers are especially good at "reading" animals, but anyone who has lived in close proximity to a pet for a number of years becomes good at it out of personal necessity. At the same time, both experienced trainers and perceptive pet owners come to realize that animals come to "read" *their* behavior in a variety of acute ways, occasionally anticipating its sequence in ways that were not originally clear to the trainers and owners themselves.

Now to the DESCARTES–KANT–DAVIDSON line of argument connecting self-consciousness to language. There are four points I wish to make about it.

The first is that only commitment to an otherwise implausible philosophical theory prevents us from seeing that animals make judgments. KANT characterizes the making of judgments in terms of the comparing of "representations". But animals compare representations all the time. How else are we to understand, for example, their indecision when faced, as was BURIDAN's ass, with alternative, and equally attractive, courses of action.

The second point concerns the claim that one cannot make judgments without using language. This claim also is false. We humans can, and for most of the time do, make judgements without engaging any linguistic capacity, although many (although I don't think all) of these judgments can be given a linguistic expression. This is, I think, the case when we make "unconscious" judgments, for example when we reach for the hard rolls instead of the croissants at the breakfast table. But even "conscious" judgments rarely have the determinate character of a sentence and, when pressed to make them determinate, we sometimes have difficulty doing so (not quite being able to find "the right words").

The third point is that self-representation can be ascribed to at least some animals, and then on grounds which have nothing to do with linguistic competence. The pioneering work of GALLUP, for example, has shown that chimpanzees and orangutans can learn to recognize mirror-images as representations of their own and not of another animal's body.[6] If having a concept or sense of self (or being able to "represent" oneself or refer to oneself) involves being able to distinguish between self and other, then these animals clearly have a concept or sense of self, and in this respect are "self-conscious". Are they aware that they are primates, among other things? Presumably not. But "self-consciousness" is necessarily a matter of degree. Certainly none of us humans is completely self-conscious, as individuals or as members of a species.

Not all primates recognize their mirror images. since there is no reason to think that they are any less self-conscious than chimpanzees and orangutans, we have to say that mirror self-recognition is a sufficient, but not also a necessary condition of attributing a minimal sense of self. Members of other species indicate that they have some sense of self in other ways. One which intrigues me is the grizzly habit of seeking out positions from which it can see without being seen. Grizzlies make a special effort to get their entire body behind the available rocks or vegetation, as if they had some (perhaps slight) conception of themselves as wholes. Whether or not the

bear has a language in which to represent and then reflect on its behavior, it would seem to me that concealment begins to reveal a sense of self, just as its capacity for being surprised reveals that it has some notion of the difference between true and false beliefs.

The fourth point is that we can begin to grasp an animal's subjective perception of the situation and thus begin to grasp the distinction that it makes between how things stand with it and how things stand with the world. What is "subjective" is measured by the distance between the world described by the behavioral text that it produces and what might be described as the "real" world. What counts as the "real" world is going to vary; to claim *our* ordinary image of things as reality *would* be unduly anthropocentric. But we do not need to. We will know about an animal's "subjective" picture of things to the extent that it differs, on interpretation, from that produced by the majority of its kind. At the very least, it would take an argument to show that this is not also what we do in the case of human beings.

8. It is thus in philosophy. We make progress by way of showing that certain claims are wrong, in this case the claims advanced by DESCARTES, KANT, and DAVIDSON that animals do not really have a mental life. The arguments for these claims are not persuasive. Now the way is open for us to look more closely at animal behavior in all its richness and to discover once again that it is in many respects much like ours, indicating intelligence and expressing personality. Of course, given the sorry record of our own species on this planet, to say this about other animals is not particularly to flatter them.[7]

Author's address

Gordon G. Brittan, Jr., Department of History and Philosophy, Montana State University, Bozeman, Montana 59717, USA.
Email: uhigb@montana.edu

Notes

1 "Spontaneity" is invoked precisely to indicate that in some cases no explanation can, or need, be given. I understand that the purring of cats is not necessarily spontaneous.

2 If one were to say that the introduction of "want" at this point is question-begging, and that at best the tidbit serves to give the animal pleasure, then I think we would still need to add that the animal wants or desires things that give it pleasure. The fact that something gives an animal pleasure is not by itself enough to motivate the animal to take steps to acquire it.

3 I have changed "presentation" to "representation," the usual translation of KANT's "Vorstellung". Otherwise, the translation is (PLUHAR 1987, p356).

4 For a summary of the evidence that language is unique to humans, see (PINKER 1995), chapter 11.

5 "Expectation" is, of course, an intentional notion and thus might seem to imply without argument that the bear has a mental life. But again, there are objective criteria which, when satisfied, allow us to apply this notion, and it is difficult to see how one might avoid its use. The "hope of eating" which DESCARTES attributes to the magpie is just another way of talking about the bird's expectations.

6 Someone might object: but don't GALLUP's experiments merely show that an animal can recognize its body (and not also that it can distinguish between itself and others)? There is undoubtedly more to self-consciousness than learning to recognize one's own body, but it's a start, just as it is a start that one begins to make a distinction between those things that are under one's control and those that are not, that one begins to take into consideration the nature and limits of one's abilities to effect changes, and so on. At least some animals besides ourselves are going to have to make just these starts on self-consciousness if they are to reproduce themselves successfully. At the same time, the distinctions involved (e.g., between what is subject to my will and what is not) invoke notions of substance, causality, and reciprocity (as KANT held) as well as truth (as DAVIDSON contends) even though languageless animals are of course not able to *state*, much less to analyse, these notions or lay out the truth-conditions for their various beliefs.

7 An earlier version of this paper was read at the College de France in 1997, under the title "La vie mentale des animaux". I'm grateful for their comments on that version to James ALLARD, Christopher PINET, and Jules VUILLEMIN.

References

Davidson, D. (1982) Rational animals. Dialectica (36) 4.

Dennett, D. (1978) Intentional systems. Brainstorms. Bradford Books: Cambridge MA.

Gallup, G. G. Jr. (1983) Towards a comparative psychology of mind. Melgren, ed., Animal Cognition and Behavior. North-Holland: Amsterdam.

Kenny, A. (1970) Descartes: Philosophical Letters. Clarendon Press: Oxford.

Pinker, S. (1995) The Language Instinct. Harper: New York.

Pluhar, W. (1987) Kant's Critique of Judgment. Hackett Publishing Company: Indianapolis.

Gordon M. Burghardt

Conceptions of Play and the Evolution of Animal Minds

Abstract

Issues of the origins of play must be separated from issues of the functions of play in animal lives. A comparative analysis of the role of play in the evolution of animal minds, however, is only as successful as the conception of play that is used. Here a set of five criteria for recognizing playful acts wherever they may occur phylogenetically is presented. In addition, primary processes deriving from life history, physiological, and psychological factors are distinguished from derived secondary processes. The latter may be the source of the behavioral, cognitive, and emotional changes resulting from play. Surplus Resource Theory is described as a means of accommodating both primary and secondary processes in play and also aids in predicting the conditions under which playlike behavior is observed. Brain size and intelligence are not necessarily linked with play, nevertheless, playful behavior may be linked to creativity and novel behavior throughout both evolutionary history and individual lives.

Key words

Play, evolution, cognition, creativity, development.

THE FOLLOWING THREE quotations from a century ago set the stage for the topic of this essay.

"A satisfactory theory of play is still wanting, and yet a man does not learn through any kind of instruction or study in later life anything like so much as the child learns in the first four years of his careless existence, through the perceptions and ideas acquired in his play" (PREYER 1893, p42).

"Science has shown that the embryonic period of physical development is a masquerade of long-vanished forms of life. In like manner the children of each new generation seek instinctively to revive the life that is behind them and in their favorite occupations and amusements re-enact the prehistoric experiences of mankind... Can we doubt that the order of history should be the order of education, and that before we teach the child to read and write we should aid his efforts to repeat in outline the earlier stages of human development" (BLOW 1894, pp125–126).

"Play being the all important business of childish life, and all play consisting more or less in acts, it is play then, above all, that we must seek for the beginnings of voluntary activity" (COMPAYRÉ 1902, p151).

The search for understanding animal mental life and cognitive abilities became a scientifically critical issue in the aftermath of the DARWINIAN revolution. DARWIN's writings legitimized evolutionary approaches to animal diversity, provided an easily comprehended process of operation (natural selection), and made a strong case for including behavioral and psychological processes as key elements of a unified evolutionary biology. Although later downplayed by many biologists themselves, philosophical and psychological issues were the critical goads to DARWIN in his compulsion to explain the origins of species. This is evident in the notebooks he kept in the 1830s and 1840s.

Although DARWIN (1859) developed detailed explanations for the origin of many instinctive behaviors in wild and domesticated animals (e.g., slavemaking in ants, hivemaking in bees, pointing in hunting dogs, and pouting in pigeons) he was less successful in explaining the origin of the complex mental life of the "higher" animals and people. His two books dealing with psychological topics (DARWIN 1871; 1872) focused much more on the similarities between human behavior and that of other mammals than on the origin of the comparable abilities. Indeed, although the evolution of intelligence and mental abilities were a prime interest of the early POSTDARWINIAN writers, play did not figure prominently in this picture (e.g., ROMANES 1883; MORGAN 1894). By the time of Mark BALDWIN's great *Dictionary of Philosophy and Psychology*, however (BALDWIN 1902), the importance of play in the origins of diverse creative and artistic expressions was seriously discussed. This

aspect of play was one of the few in which the three major early play theorists, Herbert SPENCER (1872), Karl GROOS (1901), and G. Stanley HALL (1904), could all agree. In human developmental psychology the role of play in the early learning and mental life of infants and young children was also widely accepted during this period (e.g., PREYER 1893; BLOW 1894; COMPAYRÉ 1902) as suggested by the opening quotations.

In the current century many formulations of the role of play in the development of virtually every aspect of human behavior have been advanced (HARTLEY/FRANK/GOLDENSON 1952). This interest intensified after mid-century, especially following the influential lead of PIAGET (e.g., 1962), the list of recent books is long and diverse (e.g., PIERS 1972; GÖRLITZ/WOHLWILL 1987). Theoretical attempts to derive art from play have likewise intensified (HUIZINGA 1949; ALLAND 1977; DISSANAYAKE 1992). For over a century, then, claims have been made concerning the role of play in the mental life and behavior of human and, less often, nonhuman animals. These claims, derived from theory, have remained controversial and largely unsupported with empirical data (MARTIN/CARO 1985; SMITH 1988, 1996). Empirical support that is advanced is often correlational or experimentally flawed. Nevertheless, since the time of FROEBEL and the Kindergarten movement (BLOW 1894; FROEBEL 1895), many countries have included play in daycare and educational systems as a means of facilitating learning. Claims have been made that play facilitates learning (including reading and mathematics), imagination, socialization, behavioral flexibility, mental agility, and creativity. Evolutionary arguments based on selected comparative data often accompany these assertions (c.f., FAGEN 1984). Certainly, such claims, in spite of the lack of experimental support, have been viewed as plausible by many who have studied play. Even those most responsible for disconfirming experiments find it difficult to completely dismiss these putative benefits (e.g., SMITH 1996).

To make progress in understanding the role of playfulness in cognitive life we need to ground the phenomena of play in a phylogenetic context. What are the processes leading to playfulness throughout vertebrate evolution? This question is prior to the search for putative consequences of play. Elsewhere I am preparing an extended treatment of animal play, and the historical, definitional, theoretical, and comparative aspects will be explored in greater depth and with supporting data and citations (BURGHARDT, in press). By focusing on the origins and evolution of play I think it possible to outline a testable model of both how playful activities arise and how they can shape the future behavior patterns of animals, including facilitating complex mental processing. I will present a way of looking at play as a potential source of the rapid evolution of behavioral complexity in endothermic vertebrates and as a resource for ontogenetic change in behavior that can be both conservative and novel.

In order to evaluate the role of play in creating or influencing the animal mind we must first identify the phenomena to be considered playful. This is not a trivial exercise, especially if one wants to discuss the origins of behavioral phenotypes (BURGHARDT 1998a, b). Sometimes even the most accepted and clear examples of play can be problematic. The traditional categories of play are locomotor, object (including predatory), and social; yet these are frequently not distinct. In order to compare play across species we must have some way of isolating play behavior using a definition that is not specific to any one kind of play (such as social play). Another important problem is that the causes and functions of play may differ for different kinds of play, differ throughout ontogeny, and differ across species. With humans the problems of characterizing play often seem insurmountable. Some have argued that play is a romantic notion of middle-class Western industrialized societies deriving from an artificial work-play dichotomy (SUTTON-SMITH/KELLY-BRYNE 1984), others have held that play is the real 'work' of other species (GROOS 1898)! Playfulness, conceptualized as participating in games, has been viewed as intrinsic to war, competition, and rituals of all types (cf. HUIZINGA 1949). Nursery teachers "adhere to the term 'play' to describe any or every activity of young children (excepting biological necessities such as eating, sleeping, and elimination)" (HUTT/TYLER/HUTT/CHRISTOPHERSON 1989, p221). Such rhetorical ploys, along with legitimate difficulties in defining, describing, and analyzing playful behavior, has contributed to the lack of serious and careful scientific attention to the role of play in development and evolution. Thus we first need to agree on how to recognize play.

Defining play

After reading many treatments of play one is probably best advised to throw up one's hands and say, play just cannot be usefully defined. E. O. WILSON (1975, p164) has written: "No behavioral concept

has proved more ill-defined, elusive, controversial, and even unfashionable." More recently, MITCHELL (1990, p197) has averred that "Play is the hobgoblin of animal behavior, mischievously tempting us to succeed in what, judging from the number of failed attempts, seems a futile task: defining play." Nevertheless, MITCHELL and many leading play researchers have advanced their own definitions. FAGEN (1981) has gathered together many of these. None of these list the minimal criteria necessary to set play apart from other phenomena in a rather precise way. Here I cannot review even the most recent and influential definitions of animal play in any depth. For illustrative purposes, however, I will quote what is arguably the currently most widely cited definition, one that I have used repeatedly myself (BEKOFF/BYERS 1981, p300):

"*Play* is all motor activity performed postnatally that appears to be purposeless in which motor patterns from other contexts may often be used in modified form and altered temporal sequencing."

Note that the use of "may" hedges our ability to apply the definition to phenomena that we do not already consider play; in actuality we are left with the apparent lack of purpose in the target behavior as the central and essential component. Stereotyped dysfunctional abnormal behavior such as feather-picking in birds cannot be excluded in this definition as BEKOFF/ALLEN (1998) acknowledge. The modification of this definition by MARTIN/CARO (1985) only compounded this problem.

In fact, the bottom line of most definitions is that the behavior seems to lack an immediate function or purpose. Some writers follow theorists such as PIAGET (1962) and list individual criteria and pronounce each one inadequate in defining play as the category we think it to be. But what if we package the many different criteria into sets of similar type and see if we can now specify those that are essential? Since for my purposes I need to identify possible candidates for play in a wide diversity of species, I extracted from the literature on animal and human play, and my own 35 years of behavioral study, a set of diverse criteria and arranged them in a formal and definitive manner. Here I present for, the first time in print, the following five criteria, *all of which must be satisfied in at least one respect* before a given behavior is acceptable as an instance of play. It is important to note that none of them alone are sufficient for applying the play designation.

The *first criterion* for recognizing play is that the performance of the behavior does not appear completely adaptive in the context in which it is expressed; that is, it includes elements, or is directed toward stimuli, that do not seem to contribute to current survival.

The *second criterion* for recognizing play is that the behavior appears to be spontaneous, voluntary, intentional, pleasurable, rewarding or autotelic ("done for its own sake").

The *third criterion* for recognizing play is that it differs from the "serious" performance of ethotypic behavior in at least one respect: incomplete (generally through inhibited or dropped final elements), exaggerated, awkward, precocious, or involves behavior patterns with modified form, sequencing or targeting.

The *fourth criterion* for recognizing play is that it be repeatedly observed in a similar, but not rigidly stereotyped, manner, during at least a portion of the animal's ontogeny.

The *fifth criterion* for recognizing play is that it is observed when the animals are adequately fed, healthy, and free from stress (e.g., predator threat, harsh microclimate, crowding, social instability), or intense competing motivations (e.g., feeding, courtship, resource competition, nestbuilding): In other words, the animals are in a "relaxed field." Note, however, that this does not mean that the play itself is free from stress or risk.

Now while these criteria do not provide a nice crisp one line definition, they seem to cover every accepted example of play and to exclude, with consistent reasoning, much of the problematic and controversial. It is not limited to social play, or juvenile play, or play that involves special signals. It excludes stereotyped abnormal behavior as well as exploration. Although much behavior called play is vigorous and energetic from a human perspective, this also is not required. The apparent fuzziness of the boundaries in applying individual criteria in some cases reflects the fact that playfulness is on a continuum with a variety of nonplay behaviors. In fact, the relationship of play to both exploratory behavior and curiosity (BURGHARDT 1984; HUTT et al. 1989) and stereotyped behavior need to be more fully resolved if we hope to make progress in understanding both the phylogeny and ontogeny of play. But if all the above criteria are met the behavior deserves to be called play, a distinct category.

Note the following about the above criteria. They make no assumptions about the function of play. They avoid the controversy of whether or not there is a separate motivation underlying play. By including terms that some find essential (pleasure, intentionality, voluntary) but which other scientists find

unclear, unscientific, or difficult to apply (as in a fish), the different conceptual and labeling systems scientists use can be accommodated. I am confident that if new behavioral labels arise they can be readily added to one of the five criteria sets. Of course, the reliability of the observations is also important, especially in the more exotic and less-studied animals or those that have been generally dismissed as not playing (e.g., ants, reptiles). Thus independent observations by several qualified investigators and their documentation are important in firmly accepting claims for some species.

The end result of this endeavor has been that, for me, the five criteria have been particularly valuable in evaluating claims and evidence for play in problematic groups such as ants, fishes, monotremes, and reptiles (e.g., BURGHARDT 1998b). They can also help guard against the anthropomorphism that has been endemic to the study of animal play since its inception. However, *critical anthropomorphism* (BURGHARDT 1985; 1997) can be used to retain the motivational and hedonic aspects of play without losing an objective and public assessment of play. These latter aspects of play will become clarified as physiological and neural measures of internal states become more refined (BURGHARDT 1995).

It is also important to note that there is no implication that all behaviors satisfying the criteria for being classified as play are necessarily the same in their causal (neurobiological) mechanisms or that they are homologous. Different species will play in different ways depending on their normal behavior, ecology, ontogenetic development, metabolism, neural organization, and phylogeny. In fact, it is because play is a heterogeneous category that arises repeatedly in vertebrate evolution that we need such a litmus test. The criteria help us to sort out the processes and variables influencing playfulness and point the way to experiments and critical phyletic comparisons.

Now that we have (hopefully) criteria that allow us to confidently identify play and target missing or less known aspects of the putative play, what can we do to explore the role of play in the evolution of animal mentality? One approach would be to classify the kinds of play in a species and study them closely to see what they contribute to cognitive and affective life. An alternative, followed here, is to see if there are commonalities in the life history, physiology, behavior, and psychology of the animal groups in which play has evolved and whether these commonalities can help us understand the evolution of animal mentality.

Distinguishing between primary and secondary processes issues in the study of play

Play may be both a product and cause of evolutionary processes; that is, playful activities may be both sources of enhanced behavioral and mental improvements as well as a byproduct or remnant of prior evolutionary and ontogenetic events. It is the conflict between these two views of play that I view as being at the root of the paradoxical nature of play. The early conflicts between the functional view of GROOS (1898) and the recapitulation view of HALL (1904) were over misunderstandings of these very issues, and the legacy of this controversy still haunts the study of play. To break the impasse and make genuine progress, we need to keep conceptually separate the *primary* processes by which "play" originated and evolved in ancient vertebrates and their modern descendants and the ways in which these may have provided an essential scaffolding from which the highly diverse and complex structures of mammalian play could be built through a series of *secondary* or derived processes. It is these derived processes that may provide cognitive advantages to those animals engaging in play. But it is a mistake to think that play originally evolved in order to provide such advantages.

Without going into all the supporting details, I will approach play in animals from an evolutionary/developmental perspective called Surplus Resource Theory or SRT. This view has several components including a reliance on studying animals, including ectothermic vertebrates, with less rich play in contrast to the exclusive focus on highly playful taxa (primates, canids, felids, ungulates) usually pursued. SRT incorporates physiology (e.g., activity metabolism, thermoregulation), life history (e.g., parental care, altriciality, food niche), and psychology (e.g., level of stimulation, learning). As in DARWIN's discovery of natural selection, domestication has produced some intriguing parallels to what may have taken place during vertebrate phylogeny.

The next move is to distinguish between primary and secondary processes in play. It is the latter that are most often viewed as leading to the claims made in the first paragraph. An example of a secondary process would be that vigorous rough and tumble play of young rats and dogs enhances adult performance, promotes socialization, or increases behavioral flexibility. However, it is precisely because the more primary processes of play have been ignored that predictions made from secondary processes have fared

poorly. An example of an important primary process derived from SRT would be the role of metabolic rate or parental care in production of 'surplus' behavioral "mutants" that could in turn be selected ontogenetically and phylogenetically. A secondary process derived from a detailed consideration of primary processes would be the recent correlation, by BYERS/WALKER (1995), of the onset of vigorous motor play with the period in ontogeny with the timing of permanent long-term changes in the muscular and cerebellar systems of several species of domesticated animals. With children, the most famous secondary process would be PIAGET's conception of assimilation and accommodation as the processes by which children cognitively adapt to their environment (PIAGET 1962). Play, especially during the early sensorimotor stage, is essential for this process. In this early stage reflexes and circular reactions are the basic building blocks. Piagetian processes have been explored in animals, especially nonhuman primates. However, while study of these stages can help uncover the ontogenetic origins of mental life, they tell us little about how these processes have been shaped by evolution to have these essential roles. Interestingly, PIAGET himself (PIAGET 1972, p27) does not see his work has having any significant role in education, stating "I am not an educator. I have no advice to give." He does state that active methods involving self-discovery by the child seem most useful for long-term retention.

In short, my approach is more basal than PIAGET's with children or the functional studies on highly derived play in mammals. The study of the role of play in the evolution of animal minds should initially focus on the primary processes leading to behavior satisfying the five criteria above. Some writers have argued that play is a "random process generator" and similar views abound. A more sophisticated version is FAGEN's (1974) comparison of the variability seen in play to that found in genetic systems. That is, in play we see similarities to chromosomal inheritance in the sense that in play sequences we can we find *recombination*, *fragmentation*, *translocation*, and *duplication*. Such variation can be the raw materials for natural selection to operate upon insofar as the behavioral variation also has some inherited component. However, play is far from random and it seems to operate within fairly tight boundaries. PELLIS (1993) has documented how play fighting can differ among closely related species with little or no overlap. Thus the challenge is not just to state that play creates novel behavioral phenotypes but to uncover the actual behavioral processes underlying such behavioral variation.

The phylogenetic origins of play

Unlike in the nineteenth century, in this century play has been viewed as primarily a property of mammals. GRIFFITHS (1978) even listed play as one of the defining attributes of mammals and thus a reason why monotremes should be considered true mammals and not "quasi-mammals". Birds have been acknowledged as engaging in play (reviews in FICKEN 1977; FAGEN 1981; ORTEGA/BEKOFF 1987). However, a clear discontinuity was recognized in that play was considered common in the endothermic vertebrates (especially mammals) and absent in the traditional "cold-blooded" vertebrates, including fish, amphibians, and nonavian reptiles (reptiles from here on). As someone who has spent decades studying reptile behavioral ontogeny, and being impressed with the degree of complexity and learning in their behavior (BURGHARDT 1977), I had to agree that I had seen nothing that appeared as playful. Why? I developed a list of life history, physiological, and psychological factors that appeared to be different between the typical reptile and the typical mammal. Although the full array of evidence underlying these differences is complex (BURGHARDT 1984; 1988) the major premises can be quickly summarized.

First, except for crocodilians, most neonatal reptiles are not cared for by their parents and consequently must devote their activities, immediately upon birth or hatching, toward obtaining food, finding shelter, avoiding predation, and growing rapidly on their own. As a result there is limited "safe" time or opportunity for practicing or perfecting behaviors to be used in an uncertain future. Most young reptiles are very small and selection will have shaped abilities that enhanced juvenile survival, such as remaining quiet and inconspicuous. Such behaviors are, of course, incompatible with vigorous play. There is also no source of high fat, high protein nourishment at little cost, as in juvenile birds and mammals. Nonavian reptiles are highly precocial and *superficially* resemble, and behave like, miniature adults. This is not to say that young reptiles do not learn many things as they go about their serious activities. Furthermore, reptiles grow as rapidly as food resources allow; unlike mammals and birds there is no necessary age defined period of youth. Young reptiles, unlike young mammals, do not get fat; they just grow fast. This characteristic of reptiles is termed nondeterminate growth. The last life history characteristic I will mention is that reptiles typically have large litter

and clutch sizes with high juvenile mortality. Thus delayed benefits of play would be less worth a current risk.

Second, reptiles typically are metabolically constrained from performing vigorous, energetically expensive behaviors, especially those of any duration. This is due to their low resting and maximal metabolic rates, limited aerobic capacity, and long recovery times after anaerobic expenditures. Reptiles have about 10% the metabolic rate of a comparably sized mammal. But, in fact, most reptiles are much smaller than mammals. Both metabolic rates and typical body sizes are further reduced in amphibians. Small body size leads to greater heat loss and even more efforts to behaviorally thermoregulate. Taken together, these factors constrain the performance, and selection for, vigorous playlike behavior with no current function (BURGHARDT 1984).

Third, animals in boring, unstimulating environments would be most likely to engage in behavior to relieve sensory and response deprivation and to increase arousal. Such boredom might be expected in the well provisioned and protected environments provided by endotherm parents. This factor may be the critical one in the consistent finding that well-cared for captive animals play much more than their wild counterparts. If the juvenile animals being buffered from the demands of survival have a complex repertoire of evolved and active behavior patterns, then they have a behavioral resource to draw on when deprived of stimulation.

As a guide to identifying behavior in nonavian reptiles that could be candidates for traditional appearing play, I subsequently (BURGHARDT 1988) used the above and related physiological, psychological, and life history processes, to predict that mammalian or avian-like play in reptiles, whether locomotor, object, or social, should be rare and occur only in specific contexts in which those factors facilitating play in mammals and birds are also present. A series of predictions was supported that suggest that the three factors listed above are involved in facilitating play. What they allow us to do is to predict in what mammalian groups we would expect to find the most complex play and the most time spent in play. For example, on energetic grounds one would expect that aquatic mammals would be particularly playful—and they are. The play we have found in turtles has been in aquatic turtles (BURGHARDT/WARD/ROSSCOE 1996; BURGHARDT 1998b).

Object play is found most often in predators and generalist feeders that rely on manipulation of limbs and mouth. Again, reptiles do not have the rich repertoire of possible movements of limbs and face seen in many mammals. The lack of parental care may have prevented the evolution of social bonding and affiliation to the extent seen in birds and mammals. Certain kinds of complex ecological and social interactions and operations upon the environment seem to characterize the more playful species. Interestingly it is those reptiles with more variable diets and foraging techniques in which more exploration, curiosity, and, rarely, play, are found.

The chief difference of this approach from previous ones is that the initial advantages of incipient playlike behavior did not involve any particular functions (such as perfecting later behavior, increasing endurance, and incorporating flexibility). Play can evolve independently whenever physiological (including neural), life history, metabolic, and psychological conditions, in conjunction with a species' behavioral repertoire, reach a threshold level. In mammals, however, play was also a disparate constellation of *ad hoc* mechanisms that could be used to maintain the continuity of endothermic and behavioral systems bridging the periods of juvenile dependence and adult responsibilities. The deterioration of certain neonate response systems through parental care, the lowering of thresholds and the broadening of the effective stimuli, and the increased aerobic metabolic capacities resulting from endothermy led to a reorganization of developmental processes so that play, as well as other experiential avenues, was not only available to homeotherms but may well have had to be exploited by them for continued survival, by replacing lost, suppressed, or maturationally delayed "instincts."

The role of play in the origins of "higher" vertebrate minds

If I am correct, the rapid evolution of cognitive capacities in mammals is due to processes initiated by the advent of parental care and the opportunity to use the surplus resources of time, energy, and innate propensities already present from a precocial existence. This process probably took initially many millions of years and was rather slow in coming. Many species and radiations of mammals and birds developed along these lines and today we see snapshots of where they are along a continuum of increasing amounts of such playfulness. Play behavior, as defined by the 5 criteria, became more prominent. First it was an epiphenomenal byproduct that may have been produced by the surplus

energy process postulated by SPENCER (1872), but almost immediately could be utilized as a source of behavioral neophenotypes.

Play can be viewed as providing opportunities that can be exploited in adapting to changing physiological, environmental, and developmental conditions. Perhaps in species in which play is prominent it is a directed, even intentional, source of adaptive variability. I actually think this is possible only rarely. First, play, especially social play, can itself become highly ritualized, have important signal functions (BEKOFF 1995), and show consistent species and age-related patterns (PELLIS 1993; BURGHARDT 1998a, b). This may give individuals opportunities to learn about conspecifics and how to respond to them. It may provide opportunities to exploit others. But it is also possible that social play is in large measure a response to social needs and motivational system legacies from when such behaviors needed to be performed for serious and immediate functional reasons. In this way HALL's (1904) recapitulation view has some merit once the rigid sequencing is eliminated as necessary. Only in play's most derived state does it do the kind of service that the prevailing practice theory, that goes back to GROOS (1898), requires.

Play and intelligence are not necessarily linked

Support for the role of play in cognition is often advanced in the following way. Play is found most commonly in birds and mammals that have large brains and are considered "intelligent" (FAGEN 1981). Play and curiosity are thus marks of intelligence and necessary for the development of a sophisticated mind. The rough correlations are certainly impressive. Primates and aquatic mammals have large brains and are considered highly playful, as are the big-brained birds in the parrot and corvid families (FAGEN 1981). Sloths, insectivores, and many less playful mammals and birds have smaller brains. The best evidence for fish play is in mormyrid fish (e.g., MEDER 1958) that have huge brains larger, relatively, than most mammals (BUTLER/HODOS 1996) and may even exceed the brain/body weight ratio of adult human beings (BULLOCK 1977)! Furthermore, animals with more altricial young play more or more complexly than even close relatives that are more precocial.

Yet there are some problems with this putative relationship. Dogs raised in social isolation appear to be far more social and curious than normally reared dogs, especially to novel situations; they are less neophobic. Actually, it has been shown that such dogs are rather stupid; even the most trivial items amuse them persistently (HEBB 1972). There is also considerable evidence that habituation, rapid boredom with stimuli, is a good measure of intelligence and can be used with preverbal infants with considerable success (COLOMBO 1993). If becoming bored is a sign that one has mastered what there is to know or do with an object, then choosing stimuli that are more complex might be the way to deal with this situation. This is exactly what Piagetian mastery play is directed towards. THOMPSON (1998) has advocated a similar process in animals she terms self-assessment. But curiosity killed the cat. Rodents, fish, and tadpoles that stray too far from secure retreats have all been shown to be more at risk of predation. Thus, prudent behavior has some advantages over the exuberant "joy of life" excitement many view as the essence of play.

But perhaps the most telling argument against a simple brain-size and intelligence relationship with play can be found by using domestication as a model system for testing surplus resource theory. Many of the processes that I postulate as having occurred in juvenile mammals with the onset of parental care and the consequent buffering from the demands of life, are also found in domesticated species. PRICE (1984) has documented these processes completely independently of the theory outlined here. The match is remarkable. The retention of juvenile morphology and behavior into adulthood is seen particularly clearly in dogs. Litter sizes increase and many behavioral skills shown in wild populations show deterioration. But playfulness increases! Brain size decreases! KRUSKA (1987) has shown that in many mammals domestic populations have brain sizes 5% less than wild populations after only a few generations. This is shown in both a rodent (bank vole) and carnivore (polecat). Ferrets were domesticated from wild polecats 2500 years ago and now have brain sizes 30% smaller than wild animals. It would seem that a careful study of play behavior in wild animals undergoing domestication might be a most useful method to see how buffering animals from the harsher aspects of existence may change the amount, type, and frequency of play as well as cognitive capacities. It is important that critically anthropomorphic methods be used. We may find it all too easy to conclude that the more affectionate or compliant dog is smarter than the more elusive, unpredictable high-strung wolf or devious coyote.

Back to creativity

The view of play and its origins outlined here suggests a rather slow development through out vertebrate evolution of playlike behavior under suitable ecological and physiological conditions. Using the above criteria for play, I think there is convincing evidence that play evolved independently in several lines of fishes, turtles, squamate reptiles, crocodilians, birds, and marsupials, as well as virtually all families of placental mammals (BURGHARDT, in press). Endothermy and parental care interacted with this early incipient play and gave it new roles and functions, a process still ongoing. The rapid rate of evolutionary changes in endothermic animals, especially mammals, in genome size, brain size, and behavioral complexity are remarkable and still little understood. Play may have been a major engine in this rapid cascade of evolutionary change that led to increased cognitive complexity. This may occur by moving initially playful responses to more serious endeavors and functions so that the once playful behavior has been transformed and is now outside of the realm of play using the criteria developed above. Thus a paradox: play may be essential in the evolution of mind but what remains as play may not be as valuable as often believed. Much play leads nowhere. We have to avoid the idealization of play, as SUTTON-SMITH/KELLY-BYRNE (1984) argued, but for different reasons than they advanced. I have concluded that after a slow period of evolutionary reorganization in behavioral ontogeny (BURGHARDT 1988), play facilitated rapid behavioral and mental development.

There are some intriguing parallels between recent work exploring creativity by individual human beings (CSIKSZENTMIHALYI 1996) and the role of play facilitating rapid evolutionary creativity. Creativity is often evidenced by periods of information gathering, rehearsal, and consolidation before a rapid productive phase, and this can be simulated in a neural model (MACLENNAN, in prep.). Once begun, creative achievements may flow rapidly. Thus like the model of play, after a slow process "up a mountain" a turning point is reached at the top and the trip down is much faster than that going up. We can also extend this to cultural change where, after an innovation slowly spreads across a population, a point is reached in which the society is rapidly and irreversibly transformed. This seems to have been the case with the automobile, air travel, the telephone, and, most recently, the computerization of many of our activities. But these changes, due to the efforts of countless creative individuals, also have their origins in the playful tinkering and symbolic rehearsals that derive from processes underlying the criteria for recognizing play in even the "lowliest" animals! Tracing out these connections is a most exciting and significant challenge.

Author's address

Gordon M. Burghardt, University of Tennessee, Knoxville, Tennessee 37996-0900, USA
Email: gburghar@utk.edu

Note

Based on a talk given at the "Animal Mind" workshop, Institute for Brain Research at the University of Bremen, Germany, September 19, 1997.

References

Alland, A. Jr. (1977) The artistic animal: an inquiry into the biological roots of art. Anchor Press: Garden City NY.

Baldwin, J. M. (ed) (1902) Dictionary of Philosophy and Psychology. Macmillan: New York.

Bekoff, M. (1995) Play signals as punctuation: The structure of social play in canids. Behaviour 132: 419–429.

Bekoff, M./Allen, C. (1998) Intentional communication and social play: how and why animals negotiate and agree to play. In: Bekoff, M./Byers, J. A. (eds) Animal play: Evolutionary, comparative, and ecological perspectives (pp. 97–114). Cambridge University Press: Cambridge.

Bekoff, M./Byers, J. (1981) A critical reanalysis of the ontogeny and phylogeny of mammalian social and locomotor play: an ethological hornet's nest. In: Immelmann, K./Barlow, G. W./Petrinovich, L./M. Main (eds) Behavioral development: the Bielefeld interdisciplinary project (pp. 269–337). Cambridge University Press: Cambridge.

Blow, S. E. (1894) Symbolic education. A commentary on Froebel's "Mother play". D. Appleton: New York.

Bullock, T. H. (1977) Introduction to nervous systems. San Francisco: Freeman, W. H.

Burghardt, G. M. (1977) Learning processes in reptiles. In: Gans, C./Tinkle, D. (eds) The biology of the Reptilia. Vol. 7, ecology and behavior. Academic Press: New York, pp. 555–681.

Burghardt, G. M. (1984) On the origins of play. In: Smith, P.

K. (ed) Play in animals and humans. Basil Blackwell: London, pp. 5–41.

Burghardt, G. M. (1985) Animal awareness: Current perceptions and historical perspective. American Psychologist 40: 905–919.

Burghardt, G. M. (1988) Precocity, play, and the ectotherm-endotherm transition: Superficial adaptation or profound reorganization? In: Blass, E. M. (ed) Handbook of behavioral neurobiology. Vol. 9. Plenum: New York, pp. 107–148.

Burghardt, G. M. (1995) Brain imaging, ethology, and the nonhuman mind. Behavioral and Brain Sciences 18: 339–340.

Burghardt, G. M. (1997) Amending Tinbergen: A fifth aim for ethology. In: Mitchell, R. W./Thompson, N. S./H. Miles, L. (eds) Anthropomorphism, anecdotes, and animals. SUNY Press: New York, pp. 254–276.

Burghardt, G. M. (1998a) The evolutionary origins of play revisited: Lessons from turtles. In: Bekoff, M./Byers, J. A. (eds) Animal play: Evolutionary, comparative, and ecological perspectives. Cambridge University Press: Cambridge, pp. 1–26.

Burghardt, G. M. (1998b) Play. In: Greenberg, G./Haraway, M. (eds) Comparative psychology: A handbook. Garland: New York, pp. 757–767.

Burghardt, G. M. (in press) The genesis of animal play. Cambridge, MA: MIT Press.

Burghardt, G. M./Ward, B./Rosccoe, R. (1996) Problem of reptile play: Environmental enrichment and play behavior in a captive Nile soft-shelled turtle, (Trionyx triunguis). Zoo Biology 15: 223–238.

Butler, A./B./Hodos, W. (1996) Comparative vertebrate neuroanatomy: Evolution and adaptation. Wiley: New York-Liss.

Byers, J. A./Walker, C. (1995) Refining the motor training hypothesis for the evolution of play. American Naturalist 146: 25–40.

Colombo, J. (1993) Infant cognition: Predicting later intellectual functioning. Sage: London.

Compayré, G. (1902) Development of the child in later infancy. (Translated by Mary E. Wilson.) D. Appleton: New York.

Csikszentmihalyi, M. (1996) Creativity: Flow and the psychology of discovery and invention. HarperCollins: New York.

Darwin, C. (1859) On the origin of species. Murray: London.

Darwin, C. (1871) The descent of man and selection in relation to sex. Murray: London.

Darwin, C. (1872) The expression of the emotions in man and animals. Murray: London.

Dissanayake, E. (1992) Homo aestheticus: Where art comes from and why. Free Press: New York.

Fagen, R. M. (1974) Selective and evolutionary aspects of animal play. American Naturalist 108: 850–858.

Fagen, R. M. (1981) Animal play behavior. Oxford University Press: New York.

Fagen, R. M. (1984) Play and behavioral flexibility. In: Smith, P. K. (ed) Play in animals and humans.Basil Blackwell: London, pp. 159–173.

Ficken, M. S. (1977) Avian play. Auk 94: 573–582.

Froebel, F. (1895) The mottos and commentaries of Friedrich Froebel's Mother Play. (Translated by S. E. Blow and H. R. Eliot.) D. Appleton: New York.

Görlitz, D./Wohlwill, J. F. (eds) (1987) Curiosity, imagination, and play. Lawrence Erlbaum Associates: Hillsdale NJ.

Griffiths, M. (1978) The biology of the monotremes. Academic Press: New York.

Groos, K. (1898) The play of animals. (Translated by E. L. Baldwin.) D. Appleton: New York.

Groos, K. (1901) The play of man. (Translated by E. L. Baldwin.) D. Appleton: New York.

Hall, G. S. (1904) Adolescence: Its psychology and its relations to physiology, anthropology, sociology, sex, crime, religion and education. D. Appleton: New York.

Hartley, R. E./Frank, L. K./Goldenson, R. M. (1952) Understanding children's play. Columbia University Press: New York.

Hebb, D. O. (1972) Textbook of psychology. 3rd edition. W. B. Saunders: Philadelphia.

Huizinga, J. (1949) Homo ludens. (Translated by R. F. C. Hull.) Routledge: London.

Hutt, S. J./Tyler, S./Hutt, C./Christopherson, H. (1989) Play, exploration and learning: A natural history of the preschool. Routledge: London.

Kruska, D. (1987) How fast can total brain size change in mammals? Journal für Hirnforschung 28: 59–70.

Martin, P./Caro, T. M. (1985) On the function of play and its role in behavioral development. Advances in the Study of Behavior 15: 59–103.

Meder, E. (1958) Gnathonemus petersii (Günter). Zeitschrift für Vivaristik 4: 161–171.

Mitchell, R. W. (1990) A theory of play. In: Bekoff, M./Jamieson, D. (eds) Interpretation and explanation in the study of animal behavior. Vol. 1: Interpretation, intentionality, and communication. Westview Press: Boulder, pp. 197–227.

Morgan, C. L. (1894) An introduction to comparative psychology. Walter Scott: London.

Ortega, J. C./Bekoff, M. (1987) Avian play: Comparative evolutionary and developmental trends. Auk 104: 338–341.

Pellis, S. M. (1993) Sex and the evolution of play fighting: A review and model based on the behavior of muroid rodents. Play Theory and Research 1: 55–75.

Piaget, J. (1962) Play, dreams and imitation in childhood. (Translated by C. Gattegno and F. M. Hodgson) W. W. Norton: New York.

Piaget, J. (1972) Some aspects of operations. In: Piers, M. W. (ed) Play and development. W. W. Norton: New York, pp. 15–27

Piers, M. W. (Ed.). (1972) Play and development. W. W. Norton: New York.

Preyer, W. (1893) Mental development in the child. (Translated by H. W. Brown.) D. Appleton: New York.

Price, E. O. (1984) Behavioral aspects of domestication. Quarterly Review of Biology 59: 1–32.

Romanes, G. J. (1883) Mental life of animals. Kegan, Paul, Trench, Trübner & Co: London.

Smith, P. K. (1988) Children's play and its role in early development: A reevaluation of the 'play ethos'. In: Pellegrini, A. D. (ed) Psychological bases for early education (pp. 207–226). Chichester: John Wiley and Sons.

Smith, P. K. (1996) Play, ethology, and education: A personal account. In: Pellegrini, A. D. (ed) The future of play theory (pp. 3–21). Albany: SUNY Press.

Spencer, H. (1872) Principles of psychology. (second ed.). (Vol. 2). D. Appleton: New York.

Sutton-Smith, B./Kelly-Byrne, D. (1984) The idealization of play. In: Smith, P. K. (ed) Play in animals and humans (pp. 305–321). Oxford: Basil Blackwell.

Thompson, K. V. (1998) Self-assessment in juvenile play. In: Bekoff, M./Byers, J. A. (eds) Animal play: Evolutionary, comparative, and ecological perspectives (pp. 183–204). Cambridge University Press: Cambridge.

Wilson, E. O. (1975) Sociobiology: The new synthesis. Cambridge: Belknap Press.

Colin Allen

Animal Concepts Revisited[1]

The Use of Self Monitoring as an Empirical Approach

Abstract

Many psychologists and philosophers believe that the close correlation between human language and human concepts makes the attribution of concepts to nonhuman animals highly questionable. I argue for a three-part approach to attributing concepts to animals. The approach goes beyond the usual discrimination tests by seeking evidence for self-monitoring of discrimination errors. Such evidence can be collected without relying on language and, I argue, the capacity for error-detection can only be explained by attributing a kind of internal representation that is reasonably identified as a concept. Thus I hope to have shown that worries about the empirical intractability of concepts in languageless animals are misplaced.

Key words

Animal concepts, animal learning, error discrimination, language.

A PIGEON, TRAINED TO peck for a food reward in response to pictures of trees, generalizes to new instances, responding appropriately to pictures that it has never seen before (HERRNSTEIN et al. 1976). Does it have a concept of TREE? A pig, trained to respond to a pair of objects based on whether they are similar in shape and color, also generalizes to object pairs that are new to it (KEDDY-HECTOR et al. 1999). Does it have a concept of SAMENESS? Many comparative psychologists consider it reasonable to attribute concepts to nonhuman animals on the basis of such results. But questions remain about the adequacy of such approaches to support conclusions about animal concepts.

Many of the skeptical arguments that have been put forward focus on the relationship of language to human concepts. For example, CHATER/HEYES (1994) argue that because nonhuman animals lack language it is not possible to assign content to animal concepts in a rigorous way. Reviewing theories of concept drawn from philosophy, cognitive psychology, comparative psychology, and cognitive ethology, CHATER and HEYES argue that none of the proposed theories can simultaneously satisfy the three desiderata of (1) accounting for the close relationship between language and concepts in humans, (2) being applicable to non-linguistic animals, and (3) being empirically tractable by behavioral methods. They conclude (p237) that "no clear sense has been provided for the claim that nonlinguistic animals have concepts" and that "we simply do not know how to turn the claim that nonlinguistic animals have concepts into an empirically substantive question".

In the face of such criticism, should comparative psychologists and cognitive ethologists abandon the notion of animal concepts? There would be costs to doing so. First, insofar as cognitive psychologists have considered concepts central to their theories of human psychology, the abandonment of animal concepts would make it harder to relate studies of animal cognition to attempts to understand the evolution of human concepts. Second, there is a very tight link between notions of concept and intentional content. Concepts are constituents of the so-called intentional states (beliefs, desires, etc.): for instance, the thought that there is a squirrel in the tree is constituted by a structured arrangement of the concepts of squirrel, tree, and in-ness. Giving up on the notion of animal concepts would make it difficult (although perhaps not impossible) to see the form of a theory of content for the intentional states of animals. Conversely, a workable theory of animal concepts would be a significant contribution to the theory of animal intentionality generally.

Previously, Marc HAUSER and I also disputed the adequacy of standard comparative psychological results to support claims about concepts (ALLEN/HAUSER 1991), but did not embrace the skeptical conclusion suggested by CHATER and HEYES. Rather,

we undertook to provide suggestions for the empirical investigation of concepts in nonhuman animals. Specifically, we maintained that the attribution of concepts is justified "if there is evidence supporting the presence of a mental representation that is independent of solely perceptual information" and by means of a pair of thought experiments we indicated how we thought such evidence could be collected. Both of these thought experiments involved interpreting animals' reactions to perceptual evidence that would seem to indicate conflicting conclusions. Both of our thought experiments revolved around providing information to an individual organism that a familiar conspecific was not dead despite the appearance of death, and looking for changes in future reactions to subsequent presentation of the markers for death (or the absence of markers for life).

The theoretical basis underlying the inference from these experiments to claims about concepts was not carefully articulated in that 1991 paper. Instead we implicitly relied upon a comparative approach, which justified the attribution of concepts to nonhuman animals by using human behavior in similar circumstances as the benchmark. It was not, however, our intention to provide an account of animal concepts that was essentially anthropocentric for it was our belief that the basic strategy of using responses to conflicting information as a basis for concept attribution could stand alone. The present paper attempts to give a more explicit statement of the framework underlying this suggestion about empirical research into animal concepts.

Before turning to the substance of the proposal, it will be useful to make a distinction that is fundamental to any discussion of concepts, but that is often neglected nonetheless. The distinction is marked in the questions "What is *the* (or *our*) concept of tree?" versus "What is *Fred's* concept of tree?" which deal respectively with social and individual notions of concept. Whenever the idiom "*the* concept" is used, the definite article suggests that there is one concept which many individuals share (likewise for "our concept"). But it is also common to hear it said that everyone has a different concept of *X*. From the perspective of this claim about individual concepts, it can seem unlikely that there should be such a thing as "the" concept of anything rather than just a collection of more or less related individual concepts of *X*. In commonsense psychology this tension between social and individual conceptions of concept is mostly unanalyzed.

Both social and individual conceptions of concept are important to scientific psychology. Social concepts play a role in explaining communication and cooperation among individuals, while individual concepts are implicated in the structure of individual behavior and differences between the behavior of individuals. Both conceptions also have intuitive counterparts in animal psychology. For instance, assuming that it will be possible to make sense of animal concepts, one might consider what concepts are implicated in the alarm calling systems of a particular species, such as the vervet. As in human social groups, the group concept involves a centralizing tendency: vervets communicate and cooperate with each other in ways that tend to maintain a stable reference for specific alarm calls. Yet there is also room for individual difference: some vervets may be more likely than others to distinguish between different species of raptor for the production of their avian predator alarm calls, and this may not be simply a function of differences in perception.

In scientific contexts (particularly cognitive psychology) where the focus is ostensibly on individual concepts it is quite common to find the use of idioms that are more suited to the social conception. Philosophical arguments about animal cognition are also plagued by failure to heed the distinction. For example, several philosophers have been tempted by the argument that animals do not have beliefs because they lack the constituent concepts: for example, a dog does not believe there is a squirrel in the tree because it lacks "the" (presumably the authors mean *human*) concept of squirrel. But there is no reason to think that having the belief requires that animals have that specific concept, nor that lacking the canonical concept of squirrel means that they lack any concept whatsoever (see ALLEN 1992 and ALLEN/BEKOFF 1997 for discussion).

The notion under discussion in this paper is that of the individual conception of concept. Of particular interest is the relationship between concepts and perception. Most, and perhaps all organisms are capable of categorizing stimuli into perceptual equivalence classes—a basic requirement for producing consistent responses to stimuli. But some organisms construct categorization schemes that, in a sense to be explained below, transcend particular perceptual stimuli. Concepts are the mental representations constituting the nodes in such categorization schemes.

For example, consider the representation of death (ALLEN/HAUSER 1991). Many organisms have

perceptual mechanisms whose function it is to allow them to respond differentially to dead or alive conspecifics. Ants, for example, respond to the presence of acidic byproducts of decomposition, allowing them to detect and remove dead conspecifics from the nest. Yet ants almost certainly have no concept of death; that is they provide no evidence of a capacity to represent death independently of their perception of the chemical indicator. Thus, in this respect, they do not transcend perceptual stimuli. This is indicated by the fact that ants will remove from the nest anything that has oleic acid painted on it, including other live ants. Although the ants get other information that would tend to count against the assessment that the acid-treated conspecific is dead (they are for instance capable of detecting its motion), they are incapable of using this information to modify the removal response. While it is the biological function of their chemosensors to detect dead conspecifics, this detection ability is merely perceptual not conceptual.

In contrast, humans can represent the status of an organism independently of particular perceptual stimuli. A body that appears not to be breathing may nonetheless be judged to be alive, and one that appears to be breathing may be judged dead. (There is an emphasis here on the word "appears".) Conceptual representation of death allows the organism to transcend any particular stimuli. This in turn facilitates more flexible behavior. The capacity to represent death independently of specific perceptual indicators allows the organism to learn new relationships between perceptual indicators and the underlying condition of being dead.

We are now in a position to see how one might overcome the objection that no empirical content can be given to the claim that a languageless creature possesses concepts: one seeks evidence for the non-perceptual representations just mentioned. There may be several ways to do this, but here is one suggested schema for the investigation of concepts:

An organism O may reasonably be attributed a concept of X (e.g., TREE) whenever:

(i) O systematically discriminates some Xs from some non-Xs; and

(ii) O is capable of detecting some of its own discrimination errors between Xs and non-Xs; and

(iii) O is capable of learning to better discriminate Xs from non-Xs as a consequence of its capacity (ii).

It is important to be clear that the purpose of these three clauses is not to provide a philosophical analysis of what it is for an organism to possess a concept. The question of when it is reasonable to attribute a concept to an animal is a distinct question from that of what it means for an animal to possess a concept, just as the question of when it is reasonable to believe that someone is a murderer differs from the question of what it means to be a murderer. Meeting conditions (i)–(iii) above may provide good grounds for attributing concepts to animals, even though these conditions need be neither necessary nor sufficient for concept possession, just as finding a victim's blood on a pair of socks may provide good grounds for believing the sock owner to be a murderer even though blood on sock is neither a necessary nor sufficient condition for being a murderer (ALLEN/BEKOFF 1997). To explain why satisfaction of the three conditions should be considered good evidence for concept possession in animals it is necessary to say something about what it means to possess a concept, but much less than a full analysis needs to be provided.

Investigation of the capacities specified in clauses (i)–(iii) is empirically tractable in languageless creatures. Satisfaction of clause (i) has been extensively studied in animals by the kinds of experiments referred to in the first paragraph of this paper. For instance, pigeons shown photographs of human faces interspersed with photographs of other items can be taught to peck a response key only when a face is presented. The animals generalize their categorization responses to new photographs containing faces that were not seen during training. This shows that the discriminations are systematic, not based on having learned piecemeal the particular response for each individual member of the set of training stimuli. Likewise, KEDDY-HECTOR et al. (1999) show that pigs learned to respond differentially to stimuli consisting of a pair objects, depending on whether the two objects differed in one of the properties of shape, size, or color, or whether they were identical in those respects, and that the pigs generalized the discrimination of sameness and difference to novel pairs of familiar objects and to pairs of novel objects. Although the same/different categorization is of an abstract relationship between perceptual properties, the discrimination still remains tied to perception in the sense that we have no evidence that these animals can judge that two things are really the same despite appearing different, or vice versa. Thus these discoveries satisfy clause (i) of the schema but not either of the other clauses.

The capacities specified in clauses (ii) and (iii) have not been extensively investigated, but in combination they provide a basis for empirical investi-

gation that allows us to resist the pessimistic conclusion drawn by CHATER and HEYES. By helping to determine the boundaries of the animals' representations, the joint investigation of capacities (ii) and (iii) can help to settle questions about the content of the representations. Clause (ii) concerns detection of error. There are various ways that an animal might be informed by an external cue that it has miscategorized something, but in such cases there is a challenge to any empirical inference because of the possibility that the cue indicating the error is sufficient alone to explain any subsequent behavior by an animal. (It is rather like the chemical cue being sufficient alone to explain why ants remove dead conspecifics—no cognitive representation of death is required.)

It is, however, possible for animals to show evidence of error detection without external cuing, i.e., they can show self-monitoring of their performance. Pigs again illustrate the point, although in this example it is knowledge of response error rather than perceptual error that is demonstrated. Pigs trained on the same/different task just described would still occasionally make errors (wrong choices) while performing at an overall rate at or near 90% correct. After committing to a response but before any feedback was provided, some pigs would attempt to back away from the choice they had made. Analysis of 22 cases of the backout behavior showed that in only one of these cases occurred after the pig made a correct choice (KEDDY-HECTOR et al 1999). It is intriguing to try to account for this behavior, but regardless of what it shows in this particular experimental set up, it indicates that under certain conditions some animals can provide non-linguistic evidence that supports the attribution of endogenous error detection capacity.

Clause (iii) is the most difficult to articulate and defend. Indeed, some participants at workshops based on the ideas in this paper were satisfied that concepts are reasonably attributed on the basis of the other clauses. Some were satisfied with (i) alone (and some prefer other criteria—those are fine for the point here is not to provide either necessary or jointly sufficient conditions for concept possession). It is my view, however, that clause (iii), when satisfied, provides a stronger case for the attribution of concepts because it provides a link between the first two capacities. If the capacity to satisfy (ii) is causally responsible for better performance on (i) then we have evidence that there is an integrated processing mechanism linking perceptual categorization and the recognition of perceptual error, and hence a representation system that compares the perceptual content with an independent representation of what the perception is supposed to represent, i.e., a concept.

The internal states implicated in the explanation of these capacities are worthy of being designated as concepts. For these capacities to be implemented it appears that there must be an internal standard of comparison that represents the organism's world independently of its perceptual representation at any given moment. Thus, such evidence supports the claim that organisms with these capacities possess representations of the world that are detached from immediate perceptual information.

The difference between perceptual representations and concepts correspond to longstanding philosophical ideas about concepts (WATSON 1995). Descartes, for example, discusses the power of the mind to represent a chiliagon even though the visual presentation of such an object would be indistinguishable from that of a 999-sided figure. Among cognitive psychologists too, the major current theories of concepts are all concerned with the way in which concepts serve to unite varied perceptual presentations of instances.

The close connection of language to concepts in humans has seduced many into thinking that the two notions of language and concept cannot be disentangled. This close connection may be explained on the current schema in virtue of the fact that languages provide a structure that has a vast number of degrees of freedom with respect to immediate perception. Linguistic representation is, then, the basis for the most fine-grained system of conceptual representation that we know. But it would be a mistake to think that it is the only basis for conceptual representation available. It is well within the bounds of possibility that other species have ways of structuring their experiences that transcend merely grouping those experiences into equivalence classes for the purposes of producing immediate behavioral responses. Such a capacity contains the rudiments of a conceptual scheme. Thus, I argue, there is a clear sense for the claim in which languageless animals might possess concepts, and we do know how to turn this claim into an empirically substantive question.

Author's address

Colin Allen, Department of Philosophy, Texas A&M University, College Station, TX 77843-4237, USA.
Email: colin.allen@tamu.edu

Note

1 Reprinted with kind permission of Kluwer Academic Publishers. The paper originally appeared 1999 in Erkenntnis 51: 33–40.

References

Allen, C. (1992) Mental content. British Journal for the Philosophy of Science 43: 537–553.

Allen, C./Bekoff, M. (1997) Species of Mind: the philosophy and biology of cognitive ethology. MIT Press: Cambridge MA.

Allen, C./Hauser, M. D. (1991) Concept attribution in non-human animals: theoretical and methodological problems in ascribing complex mental processes. Philosophy of Science 58: 221–240.

Chater, N./Heyes, C. M. (1994) Animal concepts: content and discontent. Mind and Language 9: 209–246.

Herrnstein, R. J./Loveland, D. H./Cable, C. (1976) Natural concepts in pigeons', Journal of Experimental Psychology. Animal Behavior Processes 2: 385–302.

Keddy-Hector, A./Allen, C./Friend, T. H. (1999) Cognition in Domestic Pigs: Relational Concepts and Error Recognition. Submitted.

Watson, R. A. (1995) Representational Ideas: from Plato to Patricia Churchland. Kluwer: Dordrecht.

Marc Bekoff

Social Cognition: Exchanging and Sharing Information on the Run[1]

Abstract

In this essay I consider various aspects of the rapidly growing field of cognitive ethology, concentrating mainly on evolutionary and comparative discussion of the notion of intentionality. I am not concerned with consciousness, per se, for a concentration on consciousness deflects attention from other, and in many cases more interesting, problems in the study of animal cognition. I consider how, when, where, and (attempt to discuss) why individuals from different taxa exchange social information concerning their beliefs, desires, and goals. My main examples come from studies of social play in mammals and antipredator behavior in birds. Basically, I argue that although not all individuals always display behavior patterns that are best explained by appeals to intentionality, it is misleading to argue that such explanations have no place in the study of animal cognition.

Key words

Social cognition, cognitive ethology, play, antipredator behavior, vigilance, canids, birds.

Introduction

Individuals in many animal taxa need to be able to exchange social information. While many details of transmitting, receiving, and processing information can be explained by appealing to hard-wired evolved mechanisms or to rules–of–thumb acquired during past exposures to various stimulus–response contingencies, in other instances explanations that appeal to the philosophical notion of intentionality (beliefs and desires) more easily and parsimoniously explain observed patterns of behavior. While there has been a good deal of concentration on the cognitive abilities of nonhuman primates (BYRNE/WHITEN 1988; BYRNE 1995; RUSSON et al. 1996), there also is a wealth of data on nonprimates that is important for studies of nonhuman animal (hereafter animal) cognition (BEKOFF/JAMIESON 1990, 1996; RISTAU 1991a; GRIFFIN 1992; ALLEN/BEKOFF 1997; BEKOFF/BYERS 1998).

In this essay I discuss two different types of behavior, mammalian social play and antipredator behavior in birds, to make the case that the time has come to expand our taxonomic interests in cognitive ethological inquiries. These examples show how individuals exchange information during ongoing social encounters—how analyses of social cognition on the run can inform comparative inquiries into what might be going on in animals' minds.

Social play behavior: How individuals negotiate cooperative agreements on the run

In recent years, the scientific study of play behavior has undergone many significant changes. Researchers in various disciplines are conducting detailed observational, experimental, and theoretical analyses of play primarily in mammals (including humans) and birds. They are concerned with such topics as evolution, ecology, development, social communication, individual well-being, neurobiology, learning, and cognition (for references see BEKOFF/BYERS 1981, 1998; FAGEN 1981; BURGHARDT 1998; BEKOFF/ALLEN 1992, 1998; ALLEN/BEKOFF 1994, 1997; PELLIGRINI 1995; SUTTON-SMITH 1998; BURGHARDT 1999). New data are forcing people to give up old ideas and set ways of thinking about this phenomenon. Most researchers recognize the importance of rigorous interdisciplinary collaboration in play research (PELLIGRINI, 1995; BEKOFF/BYERS, 1998; BURGHARDT, 1998), and the interactions among those studying humans and non-humans are producing exciting new results.

Social play behavior is performed by individuals in many mammalian, avian, and perhaps other vertebrate taxa (FAGEN 1981; BEKOFF/BYERS 1998; BURGHARDT 1998). It is a behavioral phenotype that lends itself to detailed empirical study (ALLEN/BEKOFF 1997; BEKOFF/BYERS 1998; BURGHARDT 1999). Follow-

ing BEKOFF/BYERS (1981), I will define play as follows: "*Play* is all motor activity performed postnatally that *appears* (my emphasis) to be purposeless, in which motor patterns from other contexts may often be used in modified forms and altered temporal sequencing. If the activity is directed toward another living being it is called *social play*." This definition centers on the structure of play sequences—what animals do when they play—and not on possible functions of play. It concerns what play is, rather than what it is not. The flexibility and versatility of social play make it a good candidate for comparative and evolutionary cognitive studies including those that are concerned with ways in which individuals may negotiate agreements to engage in cooperative (not merely coordinated) social interactions. FAGEN (1993, p192) has noted that "Levels of cooperation in play of juvenile primates may exceed those predicted by simple evolutionary arguments…" Play cannot occur if the interacting animals do not agree to cooperate.

Cognitive ethological approaches are useful for gaining an understanding of social play for various of reasons including: (1) it exemplifies many of the theoretical issues faced by cognitive ethologists; (2) empirical research on social play has and will benefit from a cognitive approach, because play involves communication, intention, role-playing, and cooperation; (3) detailed analyses of social play may provide more promising evidence of animal minds than research in many other areas, for it may yield clues about the ability of animals to understand one another's intentions; and (4) play occurs in a wide range of mammalian species and in a number of avian species, and thus it affords the opportunity for a comparative investigation of cognitive abilities extending beyond the narrow focus on primates that often dominates discussions of non-human cognition. For example, during social play, many nonprimates (and primates) engage in self-handicapping and role-reversals, two behavior patterns that are often used to make inferences about intentionality (and consciousness and self-consciousness).

Along these lines, POVINELLI/CANT (1995, p400) suggest that the performance by arboreal ancestors of the great ape/human clade of "unusual locomotor solutions … drove the evolution of self-conception". Many nonprimate mammals also perform complex, flexible, and unusual acrobatic motor patterns (locomotor-rotational movements; WILSON/KLEIMAN 1974) during social play, and it would be premature to rule out the possibility that the performance of these behaviors is also important to the evolution of self-conception in nonprimates. In some instances it is difficult to know whether arboreal clambering or the performance of various acrobatic movements during play may more be related to the evolution of (mere) body awareness (e.g., knowing one's place in space) and not a concept of self.

The study of social play also lends itself to Colin ALLEN's work on concepts (this volume). ALLEN explores the idea that an organism O may reasonably be attributed a concept of X (e.g., play) whenever: (1) O systematically discriminates some instances of X from some non-Xs; (2) O is capable of detecting some of its own discrimination errors between Xs and non-Xs; and (3) O is capable of learning to better discriminate Xs from non-Xs as a consequence of its capacity (2).

With respect to ALLEN's point (1), individual canids, especially coyotes, seem to be able to discriminate some cases of play from non-play. They clearly respond differently to long and vigorous bites and long stand-overs (an action that is performed significantly more by dominant individuals during aggressive interactions). Bites directed toward the body and stand-overs during play are of significantly shorter duration during play than during aggressive encounters (HILL/BEKOFF 1977). Concerning ALLEN's point (2), individuals seem to have the capacity to recognize that they have misclassified some encounters as play or non-play. For example, when it seems likely that a playful encounter will escalate into aggression, individuals tend to perform a play bow (see below). ALLEN (personal communication) pointed out that this might not be the case if play escalating into aggression is no longer perceived to be play by the interacting animals; there might be no misclassifcation in this case. With respect to point (3), individuals also appear to learn to discriminate play from non-play as a consequence of recognizing their own previous misclassifications. This is especially so in cases where they use play bows or other actions to indicate "this is play" with individuals with whom they had previously made errors. That is, individuals seem more likely to assess more carefully the situation if they have made previous category (play/non-play) errors (BEKOFF 1975).

The communication of the intention to play: The meaning of bows

When individuals play they typically use action patterns that are also used in other contexts, such as predatory behavior, antipredatory behavior, and mating. These action patterns may not be intrinsically different across different contexts, or they

may be hard to discriminate even for the participants.

How do animals know that they are playing? How do they communicate their desires or intentions to play? Individuals might know that they are playing because the actions that are performed differ when they are performed during play when compared to other contexts (HILL/BEKOFF 1977), or the order in which motor patterns are performed differs from, and might be more variable than, the order in which they are performed during the performance of, for example, serious aggression or serious predation (BEKOFF/BYERS 1981). While there still are very few empirical data that speak to either of these alternatives (BEKOFF/BYERS 1981, 1998), numerous people who study play come away with the impression that these possibilities would have strong and wide support if they were studied rigorously. Suffice it to say, playing animals seem to fine-tune on-going play sequences to maintain play and to prevent play from escalating in real aggression.

Because there is a chance that various behavior patterns that are performed during on-going social play can be misinterpreted, individuals need to tell others "I want to play", "this is still play no matter what I am going to do to you", or "this is still play regardless of what I just did to you". An agreement to play, rather than to fight, mate, or engage in predatory activities, can be negotiated in various ways. Individuals may use various behavior patterns—play markers—to initiate play or to maintain the play mood (BEKOFF 1977, 1995a; CHALMERS 1980; NEWBERRY/WOOD-GUSH/HALL 1988; ALLEN/BEKOFF 1997) by punctuating play sequences with these actions when it is likely that a particular behavior may have been, or will be, misinterpreted. There is little evidence that play signals are used to deceive others.

The "bow", a highly ritualized and stereotyped movement that seems to function to stimulate recipients to engage (or to continue to engage) in social play, provides an excellent example of a signal and it has been extensively studied in various canids in this context. Bows (the animal crouches on her forelimbs and elevates her hindlimbs) occur throughout play sequences, but most commonly at the beginning or towards the middle of playful encounters. In a detailed analysis of the form and duration of play bows (BEKOFF 1977) it was shown that duration was more variable than form, and that play bows were always less variable when performed at the beginning, rather than in the middle of, ongoing play sequences. Three possible explanations for this change in variability include: (1) fatigue, (2) the fact that animals are performing them from a wide variety of preceding postures, and (3) there is less of a need to communicate that "this is still play" than there is when trying to initiate a new interaction. These explanations are not exclusive alternatives.

Play bows occur almost exclusively in the context of play, and it is common to attribute to play-soliciting signals the message "what follows is play" or "this is still play". What is the significance of these messages for the players themselves? Are they in any way aware of the meaning of the play bows, or are they simply conditioned to respond differently, for example, less aggressively or less sexually, when a specific action such as a bite or a mount is preceded by a play bow?

One way to approach these questions is to ask whether play signals such as bows are used to maintain social play in situations where the performance of a specific behavior during a play bout could be misinterpreted or when an individual makes an error. A recent study of the structure of play sequences (BEKOFF 1995a) showed that bows in some canids (infant and adult domestic dogs, infant coyotes, and infant wolves) often are used immediately before and immediately after an action that can be misinterpreted and disrupt ongoing social play. For example, in infant coyotes, wolves, and domestic dogs, the "bow" was not performed randomly, but rather immediately preceded or immediately followed behavior patterns that can be misinterpreted (for example, biting accompanied by vigorous side–to–side shaking of the head). Recall that the social play of canids (and of other mammals) contains actions, primarily bites, that are used in other contexts that do not contain bows (e.g., agonistic, predatory, or antipredatory). Actions such as biting accompanied by rapid side–to–side shaking of the head are used in aggressive interactions and also during predation and could be misinterpreted when used in play. While there are few comparative data, in most species in which play has been described, play-soliciting signals appear to foster some degree of cooperation between players so that each responds to the other in a way consistent with play and different from the responses the same actions would elicit in other contexts.

Sharing intentions

How might a play bow serve to provide information to its recipient about the sender's intentional state—his intentions, desires, and beliefs? Perhaps one's own experiences with play can promote learning about the intentions of others. Perhaps the recipient

shares the intentions (beliefs, desires) of the sender based on the recipient's own prior experiences of situations in which she performed bows. It may be reasonable to attribute a very specific second-order inference of the form "when I bow I want to play so when you bow you also want to play" without being committed to a general capacity for the possession of second-order mental states in these animals.

GOPNIK (1993) has suggested that others' body movements are mapped onto one's own kinesthetic sensations based on prior experience of the observer. She has claimed (p275): "In particular, we innately map the body movements of others onto our own kinesthetic sensations. This initial bridge between the inside and the outside, the self and other, underlies our later conviction that all mental states are things both we and others share." TOMASELLO/GUST/FROST (1989) also have noted that some gestures in chimpanzees may be learned by "second-person imitation"—"an individual copying a behavior directed to it by another individual" (p35). They concluded (p45) that chimpanzees "… rely on the sophisticated powers of social cognition they employ in determining what is perceived by a conspecific and how that conspecific is likely to react to various types of information…" FLANAGAN (1992, pp102ff) also has been interested in ways in which mental states can be shared, and introduced the notion of a "mental detector" that is used to detect others' invisible mental states.

There is research that suggests a neurobiological basis for the understanding of others' mental states. "Mirror neurons", found in macaques, fire when a monkey executes an action and also when the monkey observes the same action performed by another monkey (GALLESE 1998; GALLESE/GOLDMAN 1998; see also JEANNEROD 1994 for discussion of possible neural bases for kinesthetic–visual matching). These findings are intriguing, and more comparative data are needed to determine if mirror neurons are found in other taxa and if they might actually play a role in the sharing of intentions between individuals engaged in an on-going social interaction.

Self-handicapping and role-reversing in play

Two behavior patterns that have been reported to occur during on-going social play in various mammals are self-handicapping and role-reversing (for reviews see BEKOFF/ALLEN 1998; BIBEN 1998; PELLIS/PELLIS 1998; THOMPSON 1998; WATSON 1998). Both might also be important for maintaining a play mood.

Self-handicapping occurs when an individual performs a behavior patterns that might compromise herself. For example, a coyote might not bite her play partner as hard as she can, or she might not play as vigorously as she can. WATSON/CROFT (1996) found that red-neck wallabies adjusted their play to the age of their partner. When a partner was younger, the older animal adopted a defensive, flat-footed posture, and pawing rather than sparring occurred. In addition, the older player was more tolerant of its partners tactics and took the initiative in prolonging interactions. (While more data are needed, this study also suggests that the benefits of play may vary according to the age of the player.)

Role-reversing occurs when a dominant animal performs a action during play that would not normally occur during real aggression. For example, a dominant animal might voluntarily not roll-over on his back during fighting, but would do so while playing.

In some instances self-handicapping and role-reversing might occur together. For example, a dominant individual might roll over while playing with a subordinate animal or inhibit the intensity of a bite. In other situations, self-handicapping and role-reversing might not be performed together—a subordinate animal restrains herself during play so as not to allow the interaction to escalate into a real fight. There are very few data that are concerned with self-handicapping and role-reversing in play. In canids, infant coyotes, when compared to infant wolves and beagles, are more aggressive and engage in serious fights that determine social rank (BEKOFF 1974). Dominant coyote pups use rolling-over as self-handicapping and role-reversing to get subordinate animals to play. In once case, 38/38 instances of rolling-over by a dominant coyote were used in the context of social play. Furthermore, 34/42 (81%) approach/withdrawals were used in social play and resulted in chasing by the subordinate coyote. In real aggressive encounters, dominant animals do not allow others to chase them.

Both self-handicapping and role-reversing are also responsible for altering the temporal sequencing of play and might serve to call further attention to the intentions of the animal who engages in these activities. That is, not only are the behavioral sequences performed in play different from those performed during serious aggression, but play sequences are punctuated with behavior patterns that indicate that an individual is willing either to handicap himself or to engage in role-reversing. From a functional perspective, self-handicapping and role-

reversing, similar to using specific play invitation signals or altering behavioral sequences, might serve to signal an individual's intention to continue to play.

The performance of self-handicapping and role-reversing suggests that an individual is able to make an assessment of himself and others and manage or regulate how she plays (PELLIS/PELLIS 1998, THOMPSON 1998), and detailed analyses of these activities will undoubtedly inform arguments about the cognitive capacities of animal who engage in them (ALLEN/BEKOFF 1997). Certainly, it appears that individuals trust others to maintain the rules of the game. BIBEN (1998) stresses the importance of developmental flexibility in the social behavior of rhesus monkeys, and suggests that individuals use play as a way to learn strategies of social interaction that are related to self-competence and self-preservation.

To sum up briefly, I found that at least some canids (and, I expect other mammals also would) cooperate when they engage in social play, and that individuals may negotiate these cooperative ventures by sharing their intentions. In general, animals engaged in social play use specific signals to modulate the effects of behavior patterns that are typically performed in other contexts, but whose meaning is changed in the context of play. These signals are often flexibly related to the occurrence of events in a play sequence that might violate expectations within that sequence. Furthermore, the relationship of play to a cognitive appreciation of the distinction between reality and pretense provides an important link to other cognitive abilities, such as the ability to detect deception (which is extremely rare in social play) or to detect errors. Given these connections, further detailed analyses of various aspects of social play might help promote the development of more sophisticated theories of intentionality, representation, communication, and (perhaps even) consciousness.

Antipredatory behavior in birds: Cognition on the fly

Antipredator behavior in birds also provides a good example of the usefulness of comparative cognitive ethological approaches. For example, RISTAU (1991b) conducted field studies of injury-feigning in piping plovers (the broken-wing display) and wanted to know if she could learn more about deceptive injury-feigning if she viewed the broken-wing display as an intentional or purposeful behavior ("the plover wants to lead the intruder away from her nest or young") rather than as a hard-wired reflexive response to the presence of a particular stimulus, a potentially intruding predator. She studied the direction in which birds moved during the broken-wing display, how they monitored the location of the predator, and the flexibility of the response. RISTAU found that birds usually performed the display in the direction that would lead an intruder who was following them further away from the threatened nest or young, and also that birds monitored the intruder's approach and modified their behavior in responses to variations in the intruder's movements. These and other data led RISTAU to conclude that the plover's broken-wing display lent itself to an intentional explanation—that plovers purposely lead intruders away from their nests or young and modified their behavior in order to do so.

In another field study of antipredator behavior in birds, I (BEKOFF 1995b, 1996) found that western evening grosbeaks modified their vigilance or scanning behavior depending on the way in which individuals were positioned with respect to one another. Grosbeaks and other birds often trade-off scanning for potential predators and feeding. Basically (but oversimplified), some birds scan while others feed and some birds feed when others scan. Thus, it can be hypothesized that individuals want to know what others are doing and learn about others' behavior by trying to watch them.

My study of grosbeaks showed that when a flock contained four or more birds, there were large changes in scanning and other patterns of behavior that seemed to be related to ways in which grosbeaks attempted to gather information about other flock members. When birds were arranged in a circular array so that they could see one another easily compared to when they were arranged in a line that made visual monitoring of flock members more difficult, birds who had difficulty seeing one another were (1) more vigilant, (2) changed their head and body positions more often, (3) reacted to changes in group size more slowly, (4) showed less coordination in head movements, and (5) showed more variability in all measures.

Behavioral flexibility and intentional explanations

The differences in behavior between birds organized in circular arrays when compared to birds organized in linear arrays were best explained by accounting for individuals' attempts to learn, via visual moni-

toring, about what other flock members were doing. This may say something about if and how birds attempt to represent their flock, or at least certain other individuals, to themselves. It may be that individuals form beliefs about what others are most likely doing and predicate their own behavior on these beliefs. I have argued that cognitive explanations were simpler and less cumbersome than non-cognitive rule–of–thumb explanations (e.g., "scan *this* way if there are *this* number of birds in *this* geometric array" or "scan *that* way if there are *that* number of birds in *that* geometric array." Non-cognitive rule–of–thumb explanations did not seem to account for the *flexibility* in animals' behavior as well or as simply as explanations that appealed to the intentions of the animals under study.

In many other instances it also may be more economical or parsimonious to assume that not everything that an individual needs to be able to do in all situations in which it finds itself is preprogrammed; intentional explanations can be simpler than cumbersome stimulus–response explanations (DE WAAL 1991; ALLEN/BEKOFF 1997; BEKOFF/ALLEN 1997; BEKOFF 1998a). While general rules of thumb may be laid down genetically during evolution, specific rules of conduct that account for all possible contingencies may be too numerous to be hard-wired. Furthermore, while behavioristic learning schemes appealing to notions such as conditioning, generalizing, and substituting can account to a limited extent for behavioral flexibility, behavioral integration, and the use of internal states and images of absent objects in some organisms (e.g., HOLLAND 1990), learning at high degrees of abstraction from sensory stimulation seems less amenable to behavioristic analysis (BEKOFF/ALLEN 1992).

Cognitive models of learning provide explanatory schemes for such cases. It might actually be more parsimonious to appeal to intentional explanations in terms of accounting for complex patterns of behavior with fewer explanations. For example, the results of my study on grosbeaks, namely that flock geometry influences various patterns of behavior, are more simply explained in terms of the representational needs or information-gathering goals of grosbeaks than by an attempt to account for the effects of group geometry in terms of numerous and diverse stimulus–response contingencies or generalizations from earlier experiences (e.g., the birds are somehow conditioned [or innately predisposed] to produce certain behavioral patterns in response to group geometry). In this case it is difficult to conceive that an individual grosbeak's experience could have included explicit conditioning about the factors that influence when and how it should scan or move its head and body.

Towards a broad and comparative cognitive ethology: The relationship between what's happening "out there" and "in here"

In his early work, Niko TINBERGEN (1951/1989; 1963) identified four overlapping areas with which ethological investigations should be concerned, namely, evolution (phylogeny), adaptation (function), causation, and development (ontogeny), and his framework also is useful for those interested in animal cognition (JAMIESON/BEKOFF 1993; ALLEN/BEKOFF 1997). BURGHARDT (1997) suggested adding a fifth area, *private experience*. He (p276) noted that "The fifth aim is nothing less than a deliberate attempt to understand the private experience, including the perceptual world and mental states, of other organisms. The term private experience is advanced as a preferred label that is most inclusive of the full range of phenomena that have been identified without prejudging any particular theoretical or methodological approach."

BURGHARDT also noted that calling the fifth area "cognitive ethology" is not radical enough because there may be some historical baggage associated with the use of this term. For one, it is frequently associated with only the study of animal consciousness. However, cognitive ethology is much more than the study of animal consciousness (ALLEN/BEKOFF 1997; BEKOFF/ALLEN 1997; BEKOFF 1998a) although many do not mark this distinction.

My studies of animal cognition have taught me a number of important lessons. The first is that detailed observational and descriptive research are mandatory, especially under field conditions. We must know what animals do "out there" in order to learn more about what is going on "in here"—in their minds. Of course, experimental work is essential as well, but before we understand the lives of the animals we study much experimental work is premature and can even be misleading. A second lesson is that the sort of research that is needed in order to study different aspects of animal cognition also is extremely tedious and time-consuming, but there are no substitutes for this work. For example, in a current project on vigilance in western evening grosbeaks, I am finding that individuals exchange information very rapidly. Information on head (and eye) and body position is lost if frames are not analyzed

every 1/30th of a second (BEKOFF, unpublished data). I hope that my colleagues who study animal cognition as well as (especially?) those who are extremely critical of the study of animal cognition and of the notion of animal minds will engage in, and in some instances return to, this type of research as we head into the 21st century. The third lesson concerns the use of old data. There are many existing data that can inform cognitive ethological inquiries and researchers might consider revisiting them while planning future studies. Sometimes a fresh look at old data can be extremely revealing.

All in all, we are a long way from having an adequate data base from which hard-and-fast claims about the taxonomic distribution of various cognitive skills, or about the having of a theory of mind, can be put forth. BYRNE (1995), who otherwise takes a strongly primatocentric view of animal cognition, agrees with the importance of conducting broadly comparative research. Consider studies that show that some monkeys cannot perform imitation tasks that some mice can (WHITEN/HAM 1992). If the point is to answer the question are monkeys smarter than mice or not it is misleading, for there is no reason to expect a single linear scale of intelligence (BEKOFF 1998b,c). In the world of mice it may be more important to be able to do some things than it is in the world of monkeys, but in other respects a monkey may have cognitive capacities that a mouse would lack. There also is much *individual* variation within species, and this also must be documented more fully.

To sum up, speciesist cognitivism must be resisted. We should strive to study individual animals to learn more about their lives in their own worlds (and their abilities to feel pain and to suffer psychologically and physically). Methodological pluralism is required: species-fair methods need to be tailored to the questions and the animals under consideration, and competing hypotheses and explanations need always to be considered. In the absence of suitable criteria and empirical data for making comparative claims about smartness or intelligence, we need to be very careful about making such statements as "apes are smarter than monkeys or dogs" for each can do things the other cannot. Thus, MARLER (1996, p22) concluded his review of social cognition in non-human primates and birds as follows: "I am driven to conclude, at least provisionally, that there are more similarities than differences between birds and primates. Each taxon has significant advantages that the other lacks." Along these lines, TOMASELLO/CALL (1997, pp399–400) summarized their comprehensive review of primate cognition by noting that "The experimental foundation for claims that apes are 'more intelligent' than monkeys is not a solid one, and there are few if any naturalistic observations that would substantiate such broad-based, species-general claims".

There are many examples other than social play and antipredator behavior that could have been chosen for illustration, but these two behavior patterns make the case that chauvinism on either side of the debate concerning how to explain animal behavior is unwarranted. Sometimes some non-humans (and some humans) behave as stimulus–response machines and at other times some non-humans (and some humans) behave in ways that are best explained using a rich cognitive vocabulary. A pluralistic approach will result in the best understanding of the minds of other animals.

Acknowledgments

I thank the organizers of this meeting for inviting me to Bremen in September 1997, and Colin ALLEN for all the discussions we have had and continue to have on animal minds and cognitive ethology. Marek SPINKA and I also have discussed numerous issues dealing with play and I thank him for these fruitful exchanges. I also thank Anne BEKOFF for allowing me to use her Peak Motus Motion Analyzer System in some of my work on avian vigilance.

Author's address

Marc Bekoff, Department of Environmental, Population, and Organismic Biology, University of Colorado, Boulder, Colorado 80309-0334 USA
Email: marc.bekoff@colorado.edu

Note

1 Reprinted with kind permission of Kluwer Academic Publishers. The paper originally appeared 1999 in Erkenntnis 51: 113–128.

References

Allen, C./Bekoff, M. (1994) Intentionality, social play, and definition. Biology and Philosophy 9: 63–74.

Allen, C./Bekoff, M. (1997) Species of Mind: The Philosophy and Biology of Cognitive Ethology. MIT Press: Cambridge MA.

Bekoff, M. (1974) Social play and play-soliciting by infant canids. American Zoologist 14: 323–340.

Bekoff, M. (1975) The communication of play intention: Are play signals functional? Semiotica 15: 231–239.

Bekoff, M. (1977) Social communication in canids: Evidence for the evolution of a stereotyped mammalian display. Science 197: 1097–1099.

Bekoff, M. (1995a) Play signals as punctuation: The structure of social play in canids. Behaviour 132: 419–429.

Bekoff, M. (1995b) Vigilance, flock size, and flock geometry: Information gathering by western evening grosbeaks (Aves, fringillidae). Ethology 99: 150–161.

Bekoff, M. (1996) Cognitive ethology, vigilance, information gathering, and representation: Who might know what and why? Behavioural Processes 35: 225–237.

Bekoff, M. (1998a) Cognitive ethology: In: Bechtel, W./Graham, G. (eds) Blackwell Companion to Cognitive Science. Blackwell Publishers: Oxford, pp. 371–379.

Bekoff, M. (1998b) Resisting speciesism and expanding the community of equals. BioScience 48: 638–641.

Bekoff, M. (1998c) Deep ethology, animal rights, and the Great Ape/Animal Project: Resisting speciesism and expanding the community of equals. Journal of Agricultural and Environmental Ethics 10: 269–296.

Bekoff, M./Allen, C. (1992) Intentional icons: towards an evolutionary cognitive ethology. Ethology 91: 1–16.

Bekoff, M./Allen C. (1997) Cognitive ethology: Slayers, skeptics, and proponents. In: Mitchell, R. W./Thompson, N./Miles, L. (eds) Anthropomorphism, Anecdote, and Animals: The Emperor's New Clothes? SUNY Press: Albany: New York. Pages. 313–334.

Bekoff, M./Allen, C. (1998) Intentional communication and social play: How and why animals negotiate and agree to play. In: Bekoff, M./Byers, J. A. (eds) Animal Play: Evolutionary, Comparative, and Ecological Perspectives. Cambridge University Press: Cambridge, New York, pp. 97–114.

Bekoff, M./Byers, J. A. (1981) A critical reanalysis of the ontogeny of mammaliansocial and locomotor play: An ethological hornet's nest. In: Immelmann, K./Barlow, G. W./Petrinovich, L./Main, M. (eds) Behavioral Development: The Bielefeld Interdisciplinary Project. Cambridge University. Press: New York, pp. 296–337.

Bekoff, M./Byers, J. A. (eds) (1998) Animal Play: Evolutionary, Comparative, and Ecological Approaches. Cambridge University Press: New York, Cambridge.

Bekoff, M./Jamieson, D. (eds) (1990) Interpretation and Explanation in the Study of Animal Behavior. Volume I: Interpretation, Intentionality, and Communication. 505 Pages; Volume II: Explanation, Evolution, and Adaptation. 465 Pages. Westview Press: Boulder CO.

Bekoff, M./Jamieson, D. (eds) (1996b) Readings in animal cognition. MIT Press: Cambridge MA.

Biben, M. (1998) Squirrel monkey playfighting: Making the case for a cognitive trainingfunction for play. In: Bekoff, M./Byers, J. A. (eds) Animal play: Evolutionary, Comparative, and Ecological Perspectives. Cambridge University Press: Cambridge, New York, pp. 161–182.

Burghardt, G. M. (1997) Amending Tinbergen: A fifth aim for ethology. In: Mitchell, R. W./Thompson, N./Miles, L. (eds) Anthropomorphism, Anecdote, and Animals. SUNY Press: Albany, New York, pp. 254–276.

Burghardt, G. M. (1998) Play. In: Greenberg, G./Haraway, M. (eds) Encyclopedia of Comparative Psychology. New York, Garland, pp. 725–735.

Burghardt, G. M. (1999) The Genesis of Animal Play. Routledge: New York.

Byrne, R. (1995) The Thinking Ape: Evolutionary Origins of Intelligence. Oxford University Press: New York.

Byrne, R./Whiten, A. (eds) (1988) Machiavellian intelligence: social expertise and the evolution of intellect in monkeys, apes, and humans. Oxford University Press: New York.

Chalmers, N. R. (1980) The ontogeny of play in feral olive baboons. Animal Behaviour 28: 570–585.

Fagen, R. M. (1981) Animal Play Behavior. Oxford University Press: New York.

Fagen, R. (1993) Primate juveniles and primate play. In: Pereira, M. E./Fairbanks, L. A. (eds) Juvenile Primates: Life History, Development, and Behavior. Oxford University Press: New York, pp. 183–196.

Flanagan, O. (1992) Consciousness Reconsidered. MIT Press: Cambridge MA.

Gallese, V. (1998) Mirror neurons, from grasping to language. Consciousness Bulletin Fall: 3–4.

Gallese, V./Goldman, A. (1998) Mirror Neurons and the Simulation Theory of Mind-Reading. Trends in Cognitive Science 2: 493–501.

Gopnik, A. (1993) Psychopsychology. Consciousness and Cognition 2: 264–280.

Griffin, D. R. (1992) Animal Minds. University of Chicago Press: Chicago.

Hill, H. L./Bekoff, M. (1977) The variability of some motor components of social play and agonistic behaviour in Eastern coyotes, Canis latrans var. Animal Behaviour 25: 907–909.

Holland, P. C. (1990) Event representation in Pavlovian conditioning: Image and action. Cognition 37: 105–131.

Jamieson, D./Bekoff, M. (1993) On aims and methods of cognitive ethology. Philosophy of Science Association 2: 110–124.

Jeannerod, M. (1994) The representing brain, neural correlates of motor intention andimagery. Behavioral and Brain Sciences 17: 187–245.

Marler, P. (1996) Social cognition: Are primates smarter than birds? In: Nolan, V. Jr./Ketterson, E. D. (eds) Current Ornithology Volume 13. Plenum Press: New York, pp. 1–32.

Newberry, R. C./Wood-Gush, D. G. M./Hall, J. W. (1988) Playful behaviour ofpiglets. Behavioural Processes 17: 205–216.

Pelligrini, A. D. (ed) (1995) The Future of Play Theory: A Multidisciplinary Inquiry into the Contributions of Brian Sutton-Smith. SUNY Press: Albany: New York.

Pellis, S. M./Pellis, V. C. (1998) The structure–function interface in the analysisof play fighting. In: Bekoff, M./Byers, J. A. (eds) Animal Play: Evolutionary, Comparative, and Ecological Perspectives. Cambridge University Press: Cambridge, New York, pp. 115–140.

Povinelli, D. J./Cant, J. G. H. (1995) Arboreal clambering and the evolution of self-conception. Quarterly Review of Biology 70: 393–421.

Ristau, C. (ed) (1991a) Cognitive Ethology: The Minds of Other Animals. Essays in Honor of Donald Griffin, R.. Lawrence Erlbaum: Hillsdale NJ.

Ristau, C. (1991b) Aspects of the cognitive ethology of an injury-feigning bird, the piping plovers. In: Ristau, C. A. (ed)

Cognitive Ethology: The Minds of Other Animals. Essays in Honor of Donald Griffin, R. Lawrence Erlbaum: Hillsdale NJ, Pages 91–126.

Russon, A. R./Bard, K. A./Parker, S. T. (1996) Reaching into Thought: The Minds of the Great Apes. Cambridge University Press: New York.

Sutton-Smith, B. (1998) The Ambiguity of Play. Harvard University Press: Cambridge MA.

Tinbergen, N. (1951/1989) The Study of Instinct. Oxford University Press: New York.

Tinbergen, N. (1963) On aims and methods of ethology. Zeitschrift für Tierpsychologie 20: 410–433.

Thompson, K. V. (1998) Self assessment in juvenile play. In: Bekoff, M./Byers, J. A. (eds) Animal Play: Evolutionary, Comparative, and Ecological Perspectives. Cambridge University Press: Cambridge, New York, pp. 183–204.

Tomasello, M./Gust, D./Frost, G. T. (1989) A longitudinal investigation of gesturalcommunication in young chimpanzees. Primates 30: 35–50.

Tomassello, M./Call, J. (1997) Primate Cognition. Oxford University Press: New York.

de Waal, F. B. (1991) Complementary methods and convergent evidence in the study of primate social cognition. Behaviour 118: 297–320.

Watson, D. M. (1998) Kangaroos at play: Play behaviour in the Macropodoidea. In: Bekoff, M./Byers, J. A. (eds) Animal Play: Evolutionary, Comparative, and Ecological Perspectives. Cambridge University Press: Cambridge, New York, pp. 61–95.

Watson, D. M./Croft, D. B. (1996) Age-related differences in playfighting strategies of captive male red-necked wallabies (Macropus rufogriseus banksianus). Ethology 102: 33–346.

Whiten, A./Ham, R. (1992) On the nature and evolution of imitation in the animal kingdom: reappraisal of a century of research. Advances in the Study of Behavior 21: 239–283.

Wilson, S. C./Kleiman, D. G. (1974) Eliciting play: A comparative study. American Zoologist 14: 341–370.

Gerhard Roth

The Neurobiological Basis of Consciousness in Man and Animals

Abstract

In humans, consciousness includes diverse states, from alertness to self-reflection. These states are produced by the interaction of many cortical and subcortical (predominantly limbic) brain centers. However, only processes taking place in the associative parietal, temporal and frontal isocortex are accompanied by consciousness. Consciousness and attention arise whenever the human brain is confronted with tasks for which no pre-established "solutions" and respective neural networks exist. Conscious states are accompanied by the formation of new neural networks based on changes in synaptic connectivity. Most centers of the human brain involved in consciousness are present in the brains of all tetrapods (amphibians, reptiles, birds, mammals). Accordingly, it is likely that most tetrapods possess at least simple states of consciousness such as awareness of sensory events, attention, knowledge representation and analogical thinking. Higher states of consciousness such as taking the perspective of the other and anticipation of future events are only found in primates and, therefore, may require a large isocortex. Finally, conscious states such as comprehension of underlying mechanisms, knowledge attribution, self-awareness and the use of simple syntactical language seem to be restricted to the great apes and may require a large prefrontal cortex. The use of a complex syntactical language seems to be restricted to humans. However, there is evidence that non-human primates have isocortical structures that are homologous to the human WERNICKE and part of the BROCA speech centers. The relationship between higher states of consciousness and syntactical language remains unclear, but both may be due to a much longer period of synaptic plasticity in humans compared to non-human primates.

Key words

Animal consciousness, consciousness brain centers, function of consciousness.

THE QUESTION OF whether animals possess consciousness is as old and as difficult as the mind–body problem. Many philosophers, psychologists and even some neurobiologists strictly deny the existence of consciousness in animals and believe that this phenomenon is unique to humans (for very different reasons, though). Others say: Maybe animals have consciousness, but we will never know for sure, because even in humans consciousness is only accessible to the individual that has it. Again others state: For a number of reasons it is very likely that at least some groups of animals have some of those states of consciousness found in humans. Now, the question is: What kind of hard evidence does exist for the presence of consciousness in at least some groups of animals? In discussing the pieces of evidence within the domain of neurobiology, I will proceed as follows: First, I will give a brief phenomenology of states of consciousness. Then I will discuss what is known about the neurobiological basis of consciousness in man and present two hypotheses about the functional role of consciousness. Finally, I will investigate whether the different groups of vertebrates possess brain centers that are homologous to those that in humans are correlated with states of consciousness and ask how this information can be used for answering the central question about the existence of consciousness in animals.

1. States of consciousness

Consciousness or awareness varies widely with respect to intensity and content ranging from deep coma to the highest degree of concentration, from alertness to self-reflection. The most general form of consciousness is wakefulness or *vigilance*. It is characterized by a general responsiveness to sensory stimuli. There are different states of *reduced* consciousness such as *somnolence* and *stupor* and different types of *coma*. Vigilance is usually combined with *subjective awareness* or *conscious experience* of something. This "something" includes external as well as internal bodily stimuli, my own emotions and my mental activity. From this awareness results the experience of my

own presence in the world. *Attention* is the most characteristic state of increased awareness. A more special type of consciousness is *body-identity awareness*, i.e. the belief that I belong to the body that apparently surrounds me. There is *autobiographic consciousness*, i.e. the conviction that I am the one who existed yesterday. There is awareness of what was going on in the past and is happening in the world surrounding me. There is awareness of *voluntary control of movements and actions*, of *being the author of my thoughts and deeds*. Finally, there is *self-awareness*, i.e. the ability of self-recognition and self-reflection.

These different aspects of consciousness can dissociate, i.e., they can be lost independent of each other after damage of different parts of the brain (KINSBOURNE 1995; KNIGHT/GRABOWECKY 1995; MOSCOVITCH 1995). Accordingly, there are patients who have all normal states of cognition, consciousness and intelligence except that they deny belonging to their own body, and others who do not know *who* or *where* they are.

2. The neurobiological basis of the different states and appearances of consciousness

It appears that we are only aware of those things that are bound to the activity of the so-called associative (i. e., parietal, temporal and frontal) isocortex. However, this is not to say that the associative cortex is the only "producer" of conscious experience; its activity is a necessary, but not sufficient prerequisite for consciousness. Many subcortical parts of the brain, the activity of which is never accompanied by consciousness, contribute in very different ways to the origin of consciousness in the associative cortex. This is true, above all, for the reticular formation, because their destruction leads to a general loss of consciousness, i.e. to coma (HASSLER 1978).

The reticular formation consists of three columns of nuclei, a median, medial and lateral column that extend from the anterior mesencephalon through the pons to the medulla oblongata and rostral spinal cord (NIEUWENHUYS et al. 1991). The medial column receives input from all sensory modalities and the cerebellum as well as descending cortical input via the pyramidal tract. This system called the "extralemniscal" or *ascending reticular activating system*, projects directly or indirectly to the intralaminar thalamic nuclei, which, in turn, project (with some topography) to the striatum and the isocortex (see below). The function of the *medial* column is the control of wake–sleep cycle and of general cortical activity. The *median* column is formed by the so-called raphe nuclei. These nuclei, predominantly the dorsal raphe nucleus, send serotonergic fibers to all parts of the limbic system that are involved in cognitive functions, e.g., hippocampus, amygdala, basal forebrain, limbic thalamic nuclei, cingulate and entorhinal cortex, frontal, parietal and occipital isocortex. The raphe nuclei are supposed to play a modulatory role in the context of behaviorally relevant events, apparently by counteracting and dampening the arousing effect of the other systems and, thus, increasing the signal–to–noise ratio in the cortex (ROBBINS 1997). The *lateral* column contains the noradrenergic locus coeruleus complex, which again projects to all parts of the limbic–cognitive system. The locus coeruleus is supposed to exert a role in controlling attention and continuously "monitoring" the external and internal environment for important events. Its projection to the prefrontal cortex in particular may mediate information about the relevance of complex sensory events and situations (ROBBINS 1997).

The next most important brain centers for the control of consciousness are the limbic *intralaminar* and *midline thalamic nuclei*, because they are the most important relay station for the ascending projections of the reticular formation. They receive input from the entire isocortex and project, with some topography, back to it, predominantly to the prefrontal cortex, and additionally to the striatum; they are also connected to the entire limbic system.

The *nucleus reticularis thalami* surrounds the entire lateral part of the thalamus in a bowl-like fashion. It receives collaterals from thalamo–cortical as well as cortico–thalamic tracts and has reciprocal connections with the sensory and limbic thalamic nuclei. It does not project to the cortex, rather, it exerts inhibitory control over most thalamic nuclei via GABA-ergic fibers. It is assumed to function as a "filter" for various kinds of information coming from the sensory periphery and brainstem, acting under the control of the isocortex and the limbic system. It may, therefore, be involved in guidance of attention (GUILLERY et al. 1998).

The brain centers mentioned so far belong to the limbic system in a wider sense, which in the vertebrate brain is the system that subconsciously controls all aspects of cognitive and emotional states and, accordingly, "voluntary" behavior. The limbic system evaluates what the organism does and stores the result of this evaluation in the different kinds of

memory. Other parts of the limbic centers discussed in the following contribute in more specific ways to the different states of consciousness.

The *basal forebrain–septal nuclei complex* is reciprocally connected with the hippocampus and the amygdala as well as with the centers of the reticular formation already mentioned. Its cholinergic fibers project to all parts of the isocortex. The basal forebrain is involved in the control of attention and of activity of neocortical neuronal network, primarily in the context of earlier experience (VOYTKO 1996).

The *amygdala* is a complex of many different nuclei and is reciprocally connected with the associative isocortex, particularly with the orbitofrontal prefrontal cortex (either directly or via the mediodorsal thalamic nucleus) and the hippocampal formation (AGGLETON 1992). It strongly influences the sensory (visual, auditory, gustatory) isocortex. Subcortically, it receives input from the olfactory system, the medial, intralaminar and midline thalamic nuclei and from the rest of the limbic system. Regarding cognitive and emotional functions, it is believed to be an important center (together with the anterior cingulate cortex, see below) for evaluation and perhaps storage of *negative* experience, e.g., in the context of fear conditioning and anxiety (AGGLETON 1992, 1993). The *hippocampal formation* (AMMON's horn, subiculum, dentate gyrus) and the surrounding parahippocampal and perirhinal (including entorhinal) cortex are important centers for the formation and the consolidation of traces of declarative memory inside the isocortex, i.e., of those kinds of memory that in principle can be consciously retrieved and reported (SQUIRE 1987; MARKOWITSCH 1992, 1999; SQUIRE/KNOWLTON 1995).

The dorsal parts of the *basal ganglia*, i.e. putamen, nucleus caudatus, globus pallidus, are closely associated with the prefrontal, premotor and parietal isocortex as well as with the entire limbic system. The basal ganglia have to do with subconscious planning and final decision in voluntary action under the influence of the limbic system. The so-called *mesolimbic system* (i.e., nucleus accumbens, lateral hypothalamus, ventral tegmental area) is characterized by the neuromodulator dopamine. This system has strong connections with the orbitofrontal cortex and is involved in the formation of positive memories and pleasure and perhaps in the control of attention in the context of new events (ROBBINS/EVERITT 1995, 1996).

All these parts of the brain outside the isocortex substantially contribute to consciousness, while their activities remain completely unconscious. Accordingly, damage to these subcortical centers usually produces either complete loss of consciousness or profound impairment of conscious cognitive and emotional functions. This may include the inability to recognize positive or negative consequences of action (for example as a consequence of lesions in the amygdala or nucleus accumbens), impairment of attention and loss of declarative memory. Importantly, patients are usually unaware of these deficits.

Although activity in the isocortex is necessary for consciousness, we are likewise unaware of processes in the primary and secondary sensory and motor areas of the isocortex, although these processes are necessary for the specific contents of awareness of events inside or outside our body. We are only aware of processes bound to the activity of the cingulate and the associative cortex, and only of some of those processes.

The *cingulate cortex* is that part of the cortex that surrounds the subcortical parts of the telencephalon and the thalamus. It is tightly connected with the prefrontal and parahippocampal cortex, the basal forebrain–septal region, the amygdala, the intralaminar and midline thalamic nuclei and the reticular formation. The anterior part (BRODMANN area 24) is involved in the sensation of pain (in combination with the somatosensory cortex, the medial thalamic nuclei and the so-called central tegmental gray), and in memory of painful events. PET and functional MRI studies demonstrate that it is always active in tasks requiring attention (POSNER 1994).

The *posterior parietal cortex* (PP) shows strong hemispheric asymmetry. The left PP is involved in symbolic–analytic information processing such as mathematics, language, meaning of drawings and symbols. Lesions impair reading and writing and respective memory functions. The right PP involves real and mental spatial orientation, the control of hand and eye movement, change of perspective and control of attention. Lesions of inferior right PP produce neglect, e.g. ignoring the contralateral half of the body or events in the contralateral visual hemifield, or anosognosia, i.e. lack of insight or denial of disturbances.

The *superior* and *middle temporal cortex* includes perception of complex auditory stimuli including (generally on the left side) WERNICKE's semantic speech center, which is crucial for the understanding and the production of meaningful written and spoken language. Perception of music usually involves the right medial temporal cortex. The *inferior* temporal cortex (IT) is decisive for complex visual information regarding non-spatial properties of vi-

sual objects and scenes including their meaning and correct interpretation. Lesions in IT produce object agnosia (left IT), color agnosia (right IT), prosopagnosia, i.e. inability to recognize faces (right or bilateral), deficits in categorization, changes in personality and emotionality and deficits in the use of contextual information.

The *prefrontal cortex* (PFC) includes the largest portion of the isocortex (about 30% in man). Usually, two major parts are distinguished in the primate, including human, brain: a dorsolateral and a ventral–orbitofrontal part. The *dorsolateral* PFC (Brodmann areas 9, 10, 46, and 8) receives its major isocortical input from posterior parietal cortex, somatosensory cortex related to trunk and limbs (superior parietal cortex), associative visual areas (V1, 3 and 4) and the anterior cingulate cortex. Subcortical input comes from the middle (small-celled) part of the mediodorsal thalamic nucleus, the ventral anterior and parafascicular thalamic nuclei, which mediate information from the basal ganglia (ALEXANDER et al. 1990; ROBERTS et al. 1998). The dorsolateral PFC appears to be involved in: (1) attention and selective control of sensory experience, (2) action planning and decision making, (3) temporal coding of events, (4) judgement and insight, particularly with respect to reality, (5) spontaneity of behavior, (6) strategic thinking, (7) associative thinking, and (8) short-term or working memory. Thus, the dorsolateral PFC is predominantly, though not exclusively, oriented toward the external world and its demands including short-term memory.

The *orbitofrontal* PFC (Brodmann areas 12–14, 10, 11, 46, and 8) receives its main intracortical input from the anterior cingulate cortex, the inferotemporal cortex, the inferior somatosensory cortical areas (head, neck, face), the anterior pole of the temporal lobe and the parahippocampal and perirhinal/entorhinal cortex. Subcortical input is mediated mainly by the medial (large-celled) part and portions of the lateral (small-celled) part of the mediodorsal thalamic nucleus (ALEXANDER et al. 1990), which mediate information coming from the entire limbic system regarding emotion, motivation, selective attention. The orbitofrontal PFC is involved in social behavior, ethical considerations, divergent thinking, risk assessment, awareness of consequences of behavior, emotional life and emotional control of behavior. Accordingly, damage to the orbitofrontal PFC results in loss of interest in important life events, loss of "ego", "immoral" behavior and disregard of negative consequences of own behavior. Thus, the orbitofrontal PFC is predominantly oriented toward "internal" emotional and social aspects of life.

In conclusion, there is neither *consciousness per se* nor *a highest brain center* producing consciousness. Different parts of the associative isocortex contribute in different ways to the high diversity and content of consciousness including awareness of external and internal sensory events, consequences of one's own behavior, autobiographic, body and "ego" identity, action planning and authorship of own actions. It does this under the strong influence of the primary and secondary sensory and motor cortices as well as of the subcortical centers mentioned.

3. The functional role of consciousness

In the following, I will briefly present two hypotheses—a lower and a higher order one—about the functional role of awareness of sensory events and particularly attention as the most prominent states of consciousness,. The lower order hypothesis is the following: States of conscious awareness are bound to a high degree of neural activity in associative areas which in turn is bound to three factors: (1) sufficient sensory stimulation through specific thalamic nuclei and primary and secondary cortical areas; (2) adequate stimulation by the reticular formation, reticular and limbic thalamic nuclei, the limbic system, basal ganglia etc. (3) high local metabolic activity based on sufficient supply of oxygen and glucose as a consequence of high local cerebral blood flow. These facts underlie the known imaging techniques such as functional MRI and PET (POSNER 1994; MAGISTRETTI et al. 1999).

Why is increased local metabolic and neural activity of associative isocortical networks a necessary prerequisite for specific states of consciousness? These states occur only when our brain is confronted with something that is sufficiently important for the brain to deal with and for which the brain does *not* have already existing neuronal networks and accordingly pre-established "solutions" (e. g., in the realm of understanding of sentences, motor skill or the recognition of faces). When such pre-set networks exist, we master these demands without paying attention or even without any awareness at all. Rather, attention and directed consciousness often will disturb the performance of automatized motor skills. Accordingly, localized metabolic and neural cortical activity is low during these routine behaviors, as can be visualized through imaging techniques (POSNER 1994).

However, when our brain is confronted with something new and important, e.g., an unexpected or unfamiliar perception, the meaning of a newly heard sentence, handling an unfamiliar tool or practicing a new piece of piano music, we can cope with these tasks only with attention. As a consequence, certain brain regions are activated above normal levels, and this activation is accompanied by an increase in blood flow and cellular metabolism at the expense of other brain regions. In this state, the brain apparently modifies the pattern of synaptic coupling among neurons in those centers that are committed to these functions. For example, when we look at a new face that is important for us, then synaptic coupling in specialized networks in the caudal portion of the lower temporal lobe are modified. This also happens in the BROCA and WERNICKE speech areas, when we try to understand the meaning of a sentence.

In contrast to "normal" neuronal activity, changes in the efficiency of synaptic coupling require increased metabolic energy (sugar and oxygen), probably due to an increase in intracellular signal processing, an increase in receptor density etc. Such synaptic changes are required to form new network properties that deal with the task at hand. If these processes are inhibited by the administration of certain drugs, then awareness, attention and the ability to master new cognitive and motor tasks are likewise hindered.

This hypothesis addresses the necessary neuronal preconditions and correlates of consciousness; it does, of course, not answer the question, *why* we are conscious *at all* when we are confronted with new and important cognitive and motor tasks. A possible answer to this question comes up, when we carefully study the properties of our phenomenal, subjectively experienced world. In this phenomenal world, we are completely unaware of the overwhelmingly complex processing of information from sense organs to the associative cortex. Rather, the sensory world seems to *directly* enter our mind. We may recall the finding of Benjamin LIBET from the late 70s that our consciousness simply ignores the 300–500 ms that are needed for sensory information from sense organs to cortical consciousness and acts as if we would directly sense the sensory world (LIBET 1978). Similarly, our will seems to directly move our muscles, and we are completely unaware of the hyper-complex actions of the cortical and subcortical motor systems. Things in this phenomenal world indeed happen radically differently from how the brain, the sense organs and the muscles work, and it would be absolutely impossible for us to consciously follow sensory processing or to consciously activate or inhibit all muscles involved even in a simple movement. The phenomenal world works on the basis of a *user-friendly code* as opposed to the machine code of the brain. The creation of a phenomenal world, of "actuality", seems to be highly favorable or even necessary for action planning, tactic and strategic thinking and speaking via a complex syntactical language, which would be impossible in terms of the actual neuronal code.

This higher-order hypothesis includes the assumption that during the evolution of great apes an increase in demands of action planning, imagination, strategic thinking and complex syntactical language took place, and this necessitated the construction of a phenomenal world, in which a virtual actor, the *ego*, plans, acts and communicates without caring about how this is actually done by the brain in terms of neuronal activity. Apparently, the invention of a conscious, phenomenal world was a necessary prerequisite for survival in a complex biological and social world. This hypothesis is largely consistent with the view of consciousness as something like a "self model" or "monitoring system", as has been developed by various authors (cf. METZINGER 1993; BAARS 1997). The model presented here in particular stresses the need for reduction in complexity at planning and execution of action.

4. Cognition and consciousness in animals

Checking recent reviews on "animal mind" (STAMP DAWKINS 1993; PARKER et al. 1994; BYRNE 1995; PEARCE 1997; for criticism see MACPHAIL 1998), it seems that all vertebrates and probably invertebrates with large brains (e.g., cephalopods) display sensory or focused attention, extended memory, categorization and the formation of cognitive maps. Whether this is accompanied by some kind of consciousness is difficult to determine in fishes, amphibians or reptiles. Many of so-called higher cognitive functions such as concept learning, knowledge representation, analogical thinking, the formation of abstract representations, and imitation in the sense of copying a behavior are found at least among birds and mammals. Even in humans these higher cognitive functions are not necessarily accompanied by consciousness

There are, however, cognitive functions that in humans require consciousness and for which it is hard to believe that animals but not humans could

carry them out unconsciously. These functions include: (1) Imitation in the sense of task structure or task principle learning and tool learning. This is found in macaque and capuchin monkeys, apes and maybe in some other mammals (otter etc.). (2) Taking the perspective of other individuals in deception and counterdeception. This is found in monkeys (e.g., baboons) and great apes. (3) Anticipation of future events (preparation of tools in advance etc.). This has been found in the great apes and may be present in some monkeys. (4) Comprehension of underlying mechanisms, for example in the use of tools: this has been reported only in great apes. (5) Knowledge attribution/theory of mind: This is found in great apes, particularly chimpanzees. (6) Self-recognition in mirror: This is found in great apes and dolphins. (7) Distinction between appearance and reality: This is present only in chimps and dolphins. (8) Playing with imaginated objects: This is found only in chimpanzees. (9) Teaching: This again is found only in chimpanzees. (10) Understanding and using simple syntactical language (up to three-word sentences): This is found in great apes and dolphins. (11) Using complex syntactical language: This is found only in humans.

Thus, a few of the mentioned "higher cognitive functions" may be found in birds and all mammals, but most of them are found either only in primates and dolphins, only in the great apes and dolphins, only in great apes or even only in chimpanzees and humans, and at least one function (viz., the use of complex syntactical language) only in man. Being aware of the risk of an anthropomorphic perspective, we might speculate about "jumps" in cognition and states of consciousness among tetrapods from amphibians and reptiles to birds and mammals, among mammals to cetaceans and primates and among primates from monkeys to the great apes and from non-human apes to *Homo sapiens*.

Functions (4), (5), (6) and (10) have been extensively studied in the recent past and are of special interest, because they might draw a borderline between monkeys and great apes among the primates. Such a borderline is less emotionally laden than that between humans and non-human apes. Primatologists almost unanimously agree that monkeys (e.g., capuchins) exhibit the use of tools (as many other animals do), but without an understanding of the underlying mechanism. They do not seem to "know" why one tool is effective and another is not (VISALBERGHI/LIMONGELLI 1994). A similar lack of "insight" has been reported by KUMMER on baboons (using a stick for getting into reach of apples). Even more important are the experiments concerning mirror self-recognition (PARKER et al. 1994). Monkeys are capable of making use of mirrors, e.g. in order to look behind objects that otherwise are inaccessible (BYRNE 1995). Kummer reports that baboons recognize a group member on a slide and identify themselves as mother or child on slides without difficulty. Yet, primatologists agree that monkeys show no sign of mirror self-recognition. Among apes, chimpanzees and orangutans show mirror self-recognition, while among gorillas only Koko does (it is assumed that gorillas often are too shy). Chimpanzees show great interest in their mirror image and pass the marking test (i. e., removing marks of paint from face or body using the mirror image) well. Interestingly, at least some dolphins are reported to show mirror self-recognition, too, but their behavior appears to be very different in front of a mirror. They show no natural interest in their mirror image and dislike the marking test (MARTEN/PSAKAROS 1994). The reasons for these differences are unclear, besides many methodological and conceptual problems.

The presence of knowledge attribution or theory of mind to others (BARON-COHEN et al. 1985) is difficult to distinguish from the capability of taking the perspective of the other. For most animals with complex behavior, it is important to guess what the other is going to do, e.g. in agonistic behavior (STAMP DAWKINS 1993). However, primatologists (e.g., KUMMER) definitely state that monkeys (e.g., baboons) do not take into account what the other is "thinking". Important in this context is the performance of very young children. According to PARKER et al. (1994), in children self-detection starts with 3 months, and mirror self-recognition is shown at an age of 18 months on average. However, little children show embarrassment and coy reactions earlier than that. Self-recognition in photos starts at 24 months, followed by signs of self-evaluative emotions, e.g., shame and pride. A true "theory of mind" is said to emerge at 4 years (MELTZOFF/GOPNIK 1993). The authors stress that these events occur in the same sequence in non-human and human apes, though at a much slower rate in the former.

Much has been written about the presence or absence of a syntactically complex, language in non-human primates (and other animals). Monkeys have complex systems for intraspecific vocal communication that are able to express relatively complicated types of meaning including symbolic information (i.e., about objects and events that are not present) or about relationships between events (e.g.,

parental relationships). Many of these calls have to be learned by the infant monkey and—as in songbirds—dialects exist (cf. ZIMMERMANN et al. 1995). Most authors agree that sentences consisting of up to three words are understood and used by chimpanzees, gorillas and dolphins. Whether this is a sign for a "simple" syntax is a matter of debate. Chimpanzees do combine words to form new words (GARDNER et al. 1989), but they do not go beyond the linguistic capabilities of a 3-year-old child, even after intense training (SAVAGE-RUMBAUGH 1984).

Of great interest in this context is the development of language in human children (LOCKE 1995; STROMSWOLD 1995). In humans, language learning starts long before speaking, viz., in utero. Facial learning starts immediately after birth; babies seem to be pre-adapted to it. At 6 months after birth, children show a preference for familiar over foreign language. First voweling sounds are produced around 4–6 months, when "babbling" starts, first consonant-like sounds occur at 9–12 months. First words are produced between 8–20 months, (average 12 months). Around 18–23 months, children begin to combine words and form 2-word utterances. Then, they gradually begin to use sentences longer than two words, but typically in a "telegraphic" style similar to that found in patients with lesions of BROCA's speech area. In the third year, children start distinguishing singular and plural and asking the famous question "what's that?". From there on, their language syntactically becomes rapidly more elaborate. Between 3 and 4 years, utterances are completely grammatical.

Thus, it seems that apes have simple linguistic capabilities comparable to a human child of 2–3 years, but from then on the child rapidly develops a fully syntactical language that is far superior to any communicative system known from animals.

5. Animal brains and human brain

Are we able to correlate the above discussed differences in higher cognitive functions including consciousness among groups of vertebrates with properties of their brains? When answering this question, I will restrict myself to the discussion of the tetrapod brain, because the components of the forebrain of cartilaginous and bony fishes are difficult to homologize with those of tetrapods. In the following, I will briefly discuss the following hypotheses which in the past have been forwarded in order to explain the alleged evolutionary increase in cognitive capabilities from frog to man. Are the above differences due to differences in (1) overall organization of brains, (2) absolute or relative brain size, (3) absolute or relative size of parts of the brain, e.g., cortex or subcortical centers, (4) size of the association cortex and prefrontal association cortex in particular, or (5) anatomical or physiological features of the cortex or subcortical centers?

(1) All tetrapods have brains that—despite enormous differences in outer appearance, overall size and relative size of major parts of the brain—are very similar in their overall organization and even in many details (ROTH/WULLIMANN 1996). All tetrapod brains possess a median, medial and lateral reticular formation inside the medulla oblongata and the ventral mesencephalon. In the dorsal diencephalon, an epithalamus, a dorsal and ventral thalamus and a subthalamus are found, and it seems that they also share the distinction between "specific" (i.e., those involved in sensory and motor functions) and "non-specific", i.e., limbic nuclei. There is a corpus striatum, globus pallidus, nucleus accumbens, basal forebrain/septum and amygdala within the ventral telencephalon, a lateral pallium, homologous to the olfactory cortex of mammals, and a medial pallium, homologous to the hippocampal formation (at least AMMON's horn and subiculum). Thus, all structures required for attention, declarative memory (or its equivalents in animals), motivation, guidance of voluntary actions and evaluation of actions are present in the tetrapod brain. These structures have essentially the same connectivity and distribution of transmitters, neuromodulators and neuropeptides in the different groups of tetrapods.

A more difficult problem is the presence of a homologon to the mammalian isocortex and associative cortex in the telencephalon of other tetrapods. Amphibians possess a dorsal pallium, reptiles have a dorsal cortex plus a dorsal ventricular ridge (DVR), birds have a wulst and a DVR, and these structures are believed by many comparative neurobiologists to be homologous to the isocortex of mammals (KARTEN 1991; NORTHCUTT/KAAS 1995). However, major differences exist with regard to cytoarchitecture and size of the dorsal pallium/isocortex. In amphibians, the dorsal pallium is small and unlaminated; in lizards it is relatively larger and in some groups it shows a three-layered structure. In birds, those parts assumed to be homologous to the mammalian isocortex (i. e., DVR and wulst) are large, but unlaminated. In mammals excluding insectivores and cetaceans, the dorsal pallium or isocortex shows the characteristic six-layered structure. But what does a six-layered cortex mean? It has not been dem-

onstrated that mammals are "smarter" than birds with the same brain size or size of cortex/pallium. To date, there is no convincing answer to this important question. Furthermore, the pigeon is considered by many behaviorists to be very smart, although it has a brain and accordingly an unlaminated dorsal pallium/cortex that is much smaller than that of a rat both in absolute and relative terms. Apparently, the same or very similar cognitive functions are performed by anatomically very different kinds of pallium/cortex.

(2) Brain size—either in absolute terms or relative to body size—has often been correlated with "intelligence" or "higher cognitive abilities", but the reasons for increase in brain size are unclear. Body size appears to be the single most important factor influencing brain size, i.e., large animals generally have large brains. Increases and decreases in body size have occurred many times in virtually all animal groups and seems to happen fast (K-selection seems to favor large, r-selection small bodies), and the brain seems to follow these changes with some delay (VAN DONGEN 1998). However, increase in brain size does not linearly parallel increase in body size, but only to the power of about 0.67, which means that small animals have relatively large brains and large animals have relatively small brains (JERISON 1973). Among mammals, body size ranges from 3 g (the insectivore *Suncus etruscus*) to 150 tons (*Balaenopterus musculus*, blue whale), which is a ratio of 1: 50 millions; brain size ranges between 74 mg (the bat *Tylonycteris pachypus*) to 10 kg (sperm whale), which is a ratio of 1: 130.000, i.e., still an enormous but much smaller range than body size. Among mammals, this is reflected by the fact that in very small rodents brains occupy up to 10% of body mass, in pigs 0.1% and in the blue whale, the largest living mammal and animal 0.01%.

Nobody really knows what relative—rather than absolute—brain size means. Since cell size does not generally increase with brain size, absolutely larger brains mean more neurons, and there is good reason to believe that more neurons are better than fewer, be that for faster and more sophisticated information processing, storage of memory traces, higher neuronal plasticity, better compensation of lesions.

Among vertebrates, we find many independent cases of increases (as well as decreases) in absolute and relative brain size. In relative terms, chondrichthyans (sharks and rays) on average have large brains (with large telencephala), while osteichthyes (bony fishes) have rather small brains (with notable exceptions). Amphibians have small to very small brains, reptiles have small brains, birds relatively large brains (particularly passerine birds and parrots; the "intelligent" pigeon, however, has a remarkably small brain), as do mammals in general. Among mammals, very large brains both in absolute and relative terms are found in primates (1.4 kg in *Homo sapiens*), in elephants (up to 5.7 kg).and in whales including dolphins (up to 10 kg). Primates generally have larger brains in relative terms than any other group of mammals except dolphins. Contrary to a common belief, new-world monkeys have larger brains than old-world ones (which include *Homo sapiens*). Apes do not exceed monkeys in relative brain size except man who has a brain that is 2–3 times larger than an ape or monkey of the same size. During 3.5 million years of evolution of *Homo sapiens*, brain size has increased by a factor of 3–4, from 350 to 1400 ccm, but despite many scenarios, the reasons for this increase remain obscure.

(3) The isocortex of "higher" mammals and particularly man has been the target of many anatomical and physiological studies, because it is believed to be the seat of higher cognitive functions (see above discussion). However, while there are enormous differences both in absolute and relative brain size among tetrapods and mammals in particular, changes in relative size of the isocortex are relatively inconspicuous, because at least in mammals they more or less strictly follow changes in brain size. In mammals, the correlation coefficient between brain volume and cortical volume is almost exactly 1 (JERISON 1997). This means that cortex size increases more or less isometrically with brain size. In man, the amount of neocortex is only slightly increased (76% compared to 72% in chimpanzee). Humans have a large isocortex because they have a large brain, and accordingly elephants and most cetaceans have much larger cortices than man. Isometric increase relative to total brain volume is found in many other brain structures including the thalamus (which is tightly connected to the isocortex), the basal ganglia (the largest subcortical structure in the telencephalon) and the hippocampus (JERISON 1997).

In the cortex of large-brained mammals, sensory, motor and association areas show differences in cytoarchitecture. Typically, with respect to the presence and thickness of layers with small or larger cell bodies, a heterotypic, granular motor cortex, a homotypic, agranular frontal, parietal and temporal association cortex and a heterotypic, granular sensory cortex (also called "koniocortex") are distinguished in the cortex of large-brained mammals

(CREUTZFELDT 1983). Besides these differences in cytoarchitecture, there is great uniformity of cortex with regard to the presence of cortical cells except for the motor cortex, where small pyramidal cells are nearly absent. Rather, the different parts of the cortex are mainly characterized by their major thalamic input: The prefrontal cortex receives its main input from the medial dorsal thalamic nucleus, the precentral–motor cortex from the ventral lateral nucleus, the anterior parietal cortex from the ventral posterior nucleus, the parieto–temporo–occipital cortex from the pulvinar, the supratemporal–auditory cortex from the medial geniculate, the occipital–visual cortex from the lateral geniculate and the limbic–cingular cortices including the temporal pole from the anterior thalamic nucleus. There is no sign of something unique to humans or other primates in regard to these cortical features including those of the association cortex.

(4) There is a common belief that during hominid brain evolution the association cortex has increased dramatically. However, there are no precise criteria for distinguishing primary and secondary sensory cortical areas from true association areas. Recently, KAAS (1995) has argued that the number of isocortical areas has strongly increased from the hypothetical insectivore-like ancestor with about 20 such areas to more than 60 in primates. However, what—according to KAAS—has increased is the number of functionally "intermediate" areas, but not the primary or "highly associative" ones. Furthermore—as KAAS rightly emphasizes –, there is the danger to strongly underestimate the number of functionally different cortical areas in small-brained mammals. Thus, whether associative cortex has in fact substantially increased in relative terms with increase in brain and cortical size, remains undecided so far.

The difficulties with defining the prefrontal cortex anatomically in the various groups of mammals are even greater. UYLINGS/VAN EDEN (1990) define—according to the neuroanatomical tradition—the PFC as the cortical area with major (though not exclusive) input from mediodorsal thalamic nucleus. Using this definition, the volume of PFC relative to the rest of the isocortex does not differ dramatically between the rat (24%) and *Homo sapiens* (30%). Furthermore, the PFC of the orangutan is equal to humans in relative terms, while in monkeys it varies between 11 and 18%. In general, PFC increases isometrically with an increase in cortical and overall brain volume (JERISON 1997).

(5) Remains the question of any cortical or subcortical anatomical or physiological specializations in the cortices of "higher" mammals that could be correlated with the observed "higher" cognitive abilities. So far, only two features have been discovered that could distinguish the human cortex/brain from that of other primates, viz., (1) differences in growth rate and (2) the presence of speech centers.

As to growth, in humans, the isocortex is immature at birth, with simple afferent connections and "primitive" intracortical connections (CREUTZFELDT 1983). Postnatal maturation consists in explosion-like increase in number of synapses in all parts of the brain, increase in size of pyramidal neurons and diameter of axons, increase in size in smaller cortical neurons, increase in number of glial cells, growth of intracortical capillary network and therefore strong decrease in neuron density. This maturation is nearly completed 3 months after birth in monkeys and at 3 years in humans, but the human brain still continues to mature until the age of 20, which is much longer than in any other primate. A critical phase in the development of the human brain seems to occur around 2.5 years. At this time, major anatomical rearrangements in the association cortex have come to a stop, and the period of "fine-wiring" seems to start, particularly in layer 3 of the prefrontal cortex (MRZLJAK et al. 1990). As mentioned above, at this time, human children cognitively "take off" compared to non-human primates.

The other exception concerns the presence of so-called speech centers in the human brain, particularly the BROCA speech center in the frontal lobe responsible for temporal aspects of language including syntax, and the WERNICKE speech center in the temporal lobe responsible for the meaning of words and sentences (although meaning is likewise dependent on syntax and grammar). However, it is to date unclear whether these speech centers really are evolutionary novelties. Several explanations have been offered for the absence or strong limitations of human-like speech in non-human primates. It has been argued that the vocal tract of non-human primates is incapable of producing sounds typical of and necessary for the human language. However, the capabilities of the vocal tract of non-human primates appear to exceed previous assumptions and the articulatory gestures associated with the non-human primate vocal output are more complicated than originally thought. Yet, some inherent sound-producing limitations of this apparatus cannot be denied, particularly with respect to vowels and formant transitions. If these were the only limitations, we would expect the great apes to naturally produce or learn from humans a much richer vocal language

mostly based on consonants. Thus, there seem to be differences in the neural control of vocal language in addition to the restrictions of the vocal apparatus. What could these be?

All mammals studied so far have a center for intraspecific communication within the temporal lobe (mostly left) which may be homologous to the WERNICKE center for semantics. It has been reported that destruction of these areas leads to deficits in intraspecific vocal communication (HEFFNER/HEFFNER 1995). In addition, it has long been argued that the posterior part (A 44) of the BROCA speech center in humans and the ventral premotor area (PMV) of non-human primates are similar in location and cytoarchitecture and probably are homologous (PREUSS 1995). PMV controls movement of forelimbs, face and mouth, which is the case with the posterior portion of the BROCA area. However, JÜRGENS (1995) emphasizes that in non-human primates (e.g., squirrel monkeys), intraspecific vocal communication does *not* involve isocortical functions. According to JÜRGENS, in primates including man, critical structures for "innate" sound production are the mesencephalic periaqueductal gray (PAG) and laterally bordering tegmentum. Destruction of these two centers alone produces mutism in monkeys and man. The PAG is controlled by sensory input via mesencephalon and medulla oblongata (e.g., colliculus superior and inferior) and in turn controls the motor nuclei innervating the laryngeal, oral and respiratory muscles involved in vocalization.

While in humans, isocortical centers (e.g., BROCA and WERNICKE) are obligatory for voluntary control of vocalization, in non-human primates, only the anterior cingulate cortex is believed to be involved in this function. Destruction of this cortical area abolishes self-initiated vocalization in the monkeys, while in man, it abolishes all kinds of emotional control of vocalization, and utterances become very monotonous, but patients can still speak. JÜRGENS (1995) concludes that the anterior cingulate cortex is the main center for voluntary control of *emotional* utterances. This structure is directly connected with the PAG and functions *in parallel* to the isocortical speech centers. According to JÜRGENS, non-human primates lack a direct connection between the motor cortex and the laryngeal motor neurons. In man, bilateral destruction of the facial motor cortex abolishes the capacity to produce learned vocalization including speech or humming a melody, while a similar destruction in monkeys has no such consequence.

According to this view, the evolutionary basis for human language was an emotionally driven language typical of non-human primates. During hominid evolution, the isocortex gained control over this system such that beyond the initiation of innate vocalization a "free" production of sounds and their sequences became possible. Such an interpretation, however, is in contrast with recent evidence of a high degree of sound learning in monkeys (ZIMMERMANN 1995) and the mentioned consequences of destruction of left-hemispheric, WERNICKE-like temporal areas.

Leaving this problem unsolved, non-human primates including the great apes are strongly limited even in non-vocal speech based on the use of sign language or symbols, and these limitations seem to concern mostly syntax. Accordingly, if something relatively new concerning language developed or became substantially modified in the human brain, it was probably the BROCA center and not so much the WERNICKE center, which already may have existed in non-human primates. Such an assumption is consistent with the fact that the most clear-cut differences between humans and non-human primates concern syntactical complexity of language (see above). Thus, a reorganization of the frontal–prefrontal cortex may have happened such that the facial and oral motor cortex and the related subcortical speech centers came under the control of a kind of isocortex that is specialized in any aspects of temporal sequence of events including the sequence of action (DEACON 1990). This would be consistent with the fact that the posterior part of BROCA's area appears to be homologous with the ventral premotor area of non-human primates. Thus, only the anterior part of BROCA's area would be a "new" structure. To date, there is no good idea when and how such an evolutionary step might have occurred.

For many reasons, it is unlikely that consciousness is strictly bound to the existence of syntactical language and originated with it. It is clear, however, that the evolution of a syntactical language strongly favored higher states of consciousness including self-reflection, thinking and action-planning. While thinking is not necessarily bound to language, most people think and plan verbally. Furthermore, many concepts typical of human mind "exist" only linguistically, i.e., because we can *talk* about them, e.g., future events or abstract entities such as society and freedom.

Certainly, the presence states of consciousness were and are of great advantage for an individual, particularly regarding social life. In this direction goes the

"MACHIAVELLIAN intelligence" hypothesis of BYRNE (1995). According to this hypothesis, social environment exerted a strong selection pressure resulting in self-recognition, empathy, imitation, pretend play, deception, theory of mind, and teaching. All this was and is used for *social manipulation*. However, equally important was and is the ability to deal with a complex natural and social environment, to make complicated choices and to carefully plan one's own actions based on previous experience, to modify behavioral strategies quickly at new demands.

Author's address

Gerhard Roth, University of Bremen, Brain Research Institute, D-28334 Bremen, Germany. Email: gerhard.roth@uni-bremen.de

References

Aggleton, J. P. (1992) The Amygdala: Neurobiological Aspects of Emotion, Memory, and Mental Dysfunction. Wiley-Liss: New York, Chichester.

Aggleton, J. P. (1993) The contribution of the amygdala to normal and abnormal emotional states. Trends in Neurosciences 16: 328–333.

Alexander, G. E./Crutcher, M. D./DeLong, M. R. (1990) Basal ganglia–thalamocortical circuits: Parallel substrates for motor, oculomotor, "prefrontal" and "limbic" functions. In: Uylings, H. B. M./van Eden, C. G./de Bruin J. P. C./ Corner, M. A./Feenstra, M. G. P. (eds) The Prefrontal Cortex. Its Structure, Function and Pathology. Elsevier: Amsterdam, New York, Oxford, pp. 119–146.

Baars, B. J. (1997) In the Theater of Consciousness: The Workspace of the Mind. Oxford University Press: Oxford.

Baron-Cohen, S./Leslie, A. M./Frith, U. (1985) Does the autistic child have a "theory of mind"? Cognition 21: 37–46.

Byrne, R. (1995) The Thinking Ape. Evolutionary Origins of Intelligence. Oxford University Press: Oxford, New York, Tokyo.

Creutzfeldt, O. D. (1983) Cortex Cerebri. Leistung, strukturelle und funktionelle Organisation der Hirnrinde. Springer: Berlin, Heidelberg, New York.

Deacon, T. W. (1990) Rethinking mammalian brain evolution. American Zoologist 30: 629–705.

Gardner, R. A./Gardner T. B./van Cantfort T. E. (1989) Teaching Sign Language to Chimpanzees. State Univ. New York Press: New York.

Guillery, R. W./Feig, S. L./Lozsádi D. A. (1998) Paying attention to the thalamic reticular nuclei. Trends Neurosci. 21: 28–32.

Hassler, R. (1978) Interaction of reticular activating system for vigilance and the truncothalamic and pallidal systems for directing awareness and attention under striatal control. In: Buser, P. A./Rougeul-Buser, A. (eds) Cerebral Correlates of Conscious Experience. Elsevier/North-Holland: Amsterdam, pp. 111–129.

Heffner, H. E./Heffner R. S. (1995) Role of auditory cortex in the perception of vocalization by Japanese Macaques. In: Zimmermann, E./Newman, J. D./U. Jürgens (eds) Current Topics in Primate Vocal Communication. Plenum Press: New York, London, pp. 207–219.

Jerison, H. J. (1973) Evolution of the Brain and Intelligence. Academic Press: New York.

Jerison, H. J. (1997) Evolution of prefrontal cortex. In: Krasnegor, N. A./Lyon, G. R/Goldman-Rakic, P. S. (eds) Development of the Prefrontal Cortex: Evolution, Neurobiology, and Behavior. Brookes Publ. Company: Baltimore, London, Toronto, Sydney, pp. 9–26.

Jürgens, U. (1995) Neuronal control of vocal production in non-human and human primates. In: Zimmermann, E./ Newman, J. D./Jürgens, U. (eds) Current Topics in Primate Vocal Communication. Plenum Press: New York, London, pp. 199–206.

Kaas, J. H. (1995) The evolution of isocortex. Brain Behav. Evol. 46: 187–196.

Karten, H. J. (1991) Homology and evolutionary origins of the "neocortex". Brain Behav. Evol. 38: 264–272.

Kinsbourne, M. (1995) Models of consciousness: Serial or parallel in the brain? In: Gazzaniga, M. S. et al. (eds) The Cognitive Neurosciences. MIT Press: Cambridge MA, pp. 1321–1329.

Knight, R. T./Grabowecky M. (1995) Escape from linear time: Prefrontal cortex and conscious experience. In: Gazzaniga, M. S. et al. (eds) The Cognitive Neurosciences. MIT Press: Cambridge MA, pp. 1357–1371.

Libet, B. (1978) Neuronal vs. subjective timing for a conscious sensory experience. In: Buser P. A./Rougeul-Buser, A. (eds) Cerebral Correlates of Conscious Experience. Elsevier/North-Holland: Amsterdam, pp. 69–82.

Locke, J. L. (1995) Linguistic capacity: An ontogenetic theory with evolutionary implications. In: Zimmermann, E./Newman/J. D./Jürgens, U. (eds) Current Topics in Primate Vocal Communication. Plenum Press: New York, London, pp. 253–272.

MacPhail, E. (1998) The Evolution of Consciousness. Oxford University Press, Oxford: New York, Tokyo.

Magistretti, P. J./Pellerin, L./Rothman, D. L./Shulman, R. G. (1999) Energy on demand. Science 283: 496–497.

Markowitsch, H. J. (1992) Neuropsychologie des Gedächtnisses. Hogrefe: Göttingen.

Markowitsch, H. J. (1999) Gedächtnisstörungen. Kohlhammer: Stuttgart.

Marten, K./Psakaros, S. (1994) Evidence of self-awarenes in the bottlenose dolphin (Tursiops truncatus). In: Parker, T./ Mitchell, R. W./Boccia, M. L. (eds) Self-Awareness in Animals and Humans: Developmental Perspectives. Cambridge Univ. Press: Cambridge, pp. 361–379.

Meltzoff, A./Gopnik, A. (1993) The role of imitation in understanding persons and developing a theory of mind. In: Baron-Cohen S./Tager-Flusberg, H./Cohen, D. J. (eds) Understanding Other Minds: Perspectives from Autism. Oxford University Press: Oxford, pp. 335–366.

Metzinger, T (1993) Subjekt und Selbstmodell. Die Perspektivität phänomenalen Bewußtseins vor dem Hintergrund einer naturalistischen Theorie mentaler Repräsentation.

Schöningh: Paderborn.

Moscovitch, M. (1995) Models of consciousness and memory. In: Gazzaniga, M. S. et al. (eds) The Cognitive Neurosciences. MIT Press: Cambridge MA, pp. 1341–1356.

Mrzljak, L./Uylings, H. B. M./van Eden, C. G./Judás, M. (1990) Neuronal development in human prefrontal cortex in prenatal and postnatal stages. In: Uylings, H. B. M./van Eden, C. G./de Bruin J. P. C./Corner, M. A./Feenstra, M. G. P. (eds) The Prefrontal Cortex. Its Structure, Function and Pathology. Elsevier: Amsterdam, New York, Oxford, pp. 185–222.

Nieuwenhuys, R./Voogd J./van Huijzen, C. (1988) The Human Central Nervous System. Springer: Berlin, Heidelberg, New York. German edition: Nieuwenhuys, R., Voogd, J./ Huijzen, C. van (1991) Das Zentralnervensystem des Menschen. Springer: Berlin, Heidelberg, New York.

Northcutt, R. G./Kaas, J. H. (1995) The emergence and evolution of mammalian isocortex. Trends in Neurosciences 18: 373–379.

Parker, S. T./Mitchell, R. W./Boccia, M. L. (1994) Self-awareness in Animals and Humans: Developmental Perspectives. Cambridge Univ. Press: Cambridge MA.

Pearce, J. M. (1997) Animal Learning and Cognition. Psychology Press: Exeter.

Posner, M. I. (1994) Seeing the mind. Science 262, 673–674.

Preuss, T. M. (1995) Do rats have a prefrontal cortex? The Rose–Woolsey–Akert program reconsidered. J. Cognitive Neurosci. 7: 1–24.

Robbins, T. W. (1997) Arousal systems and attentional processes. Biol. Psychology 45: 57–71.

Robbins, T. W./Everitt, B. J. (1995) Arousal systems and attention. In: Gazzaniga, M. S. et al. (eds) The Cognitive Neurosciences. MIT Press: Cambridge MA, pp. 703–720.

Robbins, T. W./Everitt, B. J. (1996) Neurobehavioural mechanisms of reward and motivation. Current Opinion in Neurobiology 6: 228–236.

Roberts, A. C./Robbins, T. W./Weiskrantz, L. (1998) The Prefrontal Cortex. Executive and Cognitive Functions. Oxford University Press: Oxford, New York, Tokyo.

Roth, G./Wullimann, M. F. (1996) Die Evolution des Nervensystems und der Sinnesorgane. In: Dudel J./Menzel, R./ Schmidt, R. F. (eds.), Lehrbuch der Neurowissenschaft. VCH: Weinheim, pp. 1–31.

Savage-Rumbaugh, S. (1984) Acquisition of functional symbol usage in apes and children. In: Roitblat, H. L./Bever, T. G./Terrace, H. S (eds.): Animal Cognition. Earlbaum: Hillsdale, New Jersey, pp. 291–310.

Squire, L. R. (1987) Memory and Brain. Oxford University Press: New York.

Squire, L. R./Knowlton, B. (1995) Memory, hippocampus, and brain systems. In: Gazzaniga, M. S. et al. (eds) The Cognitive Neurosciences. MIT Press, Cambridge MA, 825–836.

Stamp Dawkins, M. (1993) Through Our Eyes Only? The Search for Animal Consciousness. W. H. Freeman/Spektrum: Oxford, New York, Heidelberg.

Stromswold, K. (1995) The cognitive and neural bases of language acquisition. In: Gazzaniga, M. S. et al. (eds) The Cognitive Neurosciences. MIT Press: Cambridge MA, pp. 855–870.

Uylings, H. B. M./van Eden, C. G. (1990) Qualitative and quantitative comparison of the prefrontal cortex in rat and in primates, including humans. In: Uylings, H. B. M./van Eden, C. G./de Bruin J. P. C./Corner, M. A./Feenstra, M. G. P. (eds) The Prefrontal Cortex. Its Structure, Function and Pathology. Elsevier: Amsterdam, New York, Oxford, pp. 31–62.

Van Dongen, P. A. M. (1998) Brain size in Vertebrates. In: Nieuwenhuys, R./ten Donkelaar, H. J./Nicholson, C. (1998) The Central Nervous System of Vertebrates, Vol. 3. Springer: Berlin, pp. 2099–2134.

Visalberghi, E./Limongelli, L. (1994) Lack of comprehension of cause–effect relationships in tool-using capuchin monkeys (Cebus apella). J. Comp. Psychol. 108: 15–22.

Voytko, M. L. (1996) Cognitive functions of the basal forebrain cholinergic system in monkeys: Memory or attention? Behav. Brain Res. 75: 13–25.

Zimmermann, E./Newman, J. D./Jürgens, U. (1995) Current Topics in Primate Vocal Communication. Plenum Press: New York, London.

Zimmermann, E. (1995) Loud calls in nocturnal prosimians: Structure, evolution and ontogeny. In: Zimmermann, E./ Newman, J. D./Jürgens, U. (eds) Current Topics in Primate Vocal Communication. Plenum Press: New York, London, pp. 47–72.

Franz Seitelberger

Neurobiological Comments on Empirical Consciousness

Abstract

The unexplainable categorial difference in the structural fundament of the human mental life is characterised and discussed regarding I. KANT´s transcendental concept and some topical thories of the psychophysical problem. Respective facts of neuroscientific brain research and recent results of studies concerning the unique human capacity of language and of interindividual communication are mentioned.

Key words

Conciousness, self-sentience, object–subject–ego relation; brain–mind duality, information processing.

THE PHENOMENON OF consciousness is considered as the main crossing point between neurosciences and cognitive sciences. The wealth of pertinent literature hardly can be surveyed and the problem solving strategies to the respective paramount problems are frequently repeated without an outlook of progress.

Before starting a renewed discussion on crucial points of the subject it seems advisable to look at the primary notion of the word consciousness as central object of the following remarks. Linguistically, 'consciousness' is a so-called abstract–noun of predicative–adjectival significance. Consciousness means to be conscious. Accordingly there are facts given in the consciousness but there doesn't exist the consciousness given as an object of science. This has clearly been expressed by FICHTE: "In each consciousness something is given as conscious in somebody, which, however, is not the consciousness itself." (FICHTE 1845/46)

In WUNDT's words: "Consciousness always is consciousness of something, namely of our inner experiences of the states and processes in us!" (WUNDT 1873/74).

Empirical consciousness

What, however, in phenomenological analysis is the substrate of this abstraction? Consciousness refers to a certain quality of a living being as a manifestation of it`s individual behavior. This is the primary experience underlying the empirical or biological notion of consciousness. With psychological consciousness one understands a united individual experience consisting of two components:

(a) someone who has, i.e., disposes consciousness, a quality of the living subject, and

(b) somewhat which occurs in such a consciousness, i.e., the contents of consciousness, resp. objects for a subject.

In the psychological consciousness thus a "self-conscious" part as concrete subjective experience is given, namely not as an object of perception but as a concomitant or, obligatorily enclosed self-conscious part as concrete experience. This "self conscious" part is not a object of sensation, but a concomitant or a obligatorily enclosed self-sentience of the conscious subject connected with an indistinct bodily and intentional feeling vis–a–vis or directed to the appearing objects: in any conscious state this ego-feeling appears in different degrees of clearness. It is discussed as faint allusion of the individual´s private life process. The occurring consciousness continuously unifies both unequal parts into the process of individual consciousness. In the sensation–perception of pain which is not a sense like vision or audition the self-feeling becomes prominent, initiates self-protection and is experienced as one source of the self-consciousness, whereas the other one consists in the neuronal self-representation of the conscious ego.

From the basic duality of consciousness the permanently moved and intangible character of the ego results. The ego, because it itself is the factual subject and lacks a sensory device, due to his immediacy never can become an object to the conscious-

ness. The body too is not the ego but is the perceptually objectivated ego-experience.

The other part of consciousness consists of the conscious contents, i.e., events, imaginations, thoughts, emotions, drives etc.. These are the objects themselves, experienced in external perception, embedded into subjective conscious experience. Only within the contents of consciousness physical entities and reality is immediately given to us, as one fundamental element of our dual experience consisting of the reciprocal opposition between subject and object.

As already indicated, the quantitative portion and distinctness of the basic stationary self-awareness given in consciousness may vary considerably over time, e.g., depending on the vital impact of autonomous cognitive activities. This also reflects the asymmetric relation between the conscious subject and his surrounding world. Even more important, it points to the apparent paradox that this inner duality of consciousness regarding self-sentience versus events and self-activity forms the unifying clamp of the different aspects of individual existence, so to say the invariant function of the human living system in spatial–temporal dimension.

Excursion on transcendental consciousness

Although in this treatise a systematical philosophical discourse on consciousness is not intended KANT's transcendental notion of consciousness should be mentioned because it possesses an indispensable function for orientation within the problem of consciousness in general. The central formulation by I. KANT in the original text reads as follows:

"Das ich denke muß alle meine Vorstellungen begleiten können denn sonst würde etwas in mir vorgestellt werden, was gar nicht gedacht werden könnte welches ebensoviel heißt, als: Die Vorstellung würde entweder unmöglich oder wenigstens für mich nichts sein." (KANT 1922)

In other words: Everything existing for me has to be a phenomenon in my consciousness. 'Conscious' in this context means to be an object for a subject or for an ego, respectively. The latter is possibly not something real. Rather, it is the virtual frame of various conscious experiences. If the subject by reflection towards his self changes to an object, then it will loose the ego-character and will exist in this context only as representation of oneself but not as the adequate conscious self.

Thus the transcendental ego does not mean an empirical ego but concerns an impersonal formal subject. The transcendental consciousness, however, is not an empty idealistic construct but the result of high-level abstraction processes, in KANT's often misunderstood words "an idea of reason". This concept represents the fundamental fact of "esse est percipi" and the inherent subject–object correlation in our perception of reality it appears as constitutive for the possibility of experience and scientific exploration of the world, because, following KANT, the empirical consciousness, i.e., the subject–object relation, possesses "a necessary relation to a transcendental consciousness, namely to the consciousness of my self as the primary apperception" (KANT 1922). And: "The synthetic sentence, that each different empirical consciousness had to be connected to an only self-consciousness, is plainly the first and synthetic principle of our thinking at all." (KANT 1922)

This triangular or 'elliptic' relation between subject, object and reflection is the fundament of human conscious functions.

Neuropsychological characters and the duality of consciousness

The critical feature of the phenomenon of consciousness is "self-sentience", sometimes also called "the primarily psychic". Self-sentience is lacking any sensory quality; it is not experienced as representation of a concrete object and therefore does not possess the type of information as present in consciously perceived contents of outer or inner reality: there doesn't exist any expertise of that kind about self-sentience and equally not about respective cognitive products.

Regarding the compact abundance of its contents, the dynamics of consciousness become reduced in states of rest and high concentration. However, there doesn't exist consciousness without any contents, if not as a symptom of psychic disease. The criterion 'conscious' doesn't mean a quality of defined contents, caused by a proper neural pattern of activity, as it is sometimes assumed by reductionists. Rather it offers the transcendental situation of the perceiving subject as experiencing himself as the ego. Thus it appears to be more correct to speak of "representations in consciousness" instead of "conscious representation". The psychogenetic hypothesis of a "primary psychic apperception" consisting in vague sensation of the outer world and indistinct self-sentience therefore means

a retrograde projection of high probability; it can be recognized already in most primitive organisms in self protecting behavior as "instinctive care of ones self". During evolution towards humans this vague self-sentience in co-evolution with the brain by means of concrete conscious experience has been implemented to the conscious self-concept and and since about 200 000 years a neurobiologically important element for the strategies of human self-representation.

In the latter phase, contrary to the previous trend of evolution, a transition from hierarchical organization of neural networks to combinatorical networks with self-referential metafunctions occurred. This can be considered as one of the cornerstones of cultural evolution.

The "psycho–physical problem", i.e., the question regarding nature and connections of both parts of consciousness, distinguished as sentience and imagination, occupies occidental thinking since more than 2500 years. Due to the progress in brain research now also natural sciences are increasingly interested in the "body–mind problem". Interdisciplinary research in cognition unequivocally proved that consciousness—understood as inner and intersubjective experience and control of behavior—is connected with cerebral processes. Because these exclusively are defined as biological physico–chemical processes consciousness cannot arise from or occur assisted by processes of different essence. This conclusion is basic to the position of "psycho–physical parallelism" which interprets the psychic states and actions of behavior as special systemic processes occurring parallel to neuronal processes.

A modified version of this position appears in the monistic identity theory, considering both mentioned parts of consciousness as two aspects of one and the same brain.

In contrast the hypothesis of substantial "dualism" affirms that the immaterial soul acts upon the brain and by that on the behavior. This, however, would demand the unacceptable presupposition that in this case the laws of physics could not be valid and other influences would become effective. Also the attempt to explain the duality of consciousness by a dualism of neuronal information processing, i.e., by the supposition that self-sentience is founded in not yet known, specific neural patterns of activity would not clarify the impenetrable nature of conscious experience: The problem only would be shifted to the level of a physicalistic reduction where no sound explanation and definition can be found.

Finally the experience of time and temporal events has to be emphasized. In consciousness we perceive ourselves as immediately participating or engaged in an uninterrupted stream of events leading from the past into the future, containing remembered as well as expected contents.

What is experienced as presence is not an imaginary borderline between the gone and the coming, but a distinct entity of the conscious process between the elapsed past and the not yet arrived future of a duration of about 4 sec. This time period appears necessary for generating the neuronal correlates of the actual contents of experience. E.g., of the sensory input, events, thought processes, impulses of action etc. and likewise for fixation and evaluation of the conscious contents regarding self-sentience. In consciousness we immediately experience our own relation to the world as the frame of relevant behavior organized by higher brain functions. The basic conscious brain processes, however, cannot be clarified by our observation: there is not any insight into the conscious process. Yet in the dynamics of relevant conscious contents a comprehensive selection of disparate results of neuronal processing mechanisms is offered. This also has been discussed as the 'phenomenon of binding'. Yet conscious experience not only contains actual data but also higher psychic processes, e.g., the above mentioned time-related behavior, the phenomenon of representation of the world and the ego on several levels, connected with the reflection of these relations in the self-concept. Consciousness in general can be defined as the presentation of objectivity in the medium of subjectivity: "The being–for–me of something as something."

Neurobiological correlates of cognition

Conscious phenomena, as exposed before, are objects of neuropsychology. Due to their inherent basic duality principally new aspects or attempts are scarcely available for the next time. Cerebral correlates of cognitive processes, however, are objects of neurophysiology, a big domain of striking complexity but in rapid progress. Before presenting a few actual problematic fields the instrumental complexity of the neuronal system has to be shown.

The central position of the brain regarding control of vital functions as well as of mental performances is generally accepted. Nevertheless it has to be mentioned, that the human brain compared with the brains of our phylogenetic ancestors exhibits essential innovations and structural completions, e.g., total corticalization of the input from environment

and from the body as basis of complete central conscious as well as unconscious integration of information regarding self-representation and individual behavior—also including vegetative functions and drives. In the following only main facts and essential features of the specialized organic brain structures and of the immense realm of their functions will be discussed.

Two general aspects of the complex neurofunctional central system have to be emphasized:

(a) the neurobiological substrate, the brain, is a hierarchically organized mega-system consisting of about 12 billions of nerve cells, i.e., neurons of different types arranged in numerous variant accumulations of different size: cortices, nuclei, tractus, bundles etc. Each single neuron may have more than 10,000 contacts, i.e., synapses, to others; the total number of these amounting to several 1000 billions. Due to that the number of possible different states of brain function runs to 10^{3000} what also means that an exact decoding of micromorphological brain states is beyond the borderline of intracosmic possibilities. The instrumental complexity of brain structure is reflected in the organ-specific biochemistry of various physiological trophic and regenerative functions which can be studied exactly.

(b) However, with regard to systemic functions, in contrast to other organs and equally to the computer, the brain system beyond its organic embedment into the life process does not exhibit an own concrete physical product of its systemic activity but only metaorganic results of its proper function manifested in the discussed phenomena of the consciousness and in form of human cognition.

Information and information processing

The basic level of the brain considered as a system consists of the mentioned enormous connectivity of neurons and interneuronal exchange of 'excitation', demonstrated by neurophysiological methods. The total of respective events in this level of neural interconnected actions is named *Information Processing*. The underlying structure of this function is the neural network formed of cooperating neurons in variant designs. Network activity, i.e., information, generated by figuration, dynamics and transport of excitation patterns, lacks the usual criteria of objects of natural science: organic material qualities. Information contains the wide spectrum of different physical stimuli from the outer world as well as from the body, both reaching the brain. These network processes become selectively transferred to coded patterns of electrochemical pulses. The same is true for the products of autonomous, e.g., cognitive brain activity. The specific function of the brain completely occurs in this translated form, whose processes of transformation are not completely explored. Neuronal information is totally different to the physical quality of its natural sources and lacks a definable physical nature, yet nevertheless is a constitutive physiological phenomenon. Neuronal networks are realized in the brain in countless samples differing in relation to complexity, functional architecture and diverse patterns of connectivity with other networks of cortices, grisea and nuclei. Brain morphology and physiology demonstrates a hierarchy of structural levels, with the cerebral cortex in the top position. The prominent grey formation contains half the number of brain neurons and in humans is connected with all other brain regions. However it does not exert an absolutistic functional dominance but in contrary a widely distributed democratic regime consisting of many unexplained rules. There is no central processing unit and no localized site of authority.

In spite of remarkable progress of neurosciences and of bioinformatics the chances to follow and to localize the complicated way of single signals and all the more to identify widely distributed programs of higher level-activities are mean, although today more exact instruments for registration of activation pattern in the cerebral cortex are used. Today the exact knowledge of brain activities is very limited and their complete description and exact analysis so far impossible. Thus the metaphor of C.H. SHERRINGTON of brain activity as an "enchanted loom" still appears as pertinent.

In general brain activities do not produce images but compatible abstract reconstructions of reality. Concerning the subject–object relation it must be emphasized that this pragmatic reconstruction is not an arbitrary one but is based on procedures as outcomes from millions of years of adaptation containing creative and optimizing procedures.

Cognitive brain performances

The first step of information processing is sensory perception whose adaptability does not depend on the precision of single partial senses e.g., visual or olfactory acuity but on the quality of the unity of perception (in German: Gestalt-Qualität).

In humans the object of perception, i.e., the product of neural reconstruction of reality means an

enormously complex synthesis of multisensory input enriched and instructed by an effective concert of inborn and learned strategies, memories and respective emotional and rational evaluations. It should not be neglected that the resulting human "object of perception" contains more information than that of higher mammals, monkeys included. Moreover the complex neural patterns representing external events in humans lead to an own functional and ontological status, with increasing independence from the underlying levels, e.g., reflex-like mechanisms. The new combinatoric capacities form the basis of higher cognitive processes, like comparison and classification processes as well as related abstractive abilities.

On this step of cognition the diversity and complexity of the reality is simplified by reductive abstraction yet without loosing significance and essential contents but on the contrary offering new aspects of objects. We speak of models equivalent to reality, in short of *real-models* of objects of perception and events. In the frame of our spatial–temporal orientation (not discussed here) these models are the source of the diversity of objects and their structural arrangement in our perceived world.

Moreover single elements of the these processes become part of mentally constructed worlds. These products are named *model-objects:* together with improved memory and the extrasensory faculty of time experience, which opens the dimension of future, the mentioned model-types enable humans of integrative thinking processes, i.e., pragmatic and innovative use of products of imagination, including instant planning of behavior and future activities.

With regard to the human level of cognition the fundamental role of the body has to be sketched. The *brain–body relation* in humans is the most developed. Compared to other animals the sensorimotor representation of the human body is the most perfect one. Beside this, in humans the vegetative functions, including the vital drives are represented in the cerebral cortex, in the already mentioned archaic form of a conscious emotional feeling (German: Befindlichkeit) and accidentally in the subcognitive sensation of pain.

The so-called higher brain processes therefore can be interpreted as results of the archaic body–brain–complex and of its self-referential unity. Finally the rational constitution of an individual self-concept and the reflexive representation of the world manifested the inherent duality of the human conscious experience.

Symbol competence, language capacity and human cognition

By means of decisive improvement of the ability of symbol formation and related language capacities a new level and capacity of cognition emerged. Regarding size, dimension and universal availability this symbol faculty led to enormous progress surpassing all other types of neuronal processes.

The evolution of phonetic language is based on the non-neuronal transcodification of products of neural information processing, especially of symbolic cognitive products, into acoustic signals in linguistic, syntactical order. E.g., the word "house", as a simple acoustic pattern represents an immensely rich and branched network pattern which, however, in its non-neuronal form easily can overcome the network-cascades and produce a wealth of new relations and contents. Language is the richest source of relevant empirical conscious contents whose prominent selective importance for the human cognitive evolution appears as evident. The funds of dispositions of human behavior by the language-bound faculty of symbols as well as by increase of the neuronal capacity of learning and memory has been extraordinarily enlarged; what also promoted the representation of temporal events and opened the chance to anticipate future needs. The central part of conscious utilization of symbolic products underlies rational thinking. Last but not least by means of language also the constitutive isolation of the private inner individual world has been partially opened: language makes the consciousness of the fellow transparent and communicable, forming a key event of human empathy and of the socio–cultural evolution: "At the beginning existed the word."

In relation to the development of language it has to be noticed that one of its most important precursor faculties probably has been gesticulation—what can already be demonstrated in primates. In the primate cortex also so-called mirror neurons were found which are activated not only by direct functional excitation but also in case of sensory transmission of respective events in the environment, e.g., gestures of social partners. This type of transpersonalization is indispensable for signal communication.

The structural fundaments of the unique evolutionary event of language capacity are manifest in the brain: the human cortex cerebri not only became distinctly enlarged regarding extension, area differentiation and quantity of neurons, but before

all was qualitatively furnished by new and increased connection systems as well with equally enlarged other cortical association areas in the left hemisphere and with brain regions which that far lacked cortical connections. The result was complete corticalization of the outer and inner input. Altogether these innovations form the mighty structural axis of the integration of linguistic functions into the already lateralized functional specialization of the human brain.

Language competence includes speaking: Yet the structures of articulation of the anthropoid apes lack the ability of producing the sounds of human articulation. In the course of brain–language coevolution a new direct fibre tract from the BROCA area of the motor cortex towards the augmented muscle bundles of the vocal cords of the larynx, the organ of articulation, as the instrumental precondition for production of human language and singing was established. This innovation among others was supported by translocation and transformation of the human larynx including better separation of it from the route of breathing.

Probably synchronously with the already mentioned symbol capacity a new type of neuronal information processing mechanisms developed, i.e., linear serial processing of excitation patterns.

This enabled ordered sequences of single logical steps of decision–making processes forming the proper tool and standard form of human rationality, logical thinking and communicative interpretation of reality.

In this context the metaorganic products of the cognitive processes will not be described.

The phenomena of this realm although they miss as shown some items of physical nature are based on the organic structure and intriguing functions of the brain. Therefore, due to physical impact in case of an organic injury, disorder, aging and disease the instrumental generation of them becomes disturbed, distorted or even destroyed: Agnosias, anosognosia, split-brain syndrome etc. are causal consequences of certain structural organic defects. The respective alterations pertain to distorted or defective configurations of the contents of the consciousness which—although pathological—nevertheless are products of perception of reality and do not reach the primary ego of consciousness. On the other hand organic disorders of the so-called unconscious brain regions, most of them localized in the deeper brain parts, diencephalon and brainstem, can alternate the emotional life regarding e.g., self-feeling and/or manifestations of drives.

Final remarks on consciousness

For a further rational approach to consciousness we have to start from the assumption that scientific theories are rendered possible by language and mathematics, i.e., by simplifying coding processes of reality, mediated by conscious experience. The self-sentient component of psychic experience, however, is not a sort of reality appropriate for these procedures: it is associated with objects or features of conscious contents but is not an element of this part of consciousness. Thus consciousness is something like a medium in which the products of information processing procedures are becoming evident, or something like a forum of initiation into full membership to the conscious contents. Furthermore consciousness cannot be something belonging to single defined imaginations themselves, i.e., an attribute of contents and by that a neural pattern of activity corresponding to conscious experience.

The neurophysiological findings therefore once more take into account just the tight functional relation between certain neuronal processes and conscious events with any regard to the categorical duality evident in conscious experience and the demonstrated way and form of its synthesis. In a similar way the highly speculative assumption of the ego-feeling part of consciousness as another particular type of neural information processing, e.g., based on a hypothetical super-imposition by quantum effects occurring in neuroaxonal tubuli, only shifts the problem to the level of physicalistic reduction which does not provide a convincing explanation, all the more as probably quantum processes are also inherent in information processing procedures. (DENNET 1991) Concerning the distinctive physiological basis of self—sentience (not self consciousness), whose organic basis seems to be scientifically enigmatic it seems to manifest itself in a faint emotional form and only occasionally—e.g., in pain—as a common sensation.

With that supposition questions concerning the nature of the problematic phenomenon of self-sentience in living systems is not yet answered. In "ego-feeling" too the case of the elliptic situation of a double reflection is evident: (a) the transcendental position as subject versus objects in connection with the reflection of (b) the immediate inner certitude of experiencing the subject itself. Concerning empirical consciousness this type of a apparently indeterminable relation points to GÖDEL's principle of incompleteness, especially to the impossibility to explain consciousness on the level of consciousness.

Probably the descriptive hypothesis of a quality inherent in systems of extreme complexity and autonomous self-control might prove itself as an acceptable metaphor. Thus the earlier mentioned conjecture of an "archepsychic experience" consisting of indistinct awareness and self-reference appears as a plausible retrospective projection from the self-consciousness of human reflection. If this widened concept of consciousness would prove to be correct, consciousness could be seen as an a-priori quality of life processes in general and increasing knowledge of the nature of life would lead to a better explanation of the particular dual quality of consciousness.

Author's address

Franz Seitelberger, Klinisches Institut für Neurologie, Universität Wien, Schwarzspanierstr. 97, A-1090 Wien, Austria.

References

Dennet, D. C. (1991) Consciousness explained. Little Brown: Boston.

Fichte, G. (1845/46) Thatsachen des Bewusstseyns. In: Sämtliche Werke (Edited by J. H. Fichte). De Gruyter: Berlin.

Kant, I. (1922) Kritik der reinen Vernunft. Grossherzog Wilhelm Ernst Ausgabe. Bd. 3. Insel: Leipzig.

Oeser, E./Seitelberger, F. (1995) Gehirn, Bewußtsein und Erkenntnis. 2nd edition. Wissenschaftliche Buchgesellschaft: Darmstadt.

Seitelberger, F. (1987) Wie geschieht Bewußtsein? Psychol. in Österreich 7: 6–19.

Seitelberger, F. (1993) Gehirn, Verhalten und Bewußtsein. In: Philosophia Perennis. Teil I. Peter Lang Verlag: Frankfurt/M., pp. 368–381.

Seitelberger, F. (1994) Das Bewußtsein. Freie Akademie, Schriftenreihe Bd.16: 29–47.

Wundt, W. (1873/74) Grundzüge der physiologischen Psychologie. Engelmann: Leipzig.

Gunther S. Stent

Are Animals Persons?

Abstract

The starting point is the question of personhood in human and non human beings. Referring to KANT, naturalistic and supernaturalistic phenomena attributed to persons are distinguished, and the role of supernaturalistic phenomena, especially those in relation to the intersection between biology and ethics, is discussed.

Key words

Dualism, KANT, personhood, naturalistic phenomena, supernaturalistic phenomena, ethics, animal rights.

THE QUESTION WHETHER animals are persons or things epitomizes the problem of their ethical standing, since the moral law applies only to persons but not to things (GOODMAN 1988; LOMARSKY 1987; 1992). If the correct answer to the question were: "Animals are mere things, because only human beings are persons", it would be nonsensical to speak of "Animal Rights". But this is *not* the correct answer. An animal *might* be a person and have rights, since in moral praxis (as opposed to moral theory) the category of person neither includes automatically all human beings nor excludes automatically all non-human beings. For personhood is not an objectively inherent trait that some creatures have and others lack. Instead, it is a subjectively bestowed trait with which we endow some creatures and withhold from others. Moreover, personhood is not an all–or–none trait, which we either wholly bestow on or wholly withhold from a creature. Rather, we bestow it to a greater or lesser degree on an animal or a human being, with the ethical standing of the creature rising with the degree of personhood with which we have endowed it.

The Soul

The modern view of personhood is based on DESCARTES' notion that the person comprises a mind, which is responsible for consciousness, feelings and rational thought, as well as a body, which provides for the person's physical presence and is a kind of automaton. The phenomena that we attribute to the person's body and mind can be classified according to whether they are *naturalistic* or *supernaturalistic* (STENT 1998). "Naturalistic" refers to phenomena that exist in the physical world of things and are subject to natural laws. "Supernaturalistic" refers to phenomena that exist only in the metaphysical world of thought and transcend natural laws. According to this bipartite classification, *all* of the properties we attribute to the person's body are naturalistic phenomena and amenable to study by scientific methods, as are (at least in principle) *some of* the properties we attribute to the person's mind, such as consciousness, rational thought, intentions and feelings. However, we also attribute some other properties to the mind that are supernaturalistic phenomena, which are not amenable to study by scientific methods. They include the categorical properties identified by Immanuel KANT (1949) as applicable only to persons but not to things, such as uniqueness, sacredness (or being an end in itself), and free will.

According to KANT, persons owe their ethical standing—having moral duties as well as moral rights—to their being invested with these supernaturalistic phenomena of the mind. The ensemble of KANTIAN supernaturalistic phenomena used to be called the person's "soul". Despite the haughty derision of the soul by most modern, down–to–earth, no-nonsense biologists and psychologists as an outdated, patently absurd, irrational—at best, poetic—locution, I believe that it remains a helpful, if not downright unavoidable concept for any discussion of the intersection of biology and ethics.

As KANT showed, belief in the soul, i.e., in a mind with supernaturalistic attributes, is not only not irrational but its acceptance is actually a prerequisite for a rationally coherent account of morality. He pointed out that we spend our lives in two metaphysically distinct realms of what he called the intellgible world, both of which are constructed by our mind. One of them—the amoral realm of things—is con-

structed by our mind's *theoretical reason*. In that realm, the relations between things (or objects) are governed by the causal determination of natural laws. The other realm—the moral realm of persons—is constructed by the mind's *practical reason*. In that realm, the relations between persons (or subjects) are governed by the laws of freedom self-imposed by autonomous minds.

Thus our interpretation of a being's actions can be fundamentally different, depending on whether we encounter it in an amoral context or in a moral context. In an amoral context—especially in a scientific setting—we can regard all beings as things, whose actions are attributable to inflexible natural laws. But in a moral context—especially in a social setting—we have to give special consideration to beings who qualify as persons and whose actions are attributable to minds with supernaturalistic, autonomous free will. Although our attributions of free will to the soul and of natural causal necessity to the body are obviously incoherent, as rational beings we can no more abandon the idea of moral freedom than we can abandon the idea of amoral causal necessity. Indeed, those rare individuals who *have* abandoned either of these ideas are usually diagnosed as psychopaths, and most of them live in prisons or mental institutions.

Personhood of Humans

Abortion. The controversy regarding the ethics of abortion, in particular the controversial thesis of the "right to life" of a human fetus, provides one example of our subjective endowment of some beings with personhood while withholding it from other beings. I first became aware of the critical role of the subjective assessment of personhood in the context of the abortion debate when I attended a conference on "Biology and the Future of Man," held in the *Grand Amphithéatre* of the Sorbonne. (GALPÉRINE 1976). As I soon discovered, the hidden agenda of the conference's organizers was to create a favorable intellectual climate for the legalization of abortion in France, then under consideration by the *Assemblée Nationale*. The international roster of participants included only biologists and biomedical scientists, because the organizers had evidently thought that your woolly-headed philosophers and theologians had nothing useful to contribute to the difficult moral problem under discussion.

The participants were generally agreed that a person has a right to life and that the intentional termination of that life is murder, which they, like *bienpensants* all over, thought ought to be illegal. So their avowed aim was to provide the *Assemblée Nationale* with an objective, scientific criterion for the critical stage in prenatal or perinatal human development at which personhood begins, i.e., when the biologically. immature human being acquires its soul. Prior to that stage—so their reasoning went—there could be no moral objection to abortion.

One participant, the eminent cytogeneticist, Michel LEJEUNE—discoverer of the chromosomal basis of Down's Syndrome—held that fertilization of the human egg is that critical stage, a criterion that would rule out abortion altogether. LEJEUNE, a devout Catholic and a leader of the French anti-abortion movement at that time, argued that personhood begins with the diploid zygote and formation of the unique genetic identity of the future individual. Other participants proposed various later developmental stages as critical, ranging from gastrulation of the embryo, through the onset of fetal heart beat or electrical brain activity, to parturition. Their proposals would allow abortion at ever-later stages of human development.

It was obviously futile for the conference to seek a biological criterion for definition of the critical stage in development at which the human embryo or fetus turns into a person, since the soul is a supernaturalistic concept of KANTIAN practical reason. At the earliest developmental stages—zygote, blastula, or gastrula—the embryo is certainly *alive* and contains a human genome which (as LEJEUNE emphasized) will eventually give rise to a biologically unique person. But the embryo is not yet a person. It is a mere blob of protoplasm, which, despite its covert genetic uniqueness, is devoid of any overt signs of behavioral or morphological uniqueness. Sometime later in development, the fetus does assume human. features that may make it *look* like a person. But at which developmental stage the fetus does begin to look that way is a highly subjective call. For most, though—transhistorically and transculturally—not necessarily all people, that stage will be reached no later than the time of birth.

The subjectivity of the criterion by which personhood of an immature human creature is actually assessed is evident from the records of trials in which abortionists were charged with the murder of a fetus. Very often, the prosecution showed photographs of the abortus to the jury, and—not surprisingly—the chance of a verdict of "guilty" was the greater the more advanced was its developmental stage. When asked for the grounds on which they reached their verdict, the jurors rarely mentioned any of the (ob-

jective) biological criteria put forward at the Sorbonne conference. Rather, most often they said that they found for "guilty" when the picture evoked the (subjective) impression that the abortus had a recognizably human face.

It seemed ironic that the participants of the Paris conference (not excepting even the devout LEJEUNE) addressed the ethics of abortion by resorting to purely bio-scientistic arguments, while facing the *Grand Amphithéatre*'s bigger than–life–size statue of DESCARTES. I was half-expecting the statue, like that of Don GIOVANNI's *Commendatore*, to quicken in anger and drag the discussants to Hell for not paying heed to DESCARTES' admonition that the moral law applies only to persons and not to machines—not because only persons have this or that body part that machines lack, but because they (and only they) have souls.

War. Another example of the subjective nature of our endowment of some human beings with personhood while withholding it from others is provided by the suspension of the moral law of the sacredness of the person by soldiers in brief war time encounters with an unknown, or even invisible foe. They feel duty-bound to kill the faceless enemies, who are perceived as dangerous things, or vermin, outside the bounds of morality. But if an opportunity has been afforded for the soldier to recognize his enemy as a unique individual, the foe-thing may acquire personhood and come under the protection of the moral law.

This process of repersonalization of a depersonalized enemy was rendered in a poignant scene of the 1930 Universal Pictures movie, *All Quiet on the Western Front*. In the dark of night, a German soldier tumbles into a front-line fox hole already occupied by a French soldier. The Frenchman pulls his bayonet and is about to plunge it into the German's chest when an exploding grenade lights up the scene. The *poilu* sees the terrified face of the *boche*. He recognizes that the other man, hitherto identified generically as an enemy by his military uniform and style of helmet, is a unique person and cannot bring himself to kill him. Instead, the *poilu* offers the *boche* a cigarette.

Robots. The 1973 MGM movie, Westworld, presented a further cinematic instantiation of the subjectivity of our judgment of personhood. The scene is set in a holiday resort in Arizona for (male) millionaire guests, where, in Disneyland style, an old-time, Wild West frontier town, complete with bars and bordellos, has been re-created. On arrival, the guests exchange their business suits for ranch-hand outfits and mingle with the resort's staff, who are impersonating cowboys, card sharks, bandits and whores. The guests are informed that the town's inhabitants are actually mechanical robots, manipulated electronically by human handlers in an underground control center. The robots look and behave so realistically that only the handlers, but not the guests, can tell them apart from real humans. Since the guests believe (on hearsay, without objective proof), that the people they run into at Westworld are things rather than persons, they feel free act out their repressed immoral desires. So they have an enjoyable holiday away from the office, raping, and killing, guilt-free and ad libitum.

Racism and Slavery. There is (or at least used to be) an American saying that "all Chinese waiters looks alike", which, though it may seem like a harmless opinion, is, in fact, a deeply racist revelation. Not to recognize the unique individuality of members of the foreign race makes it possible to exclude them from the category of persons protected by the moral law.

A similar, though much more egregious instance of the failure to grant full personhood, and hence protection under the moral law, to what are obviously fellow human beings is provided by slavery. Here too the denial of personhood is usually linked with a failure or refusal to perceive the uniqueness of individual slaves (who are, in any case, often members of a foreign race). In this way, the slave masters make themselves feel morally righteous in the belief that the beings they cruelly exploit are mere, commonplace, replaceable things, rather than sacred persons with unique souls who are ends in themselves.

The bestowal or withdrawal of personhood is not necessarily an all–or–none matter, however. For instance, while prior to their emancipation, the African slaves in the Southern United States had no rights and were not recognized as legal persons by the state, some slaves were invested with partial personhood by their masters. Most of these semi-persons were house slaves, whose uniqueness was appreciated by virtue of their close daily contacts with their masters' families and with whom they established even some emotional ties.

Human clones. Ever since the rise of molecular and cell biology in the 1950s and 1960s, the prospect of cloning mammals was held out as an exciting biotechnological prospect. Cloning would replace old-time, conventional sexual reproduction and its genetically uncertain outcome with a radically novel approach to procreation that guarantees the genetic identity of the progeny. This project seemed well within the range of possible experimental procedures, by extracting cell nuclei from the somatic tis-

sues of an adult male or female donor animal and transfer them into enucleated eggs provided by a female host animal. Upon embryological development of these transgenically manipulated eggs would arise a clone of offspring individuals, all of whom would be genetically identical to the donor.

The recently claimed success in cloning mammals by this procedure after, several decades of failed attempts (WILMUT et al. 1997) has given rise to great hopes for genetically engineered animal husbandry. But it has also aroused to great fears that this procedure will be applied to humans as well and that we may be at the threshold of a monstrous Brave New World populated by hordes of genetically identical people.

Why is it that whereas I would welcome having EINSTEIN, BEETHOVEN, Marilyn MONROE and Clark GABLE reside in my neighborhood and enjoy running into them at the Supermarket or at the local bar, the idea of living in a town with a hundred of their cloned replicas seems obscene? What is the source of this revulsion? It is that these hundreds of replica EINSTEINS, BEETHOVENS, MONROES and GABLES would not *really be* Albert, Ludwig, Marilyn and Clark. They would merely be CARTESIAN automata in human form, because they could not, each of them, be, endowed with a unique soul. I would doubt the humanity of these *Doppelgänger*, because I could not tell them apart and could not conceive of so many different, unique minds all inhabiting look-alike bodies with identical faces.

Personhood of Animals

Prior to the end of the 18th century, philosophical discussions regarding personhood did not concern animals, except as mindless counterexamples to thinking human beings. DESCARTES, like his philosophical predecessors and contemporaries, still considered animals as unthinking automata whose control of behavior is wholly accounted for by bodily functions. The possibility that animals have minds was not taken seriously until the latter part of the 19th century, upon the acceptance of the idea of the evolutionary continuity between animals and humans and the subsequent rise of the discipline that came to be known as "ethology". And yet, even if animals *do* have minds capable generating consciousness, rational thoughts, intentions and feelings, it does not necessarily follow that their minds include also the supernaturalistic properties such as uniqueness, sacredness, and free will that we attribute to the soul. That is, even if animals have minds, they might still fail to qualify for personhood and lack the ethical standing claimed on their behalf by the advocates of "Animal Rights" (RADNER/RADNER 1989). But just as we subjectively depersonalize some humans and deprive them of their moral status on the basis of their perceived lack of a unique soul, so we subjectively personalize some animals, endow them with a unique soul, and confer moral status on them.

Household Pets. Relatively few kinds of animals are likely candidates for elevation to honorary personhood. Most kinds of animals belong to invertebrate phyla, such as the annelids, mollusks and insects, about which it is hard to believe that they have a mind that generates consciousness and rational thought. And even among the vertebrates there is only one class—the mammals—that includes some species whose behavior is sufficiently complex that attributing to them a conscious and rational mind might cross the threshold of credibility. Yet even among those candidate species few specimens attain personhood status, because of their lack of opportunity to be recognized as, a unique individuals by a human being and be granted moral standing by virtue of the authority vested in us to confer that honor on them.

The most common examples of personalized animals are provided by household pets, such as cats and dogs. They are in a particularly favorable position for elevation to personhood, not least because they were bred from their wild ancestor species for the very purpose of providing companionship for their human masters. Their masters generally acknowledge the animals' individuality, perceive them as unique specimens of their kind, and endow them with a mind with supernaturalistic properties, such as being ends in themselves and having free will. The protection of the moral law is extended to the pet, which is given a personal name and treated with the same consideration as are the human members of the household. Moreover, the concept of moral responsibility tends to be extended to the pet as well. The master may scold, or even punish the pet for soiling the carpet or killing a song-bird in the garden, while also praising, or even rewarding it for killing household vermin.

Animal Experimentation

The personhood of animals is at stake too in the ethical struggle about the permissibility of animal experiments. The parties to this dispute generally try to disqualify their opponents by claiming that they are morally unworthy discussion partners. On the one hand, the Animal-Rights opponents of animal exper-

imentation usually attribute selfish motives to their adversaries, even claiming that biomedical scientists do their experiments mainly to satisfy their sadistic yen for torturing animals. Accordingly, they allege that one cannot expect that just those researchers who are most qualified to evaluate the scientific or medical need for animal experimentation would actually meet their ethical responsibility vis a vis the creatures they victimize. On the other hand, the biomedical scientists under attack usually claim that they have a highly ethical attitude towards the protection of life in general, show great concern for the welfare of animals in particular, including those that they use in their experiments, and that the principal professional goal of their whole research effort is precisely to avert pain and suffering of humans as well as animals. According to them, the Animal Rights movement consists mainly of sentimental fools and hypocrites, who, though they have managed to gain a disproportionately powerful political influence, cannot be engaged in any rational discourse.

As a practicing neurobiologist, I recognize the unavoidable necessity of the conduct of animal experiments, without which it is hard to imagine progress in biomedical research. Yet, I was dismayed to note that the arguments put forward in defense of animal experiments by my colleagues were easily defeated by representatives of the Animal Rights movement skilled in ethical disputation. For instance, many biomedical scientists concede that, in view of their professing a moral duty to protect animals (in fact, life in general), they see a moral problem in our killing of animals to provide us with food and clothing, destroying them as pests, and vermin, and condemning them to extinction in their natural state by ever expanding our living space. Nonetheless, they claim that placing humans and animals on a moral par is a chimera that puts into question the human primacy embedded in Occidental moral concepts, and thus of human existence on the whole. For, as we are told in Genesis 1:26. "And God said, let us make mankind in our image, after our likeness. And let them have dominion over the fish in the sea, and over the fowl of the air, and over the cattle, and over all the earth, and over every creeping thing that creeps upon the earth." Thus, within the context of Biblical tradition, animals are not ends in themselves, and their use for human utility presents no moral problem.

In response to the scientists' claim that placing humans and animals on a par is a chimera, the adherents the Animal Rights movement point out that it was not all that long ago in the United States that putting black Africans and women on a par with white males was also considered a chimera. Admittedly, this historical comparison does not provide a logically compelling argument, yet the claim that the recognition of Animal Rights is a goal representing moral progress over Biblical tradition cannot be dismissed all that easily. After all, there do exist Oriental cultures whose moral traditions are based on the atheistic ethics of Buddhism, according to which animals and humans are considered on a par to a much greater extent than they are in the West. Since Buddhists don't believe in God in the first place, they can hardly hold that He created mankind, but not the animals, in His own image. All the same, putting mankind and animals on a moral par is a belief whose general acceptance would bring about a radical transformation of the Occidental life style that even the adherents of the. Animal Rights doctrine would be unlikely to welcome.

A comparison of Sweden and the United States is instructive in this connection. On the one hand, the regulations governing animal experimentation are much more stringent in Sweden than in the United States. As far as I know, neurophysiological experiments with higher mammals, not to mention primates, are almost impossible to carry out in Sweden. On the other hand, the regulations governing human experimentation are much more liberal in Sweden than in the United States. Thus the Swedish legislature is apparently much closer to the chimera of the moral parity of humans and animals than is the American.

Coda

I give it as my view that a creature—human or animal—has full ethical standing, i.e., has moral duties as well as moral rights, if and only if we attribute to it a mind with the supernaturalistic properties identified by KANT as applicable to persons but not to things. These properties are uniqueness, being an end in itself, and free will, which belong to what used to be called "the soul". This statement is not meant to be normative, but intended merely to call attention to the subjective operational criteria by which the judgment of personhood is actually made, even by those people who may profess other criteria on religious, biological or humanistic grounds.

Author's address

Gunther S. Stent, Department of Molecular and Cell Biology, University of California, Berkeley, CA 94708-3200, USA.

References

Galpérine, C. (1976) Biology and the Future of Man. Universities of Paris: Paris.

Goodman, M.F. (ed) (1988) What is a Person? Humana Press: Clifton, NJ.

Kant, I. (1949) Critique of Practical Reason. (Translated by L. W. Beck.) University of Chicago Press: Chicago.

Lomarsky, L. E. (1987) Persons, Rights, and the Moral Community. Oxford University Press: Oxford.

Lomarsky, L. E. (1992) Concept of Person. In: Becker, L.C./ Becker, C. B. (eds) Encyclopedia of Ethics. Vol. II. Garland Publishing: New York, pp. 950–56.

Radner, D./Radner, M. (1989) Animal Consciousness. Prometheus Books: New York.

Stent, G. S. (1998) Epistemic Dualism of Mind and Body. Proc Am. Phil. Soc. 142 (4): 578–588.

Wilmut, I./Schnieke, A. E./McWhir, I./Kind, A. J./Campbell, K. H. S. (1997) Viable offspring derived from fetal and adult mammalian cells. Nature 385: 810–813.

Holk Cruse

Feeling Our Body—The Basis of Cognition?

Abstract

The paper introduces the distinction between HIP systems, i.e., systems with first-person perspective, and NIP systems, i.e., systems not having an internal perspective. This leads to the question as to what may be the conditions which are responsible to make up a HIP system.

First it is argued that the traditional distinction between the two epistemological domains, the objective and the subjective world, should be replaced by the view that the so-called objective world is not a distinct domain, but a subset of the subjective world. It is further argued that the critical condition to become a HIP system is to have an internal world model the basis of which is a model of the own body. A neural network is explained which can be used as a manipulable body model. The output of this model is compared with psychological findings discussed to be important descriptions of the internal perspective as are perspectivalness and transparency, and with recent observations of patients suffering from brain lesions. This supports the hypothesis that the content of the world model which includes the model of ourselves is what we experience as a first-person perspective. In other words, we do not have the experience, we are the experience. This hypothesis agrees with the idea of METZINGER (2000) who states that humans "are systems, which were configured by evolution in such a way that they constantly confuse themselves with the content of their phenomenal self-model".

Key words

First-person perspective, third-person perspective, world model, body model, perspectivalness.

Introduction

To be able to feel something implies that the system which feels has an internal aspect or a first-person perspective. Biologists, of course, do not deny the existence of an internal aspect, but are trained not to consider the phenomena of the first-person perspective in their scientific work, because they do not belong to the objective world and are therefore assumed not to be tractable by scientific methods. Therefore neurobiologists when looking at brain functions keep away from this suspicious field. Philosophers, on the other hand, take this phenomenon as a given basis and investigate the rich complexity of phenomena they find within this world, but are usually not interested in any possible physiological grounding of these phenomena. As in this paper my goal is to make an attempt to contribute to close the gap between these two worlds, the paper bears the danger of being of interest neither for the neurobiologist nor for the philosopher. This even more, as, when starting to work on this problem, it is indispensable to dissect it into smaller parts. So my goal is not try to understand the first-person perspective in its full richness, i.e., including all aspects of consciousness and self-consciousness (see Introductory Chapter), but merely to obtain an understanding of the first step of the transition from a pure, though possibly complex, reflex machine-like system to a system which can have a subjective experience, even if this experience occurs in only a prototypical form.

What is meant by cognition? Cognition is not an easy term to be used, because many different definitions can be found. Nevertheless, there appears to be some agreement such that cognition has to do with feeling, thinking, problem solving, usually in connection with conscious perception and awareness. There are, however, also strict operational definitions which do not rely on awareness. MCFARLAND/BÖSSER (1993), for example, define cognition as manipulation of declarative knowledge. In order to avoid the error to restrict our view on a possibly too narrow range, I would like to leave this problem in the following as open as possible. Thus, the term cognition is only used here in an intuitive and loosely defined way.

One can distinguish between different systems which are able to control behavior. For example, cognitive systems are able to control, in particular, intelligent behavior. There are, however, also two other types of systems under discussion. These are the reactive, or data-driven systems on the one hand

and symbol using systems as developed in traditional artificial intelligence (AI) on the other hand. In the following I shall briefly describe what is meant by reactive and by symbol using systems. It should, however, be noted that the application of these two terms does not necessarily imply that there is always a clear distinction to be made between both types of systems.

Nowadays robots, and possibly also simple animals are typical reactive systems, sometimes also dubbed reflex machines. The BRAITENBERG vehicles (BRAITENBERG 1984) are often mentioned examples which, in spite of the extremely simple internal wiring connecting their sensor and motor units, can show quite complex behavior. However, in both, the field of autonomous robots, and the field of neuroethology, more and more complicated "animats" or animals of this type are investigated. These reactive systems show a rich behavioral spectrum ranging from simple reflexes to complex navigation and opportunistic behavior when following different goals. Such systems can even show displacement behavior and can solve complex motor control tasks in underconstrained situations, i.e., in situations where decisions are required (e.g., MAES 1991). Could such systems be called intelligent? It has often been said that these systems are attributed with intelligence in the adverbial sense but not in the nominal sense (LANZ 2000). They may behave intelligently, or behave as if they were intelligent, but they have no "real", nominal intelligence in the sense of creating new ideas, i.e., showing new solutions to given problems.

The second type mentioned is the symbol using system such as expert systems or general problem solvers developed in artificial intelligence. In contrast to the "analog" reactive systems, traditional AI systems cannot easily deal with blurred information. In these systems input signals are first transformed into symbols and these symbols are then treated further. Given the symbolic information, known rational methods can be applied such as the rules of logic or specific mathematical tools. The general goal of AI was to develop systems of nominal intelligence. However, although being quite successful in many cases, AI systems were found not to be able to solve many "simple", daily-life problems which can already be solved by young children or by many animals. These AI systems are, the reader may permit this biological view, like parasites. They "feed" on symbols provided by humans, but they cannot "survive" without symbols, i.e., in a natural world. They are specialists, apparently not equipped with a general intelligence.

Two points are usually mentioned when looking for reasons why these AI systems lack nominal intelligence, too. One is that, in contrast to the reactive systems mentioned, these systems are not "grounded" or "situated" (e.g., PFEIFER 1996), they have no own body and therefore cannot react to the world. This has led to postulate that, in order to gain intelligent systems, they have to be built as real robots which have not only sensory but also motor contact with the physical world (BROOKS 1991).

First-person perspective and third-person perspective

Another argument is that these AI systems, as the reactive systems, are no cognitive systems because they do not have awareness. They lack the capability of having subjective experience. As the property of a system to have feelings seems to be one prerequisite for the system to be a cognitive one, this should be treated further. What does it mean to have subjective experiences?

Apparently, the world can be divided into two types of systems. There are systems, as definitely humans, and most probably many other animals, which can have awareness, or subjective experience or internal perspective. For short they might be called HIP systems (Having an Internal Perspective). There are other systems, as most probably a pencil, a stone, or a clockwork, which do not have this ability. I will name them NIP systems (Not having an Internal Perspective). They do not feel. Of course, we cannot be definitely sure about the latter in any case, but usually people agree on that and, to simplify the following discussion, I want to take this as given. In many cases, for example for an insect, we cannot decide whether it belongs to the NIP systems acting like a reflex machine, as a clockwork, or whether it belongs to the second type, the HIP systems, and has feelings. Such a decision is the more difficult, the more the brain under view differs from a human brain.

However, one part of the problem might be solved in the following way. Philosophers usually distinguish between two epistemological domains, which they describe as the external perspective and the internal perspective, or the third-person perspective and the first-person perspective, respectively. The third-person perspective concerns our daily intellectual activities dealing with "objective" facts. They are called objective because, in principle, every human has the possibility of mentally treating these objects which means that information about

these "facts" can be transmitted between persons. Examples are the number of apples lying in a dish, or the name of the color of a green pencil, or entities used when performing a mathematical calculation such as adding two numbers. These activities are objective or public because they can be looked at by any outside observer. This "external perspective" not only concerns the behavior of the complete system, but, in principle, also includes the possibility of looking inside the brain of a subject and measuring all the interesting neuronal activities, for example, when the subject judges the color of the pencil. In this way, all details of the subject seeing "green" could be determined. (Thus, looking inside in this way still belongs to the external perspective). In general, this view could also be applied to a correspondingly programmed robot. All these data form the basis of the external perspective, the view of the third person or outside observer. On the other hand, the subject seeing green does not see these neuronal activities, but *experiences* seeing green. This is called the subjective quality of experience. In contrast to the items of the external perspective, the content of this subjective or first-person perspective is only accessible to the person himself or herself. Therefore, this is also called the private view in contrast to the public view. Nobody other than myself can judge how I see green. The difference between the external perspective and the internal perspective may become even clearer in the example of pain. Again we can consider all neuronal activities that occur when the subject's skin is stimulated, for example by a needle. One might, in principle, even look at one's own action potentials, if oneself is the subject of this experiment. But the experience when considering all these neuronal activities is completely different from the pain one experiences at this moment. Thus, self observation tells us that there exist systems, namely humans, that can experience an internal perspective. As has already been mentioned above, intuition tells us that there are other systems, like a stone or a simple machine (including some clever nowadays robots) that may not have such an internal perspective.

Interestingly, there are at least three points where one can find a transition from a system without first-person perspective, a NIP system, to a HIP system, i.e., a system having first-person perspective. At least once during the evolution, anywhere between simple algae and human beings, this change must have happened. Such a transition also occurs during the ontogeny of each individual human. A, say, four-cell human embryo most probably does not yet have the ability to experience. So, somewhere during ontogeny this change has to occur. Furthermore, our everyday change from (dreamless) sleep to the awake state shows such a transition. This means that even our brain can, to some extent deliberately, switch between these states. The assumption that there are two types of systems, some with and others without a first-person perspective leads to the consideration that having the first-person perspective is a system property—sometimes called an emergent property—that suddenly occurs if some specific conditions are fulfilled. It would be of interest to know what are these conditions?

Before I treat this question, I will expand on the distinction between the two domains determined by the external and the internal view. I would like to argue that this separation is appropriate to describe the historical development and that it is also appropriate as a didactic method to introduce into the problems, but that it is actually misleading when one tries to understand the problems raised here. This separation was probably introduced by PLATO and worked out most stringently by DESCARTES. It can still be found today in most texts referring to these problems. Even in texts tending to oppose DESCARTES' view, this separation still exists in a more or less hidden form (e.g., DAMASIO 1994). Unhappy with DESCARTES' dualistic approach, philosophers have introduced the notion of ontological identity explained with the metaphor of "these are two sides of a coin". This metaphor is definitely helpful in physics in trying to combine two items belonging to the "objective world", for example the wave–corpuscular dichotomy. However, the distinction between the first-person perspective and the third-person perspective is not cogently described by this metaphor. Looking at the coin requires the capability to look from the outside. This is just what we cannot do from the internal viewpoint. We do not have the experience, we are the experience. I thus propose the following, alternative view which might be called epistemological solipsism. In this view, it is impossible to draw a clear line separating the two worlds in such a way that an item falls into either one or the other domain. Rather, everything of our so-called objective or third-person perspective is embedded in our subjective experiences. Basically, we have subjective experiences, and those we call objective are subjective experiences that can be more easily separated from each other, therefore being measurable in many cases and more easily transmittable between different persons. Most often, but not necessarily, these entities are denoted

by symbols like words. Nevertheless, non of these experiences are non-subjective. We do not have access to the "real objective world" (provided such a world exists). Even an "objective" notion like "color is an electromagnetic wave" belongs to the subjective domain, the first-person perspective, because we do not know what an electromagnetic wave really is. To say this by another, similar example: We are usually trained to formulate a sentence like: "light of the wavelength of 450 nm appears to be blue". This formulation implies the first part of the sentence, namely "light of the wavelength of 450 mn" to represent the hard fact, the second part to represent a much weaker, subjective aspect. However, the statement could also be formulated in the opposite way: "My experience of 'blue' can, in the framework of physics, be described by using the metaphor of a wave with a given length". Clearly the first part, the subjective experience, can be experienced directly. But, as mentioned, also the content of the second part is represented in the domain of subjective experience. At the same time, however, it belongs to a mental system consisting of declarative entities or symbols which we have developed during our cultural evolution. Therefore, instead of speaking of two ontologically separate domains, a better metaphor would be to consider the set of "objective" phenomena as a part, a subset of the subjective world we experience. This epistemological solipsism means that so-called "objective" phenomena are subjective phenomena with specific properties that do not belong to all subjective phenomena: objective phenomena are easily transmittable between persons, because they can be connected to more or less clearly defined symbols. Phenomena lacking these properties are those commonly called subjective. In other words, what we call "objective" might be called "intersubjective" instead. This is illustrated in Fig. 1. The large ellipses represent the subjective world of three persons. The broken lines mark subsets of these worlds containing such entities which can be easily separated from each other and can therefore easily be described and communicated between persons (this is illustrated by the double arrows). In this way, these entities appear to be elements of an objective world. The difference between objective and subjective items might correspond to the observation that verbal expressions and memorized items allow a coarser distinction than our attentional contents do. To give an example: we are able to perceive more subtle differences in color shadings than we are able to memorize or to describe verbally.

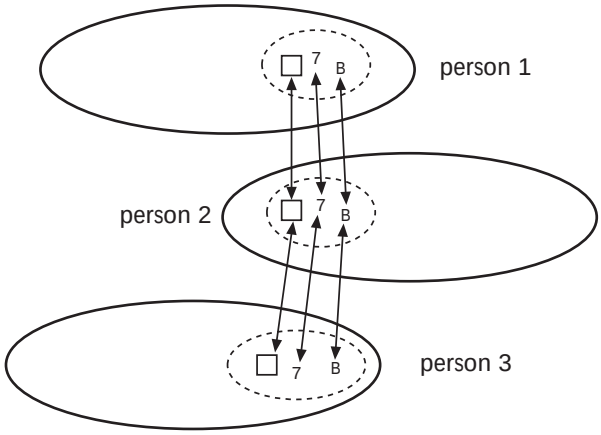

Figure 1: The large ellipses represent the subjective world of three persons. Their objective worlds form subsets of these worlds (broken lines). These objective worlds contain such entities (e.g., a square, a number, a letter) which can be easily communicated.

Simple neural network models

Let us now come back to the basic questions of how it is possible that some systems have the first-person perspective which CHALMERS (1995) has termed the really hard problem. What are the conditions allowing humans and, as I believe, "higher" animals, to experience an internal perspective ("have a soul") and thus being able to become cognitive systems whereas other systems, like simple animals, dead objects (e.g., a ball rolling in a bowl towards the lowest point, i.e., behaving as if it had a "goal"), or robots, do not experience the first-person perspective. I assume that this property depends exclusively on the architecture of the neuronal system, in particular on a special kind of world model constructed within the neuronal system. To explain this, I would like to introduce three types of systems (Fig. 2), which can be distinguished according to their internal architecture (the concepts used in the field of Artificial Neural Networks will be used to describe this architecture). The first type contains a feed-forward system, the two other types contain recurrent networks and will be explained below.

A typical feed-forward network consists of an input (sensory) layer, one or more hidden layers and an output (motor) layer. Such a reactive system is symbolized in the lower section of Fig. 2. Placed in an environment and having the appropriate internal connections, such a system can exhibit quite sophisticated behaviors (see the BRAITENBERG vehicles and their more recent and advanced relatives). The internal connections of this network may be interpreted

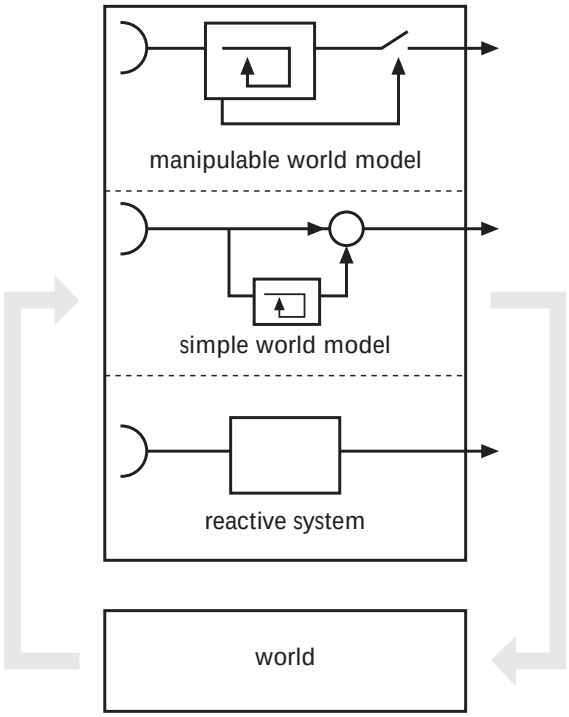

Figure 2: A three layered system acting onto and being influenced by the world. The lower part consists of a pure reactive system. The section in the middle shows a system with a simple, fixed world model as, e.g., a circadian clock. The upper part denotes a system which contains a "manipulable" world model, i.e., a model which can be uncoupled from the motor output.

as comprising an implicit world model (of that part of the world that is of importance for the system), because in some more or less indirect way it represents the properties of the world and the appropriate reactions. This corresponds to the internal representations mentioned by J. PROUST (1999). However, this "world model" is static. The reactive or data-driven mode of this system does not allow for explicit predictions in time. To make a prediction possible, additional networks with internal recurrent connections are necessary. Such recurrent networks allow for time-dependent states that are no longer directly dependent on the actual input. In this way, they can comprise world models that can be used by the system to predict sensory input before the corresponding change in the environment actually occurs and to enable the system to prepare the necessary actions accordingly. A simple example is a central pattern generator used to control time-dependent behavior (e.g., our circadian clock). This not only concerns the control of the motor output, but also the time-dependent control of sensory input. Usually these "world models" are confined to a given behavioral context elicited by a definite complex of sensory (or internal, e.g., determined by the hormonal state) stimuli, and, in this respect, are also, however, more indirectly, data-driven. These two types of systems will be described as containing *non-manipulable* models, in contrast to a third type of world model to be explained below. Both types of systems described up to now are intuitively assumed to have no first-person perspective.

As a further step, one could now imagine expanding the system such that another type of internal world model exists that can be to a large degree uncoupled from the actual sensory input and the motor output (Fig. 2, upper part). For example, this is the case when a movement is planned in order to decide whether this movement, when executed, would lead to the desired result. This movement may, for instance, concern the grasping of an apple lying on the table possibly within reach, or, to take a more abstract example, the movement of a figure in a chess game. Another example is given by the task to mentally rotate a visually given object (SHEPARD/METZLER 1971). Such an internal world model may be called *manipulable*, because after the task is given as input, it can be further uncoupled from actual sensory input and can "play around in its virtual internal world". The results this model proposes after finishing its search can be judged by given evaluation criteria. Depending on these values, a decision might be made whether or not to actually perform the proposed activity.

As long as this world model represents only information concerning the outer world and not properties of the system itself, one still might not be prepared to attribute a first-person perspective to it. (Traditional, non-situated AI systems may be considered examples of such cases). This situation might change if the system under view, first, has not only sensory but also motor contact with the physical world and, second, some of the physical properties of the system itself are also embedded into the internal world model. To explain this, I would like to restrict the following discussion to the experiences of the subject's own body including planning and performing movements, for example reaching movements with the arm and the hand. I do this because I assume that the body model that animals including humans use when planning or performing a movement may be the very basis of the phylogenetic evolution and possibly also the ontogenetic development of a manipulable internal model. This view has been proposed in a similar way by CONDILLAC (1798). If we had only the possibility to receive acoustic or chemical stimuli, according to CONDIL-

LAC, we had not the possibility to distinguish between the self and the external world. For tactile stimulation the situation is quite different. During movements of the own body self contact may occur which leads to simultaneous tactile stimulation of different body sites. As CONDILLAC states this makes a fundamental difference between the tactile sense and the other senses. Such correlated simultaneous stimulations within one sensory modality are only possible in the tactile sense. The information from the mechano–sensory input, including the position senses, can be used to form an internal representation of the own body allowing the distinction between the own body and the external world (active body moments may also play a role). This body model may then be expanded to include objects of the world outside the body. For example, if we close the eyes and hold a stick in the hand to grope the environment, we seem to feel the tip of the stick. Therefore, this stick seems to be a part of our body. Later on, other objects like a cup standing on the table but not (yet) connected to the body might be included into the world model.

The body model

How should a system having a manipulable model of its own body be constructed? To explain the hypothesis more concretely, in the following an extremely basic model will be described. STEINKÜHLER et al. (1995) proposed a model which can deal with the problem of inverse kinematics, i.e. that can determine a geometrical arrangement of a multilimbed arm to grasp a given object. This network, called MMC net, provides a holistic internal model of the kinematics of the body. Although the model can be extended to describe the geometry of a complete body (CRUSE et al. 1996), I will concentrate here on the example of moving a three-joint planar arm (Fig. 3). The control of such an arm already involves the general problem of controlling a redundant system, because the end point of the planar arm is determined by two coordinates, x and y, whereas more than two independent values, namely the three joint angles, α, β, and γ, have to be controlled (see inset Fig. 3). This "redundant" system thus has one extra degree of freedom which means that the model has to decide between different possible solutions. The necessity to cope with underconstrained situations is an important problem. In the visual domain, for example, this problem has to be solved when we interpret the two-dimensional line drawing of a Necker cube as a three-dimensional object. Here we deal with a corresponding problem in the motor domain. The model consists of a recurrent network that relaxes to adopt a stable state corresponding to a geometrically correct solution, even when the input does not fully constrain the solution[1]. This network, called body model in the following, can be used as a model for the inverse kinematics by which for any visually given target, for example, an appropriate arm configuration can be found.

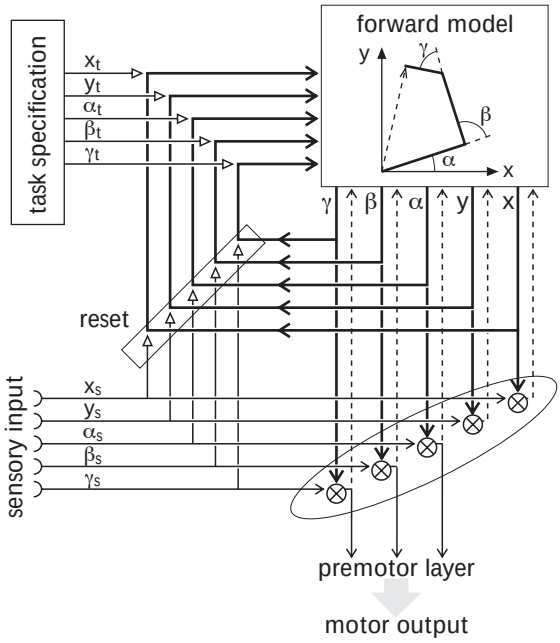

Figure 3: A schematic drawing showing the body model of a three joint arm (forward model plus bold recurrent connections). The arm is depicted in the inset, the shoulder forming the origin of the coordinate system x/y. The output of this recurrent network is assumed to correspond to the phenomenon of subjective experience. The sensory input (joint angles at shoulder, elbow and wrist, α_s, β_s, γ_s, and the end point position x_s and y_s) is compared to the output of the body model (subtraction symbols marked by the ellipse).

The MMC model will be shown in a simplified form (for a detailed description of the quantitative model see STEINKÜHLER/CRUSE 1998). This presentation permits to an explanation of the body model as a combination of a recurrent system (Fig. 3, bold lines) and a simple nonlinear multi-layered feed-forward system (Fig. 3, forward model). The forward model uses as internal variables the three angle values α, β, and γ of shoulder, elbow, and wrist as well as the x/y position of the arm end point (see inset in Fig. 3). These five values are then fed back to the input to form a recurrent system (Fig. 3, bold lines). Three sensory variables are called α_s, β_s, and γ_s, describing the actual joint angles. This input layer may

be regarded as to form a mechano-sensory layer. Furthermore, the position of the end point of the arm is given by CARTESIAN coordinates x_s,y_s provided by the visual system. These sensory values can be used to reset the recurrent connections of the body model to adopt the values of the real world. This is indicated by the open arrowheads which, when activated, suppress the recurrent signal (Fig. 3, reset). The output of the body model is compared with the sensory input (marked by the ellipse in Fig. 3) to provide the motor output.

How does the system work? When all internal variables, i.e., the angle values and the end point position coordinates, match, i.e., form a geometrically possible configuration, the recurrent network is in a relaxed state. Its "harmony value" (RUMELHART/MCCLELLAND 1986) is high. Now assume that a task is specified to move the arm to a new position x_t /y_t. This corresponds to an input from the box "task specification" in Fig. 3 which leads to a suppression (open arrowheads) of the recurrent x/y values. Now the harmony of the system is low because the actual angle values and the new x/y values do not match. However, the network relaxes to a new state with angle values fulfilling the external constraints. After having reached high harmony again, the output could be used such that the premotor values are given to the motor output to control the actual movement. By disconnecting the premotor layer from the motor output, the body model can, however, also be used to "plan" the movement. It could, in other words, be considered as the neural basis for the production of an imagined movement (which has been called Bewegungsphantasie by GEHLEN 1971).

This system can be considered to provide a solution for the problem of sensor fusion, occurring when different sensory input represent the same value. As an example, in the case of the three-joint arm, the position of the end-effector could be given by the visual system (the coordinates x_s and y_s), but also by the mechanoreceptive input (the angles α, β, and γ). These different values are integrated within this holistic system. The network concatenates both the visual and mechanoreceptively given spaces to form a common, dynamic representation. Within the recurrent model, a holistic system, no distinction between the sensory values and the motor values can be made (CRUSE et al. 1998). Interestingly, it has been reported by RIZZOLATTI/ARBIB (1998) that premotor areas are activated whether an individual is about to perform an action or observes another individual performing an action. Furthermore, in the case of mere observation of this action, a strong inhibitory influence has been found which selectively inhibits those motoneurons in the spinal chord which were involved in the observed action execution.

An open question is how the body model can adapt to changes of body geometry, e.g., during growth, or to sensory drift. At first sight, this may appear to be a mere technical problem, but the solution is of interest here. This could be done in the following way. When, in a relaxed, static state, the difference between the output of the body model, e.g., joint angles, and the actual sensory input is zero, the body model is appropriate (the symbols representing the subtractions are marked by the ellipse in Fig. 3). In this case no changes of the internal connections of the forward model, its synaptic weights, are necessary. The greater the differences, the greater the necessary changes of the weights. Using these differences as error signals, the body model could easily be updated using the classical backpropagation algorithm in accordance to a proposal by KAWATO/GOMI (1992). In Fig. 3 these error signals are indicated by dashed lines. Apart from enabling us to learn the properties of the body model, this comparison between the output of the body model and the actual sensory input might also have another effect. When the manipulable model proposes a new arm position, the latter is usually different from the actual one represented by sensory input. Functionally, this difference may be used to control a corresponding movement. Furthermore, as explained below, it may be used to direct attention.

The appearance of the first-person perspective

When applying the third-person perspective, this may be a sufficient description of the system. At this point, however, let us also consider the first-person perspective. When planning a movement we can experience a body image, we "see"—or can imagine—the position of our virtual arm moving around. Why does this experience occur? Seen by an outside observer, the situation can be described such that two neuronal layers exist that both represent the subject's own body. One is represented by the current activities of the actual sensory layers, the other by the output of the forward model. The first may be used to specify a particular behavior by direct connections to the motor output not shown in Fig. 3. Furthermore, they may be used to reset the body model and to update its weights. The second layer

may be used to plan and/or to control motor behavior. Following METZINGER (2000), I speculate that it is the output of this body model which yields the condition for the system to experience a first-person perspective, i.e., to make up a HIP system (see also CRUSE 1979; for an excellent and very elaborated description of this view, see METZINGER 1993). METZINGER (2000) claims "that—ontologically speaking—no such things as selves exist in the world. What actually exists is a special kind of self-models and their contents, and this content makes us believe that we actually do have or are identical to a self." In other words, the hypothesis is that it is the occurrence of this body model that makes the system feel. Is there a possibility to test this hypothesis? The answer is that there is probably no direct possibility to test it. The only immediate way to prove the feasibility of the hypothesis is to collect phenomenological observations on the first-person perspective and see whether corresponding properties of the model can be found. Several such observations will be explained. As mentioned, according to our hypothesis only the output of the body model can be recognized, but not the sensory representations as such. The latter are not cognitively available. This agrees with the introspective observation that the conscious I does not have the impression of viewing a movie when looking at the sensory input (like a homunculus would), but to be in the world. We "look through the screen". This is called semantic transparency (METZINGER 1993). The second, related property is the following. The I is always in the center of the experience. Philosophers call this perspectivalness, because we always see the world from our own perspective (METZINGER 1993). According to METZINGER (2000), perspectivalness means "the existence of a single, coherent and temporally stable model of reality, which is representationally centered around a single, coherent and temporally extended phenomenal subject." This also fits with the model, because the basis of the subjective experience is the data of the system's own body model. This, by the way, may provide our impression of the, subjectively, constant I, although, seen from the third-person perspective, we actually do change a lot during our lifetime. Similarly, this could explain the fact that after waking up in the morning we feel to be the same person as the day before. Probably we would have this impression even if we were dramatically changed overnight.

Falling asleep corresponds to loosing the first-person perspective. According to our hypothesis this corresponds to the model output becoming zero. Even in awake state, we are not aware of all parts of the body model in a given moment. There are "attentional islands" which can move to different regions. The attentional level may be increased in domains where the difference between the sensory input and the model output is large. This means that the sensory input is different from the expectation[2]. When undergoing mental relaxation, these islands may become flattened and this exercise is quite often followed by the transition to sleep. A spectacular case of resetting the body model has been reported by RAMACHANDRAN et al. (1995). In a patient suffering from phantom sensation in the amputated left arm, a lively subjective experience of moving this arm could be elicited when the intact right arm was both actively moved and seen in a mirror such that the mirror image of the right arm appeared at the position where the left arm would have appeared if it had been intact. The visual input appeared to be sufficient to activate the model of the arm.

Our hypothesis may be further supported by the following observation. In cases of asomatognosia an accidental destruction of brain structures may lead to the result that a part of the body, e.g., the left leg, can be visually recognized, but is not regarded as belonging to the own body. The patient considers it as a foreign body. At the same time, the leg cannot voluntarily be moved, although the muscles and peripheral nerves are intact. Following the above hypothesis, this result may be interpreted in such a way that the neuronal basis of the body model of the leg had been destroyed by the accident. As a consequence, neither a voluntary movement nor a feeling was possible. SACKS (1984) reported a self-observation in which the neurological cause was unclear. In this case, motor control was regained and this lucky experience was paralleled by the reoccurrence of the subjective experience of the leg. As both properties reoccurred together, it seems to be probable that both phenomena have a common basis. Most people may have experienced the following less critical situation: their arm has "fallen asleep",—yet the arm position that is seen and the one that is felt may not coincide. There is also another observation made by SACKS which fits the properties of the model. At the beginning of the reoccurrence of the subjective experience of his leg, it seemed to be either very long (many meters) or very short (some centimeters). As time passed these "oscillations" became less distinct and finally came closer to the real leg length. Qualitatively, this corresponds to the dynamic behavior of the MMC net when recovering from a disturbance (CRUSE/STEINKÜHLER 1993).

Taken together, these observations suggest the following hypothesis: a system comprises a HIP system, i.e., has a first-person perspective, when (a) it contains a manipulable internal world model that includes properties of the system's own body which (b) can be used to compare the ("virtual") data of this model with those provided by the "real" data from the sensory input (and the memory, see below). As a weaker form of the hypothesis, one can assume that these conditions are necessary for the first-person perspective to appear as an emergent property. The stronger version assumes that these conditions are also sufficient. In any case, our hypothesis suggests that only such items have the ability to become experienced that are represented in the manipulable world model. In this way, such a system includes all the elements proposed earlier by the different definitions of cognition. This, in particular, means that the operational definition of MCFARLAND/BÖSSER is justified insofar, as the latter concentrates on the functional aspect. Following our hypothesis, the neuronal system underlying the body model has a functional meaning, whereas the property of having a first-person perspective comes for free as a kind of epiphenomenon. This means that, the fact that we experience an internal perspective has, as such, no functional meaning, and therefore has not necessarily to be included in the definition. Another consequence, following from this hypothesis is that, if it was correct, also artificial systems could be constructed such as to have first-person perspective, i.e., to become HIP systems.

I do not see a way to make it intuitively clear that, under the conditions described, a first-person perspective necessarily develops. However, this is similar in other cases where unexpected system properties emerge as, for example, the property of resonance frequency in an electronic circuit. After having observed and described the resonance phenomena, including constructing a pendulum-like mental model, we have the impression of having understood the property of resonance. Another, possibly even better example concerns the occurrence of the phenomenon of life. It was a serious and intensively debated question between Vitalists and Mechanists at the beginning of our century, how nonliving matter could be transformed into living matter. The Vitalists argued that a special force, termed vis vitalis, is required. After intensive investigations, we now have developed a mental model showing that a special collection and arrangement of molecules has the property of living. We do not need a vis vitalis. Liveliness is understood as an emergent property.

The disadvantage of both examples is—as it is the case for the above mentioned metaphor of the two sides of a coin—that they describe the transition between entities which both belong to the domain of the third-person perspective. Therefore, these examples may not convince anybody taking a dualistic view. The following example might help to give a better intuition of a transition from the third-person perspective to the first-person perspective. Imagine you join for the first time a group of persons who already have well established social connections. At the beginning, you look at this group and the personal relations between the group members in a kind of third-person perspective. After becoming acquainted with the different people, you develop an internal model of the group and its internal relations, and, as you develop own relations to group members, you gradually experience the transition of an individual becoming a member of a group. When doing so, you develop a kind of first-person perspective for the group.

Expansions of the model

This model does, of course, not describe all necessary elements of a system which is able to control behavior. It, if at all, describes an atom, a basic element of a cognitive system. Therefore, several expansions could be envisaged. First, there are reasons to believe that not all movements are controlled by such a body model. On the contrary, for urgent actions there are presumably also many parallel systems for direct specifications of actions. These already mentioned reactive systems, however, are normally not cognitively penetrable and are therefore called non-manipulable (see Fig. 2). They form direct connections between sensors and motor layers, possibly including hidden layers. These connections are not shown in Fig. 3. The alien limb phenomenon might be explained by such influences. In these cases a limb of the patient, for example the left hand, moves, but the patient has the impression that this movement is not influenced by himself, but by some external agent. This means that the body model appears to be intact because the hand is perceived as the own limb. Voluntary control of movement, however, is not possible. As a hypothetical explanation one could assume that in this case the connection between body model and motor output is disturbed and hand movement is only controlled by such direct, non-cognitive influences.

Second, the body model itself has to be expanded, too. The exact border dividing objects belonging to the own body and belonging to the environment is not strictly fixed. A hammer in the hand can be experienced in the body image as an extension of the body. Also, a car driver may feel his body extended to the outer margin of the car's body. One may even feel kind of sick when the car does not work well. Depending on cultural influences, these borderlines may even include other people; an extreme example, for members of the western culture, may be the experience of labor-pains by males in the case of couvade.

Furthermore, as mentioned earlier, the body model is considered as being part of the world model. From this point of view, the model described up to now which only has knowledge concerning its own body including the possible expansions mentioned, comprises the most primitive version of a cognitive system. However, to be applicable for controlling the movement to reach an object, the model has to be enlarged by the capability of imbedding external objects which occur in the workspace. In the course of further development, more distant objects could be integrated to permit, for example, orientation. Already at this stage of development, this model contains internal representations or "concepts" (cf. C. ALLEN, this volume) of these objects, although application of language is not yet considered. The invention of spoken words by humans increased the manipulability of these virtual objects, allows for finer distinction between items, and improves the possibility of communicating between subjects. In this way, the internal world model starting with representations of only its own body, is expanded to contain near and distant extracorporeal objects as well as abstract entities. At least on this level, items stored in the manipulable model can be called "objective".[3]

Apart from this functional view, one might also ask in which part of the brain this model could be realized. As mentioned by ROTH (this volume), conscious experiences appear to be correlated with neuronal activities in the whole cortex (see also FLOHR 1996), but this may not be true for the primary areas. Furthermore, the content of consciousness is correlated with the locus of that activation. As shown by brain imaging techniques, imagination of arm movements, for example, are correlated with activity in parts of the premotor areas and the parietal lobe. Imagination of a visual stimulus, the picture of a tree, for example, activates the visual areas. During calculation parts of the prefrontal area and of the temporal lobe are activated, and so on. Therefore, the model that is explained here for a special case only, namely the representation and the control of arm position, may be applied in a similar way to other items and other cortical areas. Thus, these world models might form the basis of memory. Various of such models might be connected such that the output of one model ("a concept") forms the input to another model by replacing the sensory input of the latter. For example, instead of the sensory input describing the actual arm position the angle values of an arm position memorized in another context might be used.

Discussion

Whereas the considerations of ROTH (this volume) and FLOHR (1991; 1996) concern the question in which part of our brain the conscious events take place and whereas FLOHR in particular asks for correlated changes on the molecular level, and, furthermore, convincingly argues that under some conditions a first-person perspective necessarily will occur, I address the question of how the local mechanisms might work and how the transition between the third-person perspective and the first-person perspective could be understood. The proposal is made that it is the output of the world model which corresponds to the first-person perspective.

Are we really prepared to attribute a first-person perspective to an artificial system? There are two "opposite" ways to come to a wrong conclusion. We may consider even a simple robot in an animistic way, i.e., as having a plan or a will, at least, as long as we observe only the behavior of the system (think of the arguments you have with your computer). Being social animals we apparently are constructed that way. So when applying common sense, we may be inclined to falsely assume the existence of a "soul", an internal view even in the case of an extremely simple system. On the other hand, the reader may not be intuitively prepared to attribute a first-person perspective to the simple system shown in Fig. 3. Indeed, most probably each concrete circuit, even if it really described a HIP system, would probably lead to the intuition of a system not having a first-person perspective. We

Author's address

Holk Cruse, Fakultät für Biologie, Universität Bielefeld, Postfach 100 131, D-33501 Bielefeld, Germany
Email: holk.cruse@biologie.uni-bielefeld.de

tend to consider any hardwired system as possibly being complicated, but still of the clockwork type. Therefore, both observations show that intuition is a problematic advisor.

One reason why our intuition might have these problems could be that we compare the possible "inner life" of this primitive model to our rich subjective experiences. However, this direct comparison may not be justified. Such a simple system may indeed have crossed the principal threshold, but the liveliness of its first-person perspective might actually be very weak. The liveliness might become more intense and the feeling of the system may become more and more real, the more different sensory modalities contribute to the description of a given situation. This is supported by the observation of patients with sensory deficiencies who report that their feeling seems to be less real. In this way, a gradual change is conceivable, ranging from a very weak first-person perspective to full awareness, a change which we also experience during our daily life, in particular, when we change from sleep to the awake state.

The MMC net for the arm can be regarded as a "general purpose" representation of the Euclidean space in which the arm can move. The angle values (together with the fixed segment lengths) represent the workspace of the arm. The x–y coordinates describe a larger, visually given space. The recurrent network concatenates both spaces to form one unique space. In this way the MMC net corresponds to a mental representation as proposed by J. PROUST (1999).

Acknowledgments

I would like to thank the Wissenschaftskolleg zu Berlin for providing the possibility to work on this subject and Dr. Thomas METZINGER for very helpful discussions.

Notes

1 The underlying idea is that several geometrical relationships are used to calculate the same value (e.g., the angle of the elbow joint) several times in parallel. The final output is then obtained by calculation of the mean value of these multiple computations (MMC). This is then fed back for the next iteration. Any special constraint of the variables (e.g., joint limits) can be introduced.

2 According to GOMBRICH (1960), only those sensory inputs trigger our attention which are not expected.

3 Another possible expansion, not considered here, might be that the signals to be compared with the output of the world model may not only come directly from the sensors, but also from other memories which then may replace the sensory input.

References

Allen, C. (1999) this volume.

Braitenberg, V. (1984) Vehicles: experiments in synthetic psychology. MIT Press: Cambridge MA.

Brooks R. A. (1991) Intelligence without reason. Proc. Int. Joint Conf. on Artificial Intelligence. pp. 569–595.

Chalmers, D. J. (1995) The puzzle of conscious experience. Sci. Am. : Dec.: 62–68.

Condillac, E. B. de (1798/1983) Abhandlungen über die Empfindungen. Felix Meiner Verlag: Hamburg.

Cruse, H. (1979) Modellvorstellungen zu Bewußtseinsvorgängen. Naturw. Rundschau 32: 45–54.

Cruse, H./Steinkühler, U./Burkamp, C. (1998) MMC—a recurrent neural network which can be used as manipulable body model. In: Pfeifer,R./Blumberg, B./Meyer, J.-A./Wilson, S. W. (eds) From animals to animats 5. MIT Press: Cambridge MA, pp. 381–389

Cruse, H./Bartling, C./Dean, J./Kindermann, T./Schmitz, J./Schumm, M./Wagner, H. (1996) Coordination in a six-legged walking system. Simple solutions to complex problems by exploitation of physical properties. In: Maes, P./Mataric, M. J./Meyer, J.-A./Pollack, J./Wilson, S. W. (eds) From animals to animats 4. MIT Press: Cambridge MA, pp. 84–93.

Cruse, H./Steinkühler, U. (1993) Solution of the direct and inverse kinematic problem by a unique algorithm using the mean of multiple computation method. Biol. Cybern. 69: 345–351.

Damasio (1994) Descartes' error. Emotion, Reason and the human brain. G. P. Putnam's Son: New York.

Flohr, H. (1991) Brain Processes and Phenomenal Consciousness. Theory and Psychology 1: 245–262.

Flohr, H. (1996) Ignorabimus? In: Roth, G./W. Prinz, W. (eds) Kopf-Arbeit—Gehirnfunktionen und kognitive Leistungen. Spektrum Akademischer Verlag: Heidelberg, Berlin, Oxford.

Gehlen, A. (1971) Der Mensch. Athenäum: Frankfurt.

Gombrich, E. H. (1960) Art and Illusion: a study in the psychology of pictorial representation. Princeton University Press: Princeton.

Kawato, M./Gomi, H. (1992) The cerebellum and VOR/OKR learning model. Trends in Neurosciences 15: 455–453.

Lanz, P. (2000) The concept of intelligence in psychology and philosophy. In: Cruse, H./Dean, J./Ritter, H. (eds) Prerational Intelligence: Interdisciplinary perspectives on the

behavior of natural and artificial systems. Kluwer Press (in press).

Maes, P. (1991) A bottom-up mechanisms for behavior selection in an artificial creature. In: Meyer, J.-A., Wilson, S. (eds) From animals to animats. MIT Press: Cambridge MA.

McFarland, D./Bösser, Th (1993) Intelligent behavior in animals and robots. MIT Press: Cambridge, MA.

Metzinger, T. (1993) Subjekt und Selbstmodell. F. Schöningh: Paderborn, München, Wien, Zürich.

Metzinger, T. (2000) The subjectivity of subjective experience: A representationalist analysis of the first-person perspective. In: Metzinger, T. (ed) Neural Correlates of Consciousness: Empirical and Conceptual Questions. Cambridge MA: MIT Press.

Pfeifer, R. (1996) Building "Fungus Eaters": design principles of autonomous agents. In: Maes, P./Mataric, M. J./Meyer, J.-A./Pollack, J./Wilson, S. W. (eds) From animals to animats 4. MIT Press: Cambridge MA.

Proust, J. (1999) Mind, space and objectivity in non-human animals. Erkenntnis 51: 41–58.

Ramachandran, V. S./Rogers-Ramachandran, D./Cobb, S. (1995) Touching the phantom limb. Nature 377: 489–90.

Rizzolatti, G, Arbib, M. A. (1998) Language within our grasp. Trends in Neurosciences 21: 188–194

Roth, G. (1999) this volume.

Rumelhart, D. E./McClelland, J. L. (1986) Parallel distributed Processing. Explorations in the microstructure of cognition, Vol. 1. MIT Press: Cambridge MA.

Sacks, O. (1984) A leg to stand on. Gerald Duckworth: London.

Shepard, R. N./Metzler, J. (1971) Mental rotation of three-dimensional objects. Science 171: 701–703

Steinkühler, U./Beyn, W.-J./Cruse, H. (1995) A simplified MMC model for the control of an arm with redundant degrees of freedom. Neural Processing Letters 2: 11–15.

Steinkühler, U./Cruse, H. (1998) A holistic model for an internal representation to control the movement of a manipulator with redundant degrees of freedom. Biol. Cybernetics 79: 457–466.

Hans-Johann Glock

Animal Minds: Conceptual Problems

Abstract

This article discusses the problems which concepts pose for the attribution of thoughts to animals. It first locates these problems within a range of other issues concerning animal minds (sct. 1). It then distinguishes three positions: DAVIDSONIAN lingualism denies that non-linguistic animals have any thoughts or concepts; mentalism maintains that their thoughts differ from ours only in degree; an intermediate position maintains that animals can have thoughts of a simple kind (sct. 2). I opt for an intermediate position on the basis of considering those arguments for lingualism which trade on the connection between thoughts and concepts: the argument from the intensional nature of thought (sct. 3); the idea that thoughts involve concepts (scts. 4–6); the argument from the logical connections created by concepts (sct. 7). I conclude that there is a kind of perceptual thought that does not require concepts and that concept-possession is not tied to language, but to the capacity of making discriminations that are subject to normative assessment. At the same time, even the ascription of simple thoughts to animals employs a rich idiom with conceptual connections that go beyond the phenomena to which it is applied (sct. 8).

Key words

Animal Minds, concepts, normativity, DAVIDSON.

1. Conceptual problems and the problem of concepts

Concepts have a two-fold significance for animal minds. Like most philosophically interesting topics, animal mentality poses problems of a distinctively conceptual kind. What mentality, if any, one can attribute to non-human animals (henceforth "animals") depends not just on empirical findings (whether observations in the field or experiments in the laboratory) and scientific theories, but also on what one makes of heavily contested concepts like that of a mind, of thought, of consciousness, of behavior, etc. The philosophical task in this area consists not in collecting new empirical data about animal behavior, its neurological causes or its evolutionary origins, but in clarifying what it is to possess various mental properties, and hence in clarifying under what conditions such properties can be ascribed to organisms. This is not to say that one needs a cast-iron precise definition of these properties in advance of empirical theory-building. But it is to say that such theory-building must be accompanied by conceptual reflection on the provisional understanding of mental concepts that informs specific lines of research, methods and conclusions (see ALLEN/BEKOFF 1996, pp236–7).

Among the contested concepts that are important to the issue of animal minds, the concept of a concept occupies an important place. More specifically, concepts seem to pose a distinct obstacle to those theories that ascribe to animals a wealth of mental properties and capacities. On the one hand, many of these phenomena seem to involve or presuppose the possession of concepts, but on the other hand, concepts and conceptual capacities seem to be highly sophisticated features closely connected to language, and hence beyond the ken of animals. In this essay I want to discuss both *conceptual problems in* the field of animal minds, and the *problem of concepts for* animal minds.

Do at least some animals have minds which are comparable to those of humans? The question is both complex and vexed. It involves many different problems that have been treated from a variety of methodological perspectives, in subjects ranging from evolutionary biology and neurophysiology through ethology, psychology and linguistics to the philosophy of mind and language. Particularly baneful has been the intrusion of moral concerns. It is perfectly legitimate to discuss what implications the possession or lack of mentality should have for our treatment of animals. What is illegitimate is to tailor one's views of animal minds to suit one's ethical (or unethical) prejudices. Alas,

in many cases it is obvious that verdicts on animal minds are motivated neither by empirical observations nor by conceptual clarifications, but by a desire to affirm or deny that animals are worthy of various degrees of moral and legal consideration.

As one leaves behind the murky waters of applied ethics and animal welfare, other complications loom. The concept of a mind may be vague, but it is certainly wide-ranging. In dealing with the question of animal minds, we must therefore specify what mental phenomena and capacities are at issue. One possibility here is to simply provide a list, e.g., sensation, perception, memory, mental imagery, belief, emotion, desire, intention, consciousness, self-consciousness, awareness of other minds, reason, etc. But it may also be possible to proceed in a more systematic fashion.

For one thing, we must distinguish cognitive capacities, namely capacities to gain information or knowledge, from conative capacities, that is, capacities to have evaluative attitudes of pursuit or avoidance. The popular term "animal cognition" is misleading in that it seems to refer exclusively to the former, when in fact most of the debate under that title is as much concerned with animal desires and intentions as it is with animal belief.

Both cognitive and conative states are what philosophers call intentional; that is to say, they have a particular object or content. One believes or knows that something is the case, one desires that something be the case, one intends to do something. In cases in which this content is expressed by a that-clause, it is common to use the (in my view misleading) term "propositional attitudes". However, there are other mental phenomena which have no such object or content, whether propositional or other. This holds not just for sensations like pains or tickles, but also for relatively complex phenomena such as moods. One of the problems with concepts like experience and consciousness is that they seem to cover a variety of phenomena ranging from simple sensations through perception to full-blown propositional attitudes. It is one thing to be conscious in the sense of being capable of experiencing pain, which may simply be a mark of all sentient life. It is quite another thing to be capable of perceiving certain features of one's environment, and possibly yet another thing to be capable of believing that such–and–such is the case.

I shall refer to these intentional states as thoughts rather than propositional attitudes. It is unclear whether they are attitudes towards entities (see sct. 4); and even if they are, it would be sheer dogmatism to insist that loving MALCOLM X or intending to climb a tree are in the final analysis attitudes towards propositions. In any event, the capacity for thought is a precondition for the capacity of reasoning. In both theoretical and practical reasoning, one moves from one or more thoughts, the premises, to another thought, the conclusion. Reasoning for its part must be distinguished from intelligence. Roughly speaking, intelligence is the ability to solve hitherto unencountered problems in a flexible way, notably one which is not predetermined by the genetic outfit of the species (similarly DUPRÉ 1996, pp329–30). Intelligence in this general sense covers diverse phenomena ranging from problem-solving through trial and error, as with the Cappuchin monkeys in the trap tube task, to the kind of insight and foresight displayed by chimpanzees in the fashioning of tools (see TOMASSELLO/CALL 1997, pp10–11, ch. 3). But even in the latter case, it is an open question whether reasoning from premises to conclusions is involved (see sct. 6).

Because of my interest in concepts, I shall focus on animal thought, addressing the issues of sensation, perception, intelligence and reason only in so far as they bear on the theme of concepts and thought. That vertebrates have sensations and are capable of perceiving their environment should be obvious, and so is that many of them, especially higher mammals, are capable of intelligent behavior of various degrees. But only in the case of thought is there any prima facie reason for believing that concepts must be in play. Conversely, the capacity to reason clearly presupposes conceptual capacities, but only because reasoning involves thoughts.

Even within the area of animal *thought* as here understood, there are several issues. From DESCARTES onwards, the main philosophical argument against crediting animals with thoughts and concepts has been that both require language. However, recent research in cognitive ethology suggests that chimpanzees and bonobos are capable of acquiring elementary linguistic skills (SAVAGE-RUMBAUGH/LEWIN 1994; SAVAGE-RUMBAUGH/SHANKER/TAYLOR 1998). These findings have been disputed by CHOMSKIAN linguists (PINKER 1994, ch.11), who complain in particular that these apes lack the syntactical capacities characteristic of language. So one question is whether some animals possess language. If that question is answered in the affirmative, then several arguments against animal thought simply do not apply to these cases (see STEPHAN 1999, pp88–90).

The other question, which is my exclusive concern here, is whether *non-linguistic* creatures are ca-

pable of thought. This keeps out signing apes, but brings in non-linguistic humans. However, with the exception of occasional side-glances, I shall focus on *non-linguistic animals*, partly because non-linguistic humans raise problems of an entirely different kind. For example, it is unclear to what extent our attribution of thoughts to pre-linguistic children is based on an anticipation of their later linguistic skills. At the same time, I shall invoke not just actual forms of animal behavior, but also forms of behavior that a non-linguistic creature *could* engage in. For it is this possibility which is crucial to the conceptual question of whether our mental concepts preclude *ab initio* the possibility of non-linguistic animal thought.

2. Options and perspectives

At long last, we have a question which may shed light on the role of concepts for animal minds without being absurdly general. Can animals without language have thoughts? For most of this century, there has been a tendency to assume that any positive answer must be based on anthropomorphism, an illegitimate projection of human features onto animals. However, with the rise of cognitive psychology and ethology the tide has turned. It has been shown, for example, that the blanket charge of anthropomorphism is both diffuse and ill-founded (FISHER 1996).

Alas, at present there is an opposite danger. Current debates about the mind are increasingly dominated by evolutionary theory and by naturalism, the idea that philosophy has no distinctive contribution to make over and above that of the special sciences. In this climate it appears that opponents of animal thought are misguided "human exceptionalists" who ignore the "continuity across species" that has been discovered by evolutionary biology and neurophysiology (JAMIESON 1998). In this vein, various biological principles of continuity have been invoked to show that the mental differences between humans and animals can only be a matter of degree (CRISP 1996). From this perspective, any attempt to draw qualitative distinctions between humans and animals appears deplorably anthropocentric and out of touch with natural science.

This blanket charge of anthropocentrism is no more warranted than the opposing charge of anthropomorphism. There is no gainsaying the empirical fact that there is both biological (in particular, genetic and neurophysiological) similarity and evolutionary continuity between us and certain non-linguistic animals. But it does not follow that they must approximate to our mental life. For one thing, the extent to which biological evolution can involve genuine leaps is controversial, as the debate over the possibility of punctuated equilibria shows (see GOULD/ELDRIDGE 1977) For another, even if *natura non facit saltus*, continuity along lineages of evolutionary development has no implications for the mental capacities of the animals around us. Although it is probable that our closest evolutionary ancestors without language shared many of our other mental capacities, these ancestors are extinct; and there is no guarantee that the biologically closest extant species is mentally close to us. If all vertebrates except *homo sapiens* had been exterminated by a wayward meteorite, it would be absurd to conclude that starfish and sea cucumbers must be mentally close to us (see PINKER 1994, p346).

It so happens that our closest extant relatives, the chimpanzees, share 98% of our DNA. However, it does not follow that they share 98% of our mental life, simply because small biochemical differences in genotype may lead to significant differences in phenotype. In fact, this already holds at the level of neurophysiology. Although the brains of chimpanzees are made of the same matter, they are significantly smaller than those of humans (on average, 400ccm to 1400ccm), even if body-size is taken into consideration. In any event, however, our mental vocabulary captures neither genetic nor neurophysiological differences, but differences in the kinds of behavioral and perceptual capacities we humans are interested in. To that extent, our mental concepts themselves may be anthropocentric; yet it does not follow that it is anthropocentric to insist that these concepts preclude application to non-linguistic creatures (see DAVIDSON 1985, p473). Of course, some naturalistic philosophers have suggested that we should replace our mental vocabulary with a more scientific jargon based on neurophysiology. But whatever the merits of that proposal, it will not solve the problem that fuels the debate about animal minds, namely whether it is legitimate to apply our current mental concepts to animals.

As regards that question, one can distinguish not just proponents and opponents, but a whole spectrum of views. At one end we find lingualists like DAVIDSON, who deny that non-linguistic animals have any thoughts. The other, mentalist end is occupied by empiricists like HUME, who think that the thoughts of animals differ from those of humans only in degree, due to their different perceptual inputs. Oysters don't have thoughts about bicycles, simply because they cannot perceive bicycles. In one

sense it is also occupied by those cognitive representationalists who explain even simple animal behavior by reference to a rich variety of complex thoughts and calculations, except that these thoughts are held to be in a language of thought, not in a public language. In addition to these two extremes, however, there is an intermediate position, adopted by a coalition (rare, some might say) of common sense and WITTGENSTEIN. It holds that animals are capable of having thoughts of a *simple* kind, namely those that can be expressed in non-linguistic behavior.

A second complication arises because of the various dimensions of the concept of thought. Even if we leave aside thought as the capacity for reasoning and confine ourselves to intentional states, we have to distinguish between two parameters, namely the kind of state on the one hand, the kind of "content" on the other. One question is what intentional verbs can be applied to animals; another question is what that-clauses, singular terms or infinitives can follow these intentional verbs. Concerning the first parameter, it is possible to grant that a dog can know, believe or see that *p*, but to deny that it can think or hope that *p*. Concerning the second, WITTGENSTEIN famously suggested that a dog can believe that its master is at the door, but not that its master will return the day after tomorrow (1967, p174; see also RUNDLE 1997, ch. 3; SAVIGNY 1995).

Finally, there are diverse views on the status we should accord to ascriptions of thoughts to animals. It is possible to regard the ascriptions of some types of thoughts to some species of animals as empirically false. This is the case, for example, when ethologists like SEYFARTH/CHENEY (1996, pp340–3) use observations to deny that vervet monkeys have a "theory of mind", that is, beliefs about the beliefs of conspecifics. It is not easy to see how one could regard the ascription of *all* types of thoughts to *all* species of animals as empirically false. In any event, however, most opponents of animal thought do not take this line. They often regard such attributions not as false, but as suffering from a more basic defect, namely that of being meaningless, nonsensical or based on a category-mistake (see FISHER 1996, pp4–8). This is part and parcel of the aforementioned fact that disputes about animal minds are often rooted not in disagreement about the empirical evidence, but in a diverse understanding of the concepts involved. If the concept of thought is such that it precludes application to non-linguistic creatures, nothing could count as evidence that some animals have thoughts. Ascribing thoughts to animals would then make no more sense than ascribing a color to a number.

These days, the harsh charge of nonsense is much less popular among analytic philosophers than it used to be. But even those who are prepared to accept that some thought ascriptions to animals are neither conceptually incoherent nor empirically false often insist that they cannot be taken literally, but must be regarded as figurative, metaphorical or secondary. In a similar vein, Bede RUNDLE has suggested that ascribing thoughts to animals is a mere description or redescription of behavior, whereas ascribing thoughts to humans can be a genuine explanatory hypothesis, to be confirmed by what the subject is prepared to divulge (1997, ch. 4). On either view, to say that an animal believes that *p* may be no more than a *façon de parler*.

This stance is illustrated by DAVIDSON's reaction to the following tale from Norman MALCOLM.

"Suppose our dog is chasing the neighbor's cat. The latter runs full tilt toward the oak tree, but suddenly swerves at the last moment and disappears up a nearby maple. The dog doesn't see this maneuver and on arriving at the oak tree he rears up on his hind feet, paws at the trunk as if trying to scale it, and barks excitedly into the branches above. We who observe this whole episode from a window say, "He thinks that the cat went up that oak tree" (1972–73, p13).

MALCOLM claims that we would be right to say this, and DAVIDSON acknowledges that it is *prima facie* plausible. Nevertheless, he insists that strictly speaking MALCOLM's dog cannot believe anything, because he does not possess a language (see 1985, pp474–6, 478; 1984, p155).

An immediate objection to DAVIDSON is this: animals must be capable of having thoughts because we have no better way of explaining and predicting their behavior than by attributing thoughts to them (e.g., BENNETT 1976, §§7–8; FODOR 1975, ch. 1). According to DAVIDSON, this provides a pragmatic justification for our attribution of thoughts to animals, but it does not show that animals can have thoughts. In attributing thoughts to animals, we merely treat them *as if* they were capable of acting for reasons (beliefs and desires), just as one might explain the movements of a heat-seeking missile by ascribing to it the desire to destroy an airplane. In this way, "we can continue to explain the behavior of speechless creatures by attributing propositional attitudes to them while at the same time recognizing that such creatures do not actually have propositional attitudes. We will be bound to acknowledge that we are applying a pattern of explanation that is far stronger than the observed be-

havior requires, and to which the observed behavior is not subtle enough to give point" (1985, pp477–8).

This account treats thought-attributions to animals as useful fictions. However, DAVIDSON himself concedes that animal behavior is much closer to human behavior than the movements of heat-seeking missiles, and that we know of no better way of explaining the behavior of animals than by attributing thoughts to them. But these concessions invite an objection DAVIDSON ignores.

We regard attributing thoughts to animals not just as *convenient*, as he would have it, but as entirely *justified*. For, unlike attributing desires and beliefs to complex missiles, such attributions are not based on *technological ignorance* but on a *biological insight*, namely that the life and behavior of animals shows them to have both wants and perceptual capacities. DAVIDSON might reply that the alleged insight is merely an illusion of ARISTOTELIAN folk-biology, since animal behavior could be fully explained by reference to physiological processes, if our knowledge of physiology were sufficiently advanced. However, this invites the question of why human behavior should resist such explanation. And a lingualist who responds that human behavior resists mechanical explanations because it rests on thoughts would be begging the question at issue.

Nevertheless, there may be a kernel of truth in DAVIDSON's suggestion that the pattern of explanation we employ with respect to animal behavior in some sense outstrips the explanandum, because it is originally tailored to the explanation of the more complex behavior of linguistic creatures like ourselves. In this article I want to bring out that kernel of truth. At the same time, I shall resist DAVIDSON's lingualism. Instead, I opt for a version of the intermediate position, both with respect to the kinds of thoughts one can ascribe to animals, and with respect to the status of such descriptions. We can ascribe thoughts and concepts to animals, but these are restricted to a simple kind. For only simple thoughts can be identifiably manifested in non-linguistic behavior. Furthermore, although the ascription of thoughts and concepts to animals can be more than a useful fiction, it is incongruous in that the rich mental idiom we employ has conceptual connections that go beyond the phenomena to which it is applied.

I shall argue for this position from a third-person perspective. That is to say, I shall not appeal to phenomena—whether mental or neurophysiological—that *cannot* be manifested in behavior even *in principle*. For one thing, one cannot make sense of the notion of a belief as a private attitude completely detached from behavior and its explanation (DAVIDSON 1984, p170; 1985, p476). For another, although thoughts can be more or less definite, we can ascribe a specified thought that *p* to a creature *a* only if something counts as *a* thinking that *p* rather than that *q* (see GLOCK 1997, pp166–8). Finally, even if attributing beliefs or concepts to creatures which are totally incapable of manifesting them is not strictly meaningless, it is vacuous and cannot be part of a proper behavioral science.

This third-person perspective contrasts with a mentalist approach that posits mental faculties and processes which are unreflected in behavior and lie forever beyond human consciousness. I shall not attempt to defend the third-person approach against this alternative here.[1] But it is important to note that such a third-person perspective is not tantamount to behaviorism: thoughts can be ascribed on the basis of behavior, without therefore being reducible to behavioral dispositions (DAVIDSON 1985, p476). Furthermore, the third person perspective is adopted not just by common folks, but also by cognitive ethology in its scientific investigations of animal mentality.

To be sure, many cognitive ethologists make free use of the notion of a mental representation, especially with respect to concepts. One reason is that they regard this as an anti-dote to the behaviorist perspective which repudiates mental predicates and describes animal behavior exclusively in terms of stimulus and response. But in contrast to mentalist philosophers (e.g., CARRUTHERS 1989), their postulation of mental representations remains tied to the explanation of animal behavior. Furthermore, circumspect practitioners stress that this notion of a mental representation is a modest one, which does not require the imputation of "pictures in the head" or of symbols in a "language of thought" (TOMASELLO/CALL 1997, pp7–12). Instead, the main purpose of the terminology seems to indicate that animal cognition goes beyond immediate sense–perception.[2] For this reason, in most contexts talk of mental representations can be replaced by talk about higher cognitive and mental capacities, such as the ability to remember, to reason with foresight, etc. As we shall see (sct. 6), such abilities are importantly different from behavioral dispositions. Whether the causal explanation of those abilities refers to physical tokens of computational types in the brain, as cognitive representationalists have it, or whether it involves other factors, as connectionists and many neuroscientists maintain, is a separate question.

Moreover, it is a question that can and should be left be open when the issue is whether animals have thoughts or concepts.

In any event, to talk about abilities rather than representations avoids the suggestion that humans and animals alike relate to the world not directly, through exercising their cognitive and conative capacities, but through "inner" intermediaries. In my view, this shift of perspective is especially apposite in the case of concepts. There is a mentalist tradition, which regards concepts as representations in the minds or brains of individuals, and a PLATONIST tradition, according to which they are building-blocks of abstract entities. The former view has problems in doing justice to the objective nature of concepts, the fact that they can be shared between subjects which could not share private mental ideas and need not share patterns of electro–chemical activity in the brain.[3] The latter view has difficulties in explaining the role concepts play in thought and action. Both types of problems are avoided by a third position, going back to ARISTOTLE and KANT, which regards concepts as principles or rules of classification and explains the possession of concepts as the possession of classificatory abilities.

In this essay, I cannot justify either my animadversions against inner representations or my preference for the notion of abilities that can be manifested in behavior. My aim is more modest. I want to show that even from such a third person perspective, the various connections between thoughts and concepts point in the direction not of lingualism, but of an intermediate position. The three connections that threaten the possibility of animal thought have featured in several authors, but are especially prominent in the work of DAVIDSON, who is the most important current lingualist.[4] I shall discuss in turn:

- the intensional nature of thought (sct. 3);
- the claim that thoughts must involve concepts (scts. 4–6);
- the holistic connections between thoughts (sct. 7).

3. The intensional nature of thoughts

One problem with attributing thoughts to animals is that without verbal responses we cannot make the fine distinctions between different thoughts (beliefs, desires) expressed in the same non-verbal behavior. Thought-attributions to humans create intensional contexts: if we substitute co-referential terms within the content-clause, this may lead from a true attribution (e.g., "Sarah believes that Cicero was Roman") to a false one (e.g., "Sarah believes that Tully was Roman"). In the case of animals, by contrast, substitution of co-referential expressions often leads from attributions which we commonly regard as true to attributions which are absurd or unintelligible. The oak tree that the cat went up also happens to be the oldest tree in sight and the same tree the cat went up last time the dog chased it. But does MALCOLM's dog believe that the cat went up the oldest tree in sight, or the one it went up last time? Equally, a dog can know that its master is at the door. But does it also know that the president of the bank is at the door? "We have no real idea how to settle, or make sense of, these questions" (DAVIDSON 1984, p163; see also CHATER/HEYES 1994). The reason is that the dog can think neither that its master is the president, nor that he is not.

One response to this failure of intensionality is to hold that in the sentence

(1) The dog thinks that the cat went up that oak tree

the expression "that oak tree" occurs transparently (in QUINE's terminology). Accordingly, (1) is paraphrased so as to avoid problems of intensionality, e.g., as

(1') The dog thinks, with respect to that oak tree, that the cat went up it.

But as DAVIDSON (1985, pp474–5) points out, this response presupposes that "the *de re* description picks out an object the believer could somehow pick out". If MALCOLM's dog could not distinguish the oak tree from among other objects (e.g., the pine tree or the garden fence), we might still causally explain his behavior by reference to the oak tree, just as we explain the convulsions of an oyster by reference to its being pricked with a needle. But (1') would no longer be appropriate. For *de re* constructions like "with respect to" or "of" require an anaphoric referent in the subsequent content-clause, an "it" which the now disabled dog could not distinguish from other things.

However, it remains an open question whether this requirement might not be met by non-linguistic creatures on account of their possessing certain *discriminatory capacities*. The dog believes something of the tree "under some description", namely one that expresses those features by which the dog recognizes the tree and distinguishes it from other objects. This is the point at which concepts come in. If the dog is to be capable of recognizing the tree by certain features, it seems that it must have the concept of a tree. It needs not only to distinguish trees from other objects, but to recognize them by certain features. *If* this is granted, it might even provide for

a non-linguistic analogue of intensionality, at least in simple cases. Animals cannot know objects by different descriptions, but they might recognize them by different features, and they might fail to recognize that these are features of one and the same object. Thus it is conceivable that MALCOLM's dog reacts in one way to a man with heavy foot-steps being at the door, in another to his master being at the door, because he has not yet realized that the heavy stepper is his master. In that case it may be true that he believes that the heavy stepper is at the door, but not that his master is at the door, or *vice versa*.

But according to lingualists, crediting the dog with the concept of a tree makes no more sense than crediting him with beliefs about trees. The dog cannot believe of an object that it is a tree, "unless we suppose the dog has many general beliefs about trees: that they are growing things, that they need soil and water, that they have leaves or needles, that they burn. There is no fixed list of things someone with the concept of a tree must believe, but without many general beliefs there would be no reason to identify a belief as a belief about a tree, much less an oak tree" (DAVIDSON 1985, p475).

Two lingualist objections emerge here: a general one concerning the holistic connections between thoughts created by concepts, which will be discussed in section 6, and a more specific one concerning concepts to which I turn now. Consequently, the force of the argument from intensionality depends on the arguments from concepts and from holism. The original argument questions whether there is a that-clause which would capture the content of an animal's thought. This question now gives way to the question of whether such clauses would not be implying too much conceptual sophistication, and to the question whether they would not drag in too many other thoughts.

4. Thoughts and concepts

One prominent obstacle obstacle which concepts create for animal thought is this. The thoughts we ascribe to animals in common parlance involve *concepts* with which the animal cannot be credited. Attributing thoughts to animals on the basis of non-linguistic behavior is misguided, since these thoughts involve concepts which themselves cannot be attributed on such a basis, because they require general beliefs which could only be manifested through language. Accordingly, concept-possession and the ability to have thoughts amount to one and the same thing, and both are confined to language-users (DUMMETT 1993, chs. 12–3; DAVIDSON 1997, pp24–5; 1999, pp7–8).

This raises two problems. Firstly, can animals possess concepts at all, and, if so, what kind of concepts? Secondly, if they cannot have concepts of any kind, does that really preclude them from having thoughts or beliefs?

One will have to answer the second question in the affirmative if one accepts the mentalist idea that thoughts are occurrences in the mind or brain, or the PLATONIST idea that they are abstract entities which have concepts as their components. On either account, one cannot have or grasp the thought without having or grasping its constituent concepts. However, the PLATONIST picture has no explanatory value. It transposes the relation of spatial part and whole from an area where it is perfectly clear—material objects—onto an area where spatial concepts *ex hypothesis* make no sense. At least in its modern version, the mentalist picture looks very concrete: concepts are temporal stages of physical processes in the brain. But even *if* identifiable neurological processes correspond to having parts of a thought as opposed to having a whole thought (something which has been doubted on both empirical and conceptual grounds), these are only stages of the thinking of a thought, not of its content.

Furthermore, DAVIDSON himself has maintained that "propositional attitudes" are not relations between a subject and an abstract or mental phenomenon, but rather modifications of a person (1994, p232). In my view, he is right. Although that-clauses are grammatically speaking noun-phrases and can function as accusatives, they no more refer to a genuine object than noun-phrases like "everything" or "the past" (see GLOCK 1997a). If this is correct, for *a* to believe that *p*, *a* need not stand in a relation to an object (a proposition) which would involve standing in a relation to components of that object (concepts).

The building-block model according to which small entities—concepts—combine to form large entities—thoughts—is not a truism but a problematic metaphor. What seems to give content to that metaphor is exclusively the fact that the linguistic expressions of thoughts—namely sentences—have components—namely words. But as lingualists like DUMMETT and DAVIDSON themselves have stressed, at the linguistic level the building-block model is problematic. We do not construct the senses of sentences out of the meanings of words; rather, we assign meanings to words by noting their role within sentences. For this reason, DAVIDSON's approach to hu-

man belief is "holophrastic": we ascribe beliefs to linguistic creatures on the basis of their assenting to sentences as a whole (1984, pp4, 22, 220–5; 1997, p25).

The possibility he ignores is that rejecting the building-block picture in the linguistic case invites an analogous move in the case of animals, a *holodoxastic* approach that starts out from the whole belief. Concepts are not the building-blocks of thoughts, they are abstractions from thoughts, because, as KANT noted, their sole function is to be used in judgements. According to such an approach, what matters is precisely a kind of "modification": if a creature can be correct or mistaken as to how things are, it can have beliefs. Although the sentences we use in ascribing thoughts have components, our ascriptions are not based on a prior ascription of these components. Instead, they are based on the subject manifesting certain perceptual capacities, attitudes and emotions.

In the non-linguistic case, these manifestations will obviously not include assent to sentences. But they will include forms of behavior, postures and facial expressions which higher animals share with human beings. When we say that MALCOLM's dog believes that the cat went up that oak tree, we do not do so on the grounds that it picks out objects and classifies them in a way that corresponds to the singular and general terms we use in the attribution—that is why, for philosophical purposes, it may be more accurate to rephrase (1) in the transparent manner of (1'). Rather, we simply note the dog's reaction to its environment. We regard these reactions as directed towards particular objects, creatures and events, because we assume that dogs have certain perceptual capacities and wants, assumptions which require rudimentary knowledge of the way dogs live (what they can recognize, what they tend to dislike, etc.).

Because of its reliance on behavioral reactions, the holodoxastic move is confined to simple beliefs, notably about perceptible features of the subject's environment. But it suffices to blunt the force of the line "No thoughts without concepts!". It may seem that a problem remains nonetheless. Granting thoughts to animals while denying them concepts suggests that there is an incongruity between ascribing thoughts to animals and ascribing thoughts to linguistic creatures. In the second case, our ascriptions impute to the believer a grasp of the concepts involved, whereas in the first they do not. This creates a pressure for holding that intentional verbs like "believes" are ambiguous, referring either to a holodoxastic, behavioral phenomenon—notably perceptual experience—or to a conceptual, linguistic one like full-blown thought.

In this vein, MALCOLM suggests that while the dog can "believe" that the cat went up the oak tree, only humans can "have the thought" that it went up the oak tree. Similarly, DUMMETT maintains that while humans can have thoughts consisting of concepts, animals have mere "protothoughts" consisting of spatial representations. But this type of distinction seems to count against ascribing one and the same belief to humans and animals. It suggests that "Both Sarah and the dog believe that *p*" is not so much a falsehood as a zeugma, like "Both the exam and the chair were hard". For "Sarah believes that *p*" comes out as "Sarah has the thought that *p*" while "The dog believes that *p*" comes out as "The dog has the protothought that *p*".

In my view, however, we can grant that there are important differences between the beliefs of conceptual and those of non-conceptual creatures, yet resist the pressure towards postulating distinct objects and hence distinct attitudes. A certain disparity between the terms used in a belief report and those that could be used by the subject is present even in the linguistic case, without constituting a fundamental incongruity. The terms which occur in the content-clause are in general dictated not so much by the creature whose belief we report, but by the concerns of speaker and audience. Thus, "Sarah thinks that the charlatan you introduced me to is about to give her a biscuit" can be in order, whether Sarah is an adult, a child that lacks the concept of a charlatan, or a dog (RUNDLE 1997, p83).

Consequently, it is far from obvious that attribution of beliefs requires attribution of concepts. Still, the question whether animals can possess concepts remains relevant, for two reasons. When it comes to attributing beliefs to animals, some terms are more absurd than others. This suggests that animals can be credited with some concepts but not others. Furthermore, if some animals have conceptual capacities, the lingualist argument from concepts fails even if it is right to tie beliefs to concepts.

5. Animal concepts

With respect to animal concepts, one finds the same spectrum of opinion we already encountered concerning animal thoughts. KANT, FREGE, DAVIDSON and DUMMETT occupy the lingualist corner. According to them, animals can perceive, but lack concepts of *any* kind. In the other corner are cognitivists, who have no qualms about ascribing complex concepts

to animals. An intermediate position is occupied, for example, by Anthony KENNY, who maintains that animals can possess some concepts, namely those that can be manifested in non-linguistic behavior (1989, pp36–7; see also DEGRAZIA 1996, pp154–6).

Proponents of this position have to concede that the concepts animals have are often not the ones we use in ascribing thoughts to them. The discriminations which underlie animal behavior may not coincide either extensionally or intensionally with our verbal classifications. A dog might group cats together with hamsters or distinguish black cats from all others; and even if it groups all and only cats together, it might recognize them by smell rather than visually. But this by itself is no obstacle to ascribing to them concepts that *differ* from ours. For example, when Sue SAVAGE-RUMBAUGH's chimpanzees distinguish foodstuffs and tools, the operative difference seems to be simply that between the edible and the inedible (1986, p257). Accordingly, what kind of concepts we should ascribe to avoid anthropocentrism is a matter for ethological discoveries. We discover animal concepts by discovering the parameters governing their discriminatory behavior. Such considerations are likely to indicate that our ordinary ascriptions require qualification, but not that they involve the kind of convenient pretence lingualists have diagnosed.

Whether this criticism holds water naturally depends on what one makes of concepts and concept-possession. According to one construal, concepts are principles of discrimination, and to possess a concept is to have the ability to recognize or discriminate different types of things (PRICE 1953, p355; DUPRÉ 1996, p331). This construal is *prima facie* plausible, and has served as a starting point even for many who regard concepts as a type of mental representations. It certainly implies that animals can possess concepts. Both in the wild and in the laboratory, animals distinguish between a host of different colors, tastes, sounds, shapes, stuffs, quantities, types of creatures, etc. Moreover, many of these capacities are learned rather than innate.[5]

DAVIDSON thinks that this account of concepts falls prey to a *reductio ad absurdum*. "Unless we want to attribute concepts to butterflies and olive trees, we should not count mere ability to discriminate between red and green or moist and dry as having a concept, not even if such selective behavior is learned" (1997, p25). Leaving aside butterflies for the moment, I agree that it would be absurd to credit olive trees with concepts. But I do not accept that this absurdity follows from treating concepts as powers of discrimination. Olive trees do not discriminate between moist and dry soil, since discrimination is a prerogative of *sentient* creatures, that is, animals. We must distinguish between mere differential reaction to causal inputs, which is a universal feature of physical phenomena, and discrimination, which is tied to creatures with perceptual capacities.

Nevertheless, even proponents of animal concepts have come to accept DAVIDSON's conclusion, namely that conceptualization requires more than discrimination (e.g., ALLEN/HAUSER 1996; ALLEN 1999; STEPHAN 1999). But there is much less agreement on what more is needed.

One noteworthy proposal is that of ALLEN and HAUSER. They distinguish between "recognizing an X" and "recognizing something *as* an X or recognizing it to be an X". The former is merely the "extensional characterization of a discriminatory ability", while the latter requires the ability to recognize an X on the basis of several different properties, notably on the basis of properties which transcend perception. Perhaps these properties should even be essential rather than accidental to Xs (1996, p51).

I find this proposal unconvincing. For one thing, it is far from clear that "recognizing an X" is extensional: it would seem, for example, that one can recognize a flash of lightning without recognizing a certain type of electric discharge. For another, the proposal rules out the possibility of distinguishing either between perceptual and more abstract concepts or between having more or less rich concepts of an X. Thus it implies that before the advent of instruments to measure the length of electromagnetic waves, humans had no color concepts, because they could recognize colors in only one way—by looking—and one which is purely perceptual. I find it even less plausible to suppose that one has the concept of an X only if one distinguishes Xs by those features which we regard as essential to Xs.

What is correct, however, is that for creatures that are capable of distinguishing between essential and accidental properties, what concept of an X they possess is determined by what properties they regard as essential to being an X. Now, the distinction between essential and accidental properties is beyond the grasp of animals. But this should not count against their having concepts, since even many philosophers, especially radical empiricists, have at best a tenuous grasp of the difference. Distinguishing essential from accidental properties is, in my opinion, crucial to a proper understanding of concepts, but it cannot be crucial to their possession.

6. Concepts and normativity

For these reasons, I favor a different way of separating conceptualization from discrimination, namely one which stresses the *normative* dimension of conceptual classification. This idea goes back to WITTGENSTEIN, and it has been succinctly expressed by DAVIDSON. "To have a concept is to classify objects or properties or events or situations", or, more accurately, to be able to do so. Powers of discrimination, by contrast, are mere "dispositions", and therefore, "as WITTGENSTEIN emphasized, have no normative force". Such dispositions do not involve the ability to *recognize a mistake*, and hence no knowledge of the difference between correct and incorrect behavior (1997, pp24–5; see 1985, p480).

It is at least partly the absence of such classification which makes us reluctant to credit butterflies with concepts. Furthermore, the normativity required for such classification presupposes that the classifier can make a mistake which she is capable of recognizing as such. A WITTGENSTEINIAN distinction can help to clarify this point.[6] To be capable of classifying or misclassifying things, a creature *a* must not just have a disposition to behave *in accordance with a rule*—as butterflies do when they land only on red petals—but of *following a rule*. That is to say, the principle which distinguishes *F*s from non-*F*s must be part of *a*'s *reason* for differentiating between *F*s and non-*F*s, not just a law to which its discriminating behavior conforms.

If *a* can classify things into those which are (an) *F* and those which are not, it must be possible that *a* should be mistaken, namely in taking something as not being *F* which is in fact *F*, or in taking something as *F* which is in fact not. But *a* can be accused of making a mistake in applying the rule which distinguishes *F*s from non-*F*s only if *a* is also capable in principle of recognizing that she has violated that rule. Only given that possibility can *a* be said to diverge from a rule which she was trying to follow, i.e., to have acted contrary to her own intentions. Otherwise, *a* is merely diverging from our expectations or from a statistical norm. As DAVIDSON points out, a slippery road may be a danger or a nuisance, but it does not commit a mistake. *Mutatis mutandis*, a butterfly that fails to discriminate between red and green may reduce its biological fitness, but it does not violate a principle to which it has committed itself. There are of course types of mistakes that do not require this possibility, e.g., failures to perform in line with evolutionary design. But without an intention to perform in this way, such failures are not misapplications of a rule.

However, the crucial question is whether such normative behavior is the prerogative of linguistic creatures. DAVIDSON thinks it is, because the behavior of animals and children is purely mechanical, no different in principle from that of a heat seeking missile (1997, p25). This argument rightly assumes that only intentional discriminations can be corrected in the relevant sense, because only *intentional* behavior can be accused of misapplying a principle of classification. It also intimates, again rightly, that such behavior must be voluntary in the sense that the agent could have done otherwise. Mechanical behavior—no matter whether unconditioned or conditioned—cannot be accused of failing to live up to a principle, roughly because ought implies can. This is why classification is not the exercise of a mere disposition to react differently to distinct external influences, even if, as with animals, these influences are stimuli perceived by a sentient creature. Rather, classification is the exercise of an *ability*. Unlike dispositions, abilities are not automatically exercised under specified conditions; the agent can intentionally exercise them or refrain from exercising them.[7]

At the same time, DAVIDSON is wrong to hold that non-linguistic creatures have only dispositions and lack abilities. The behavior of non-linguistic creatures is not always explicable solely by reference to immediate biological imperatives. Both prelinguistic children and the great apes are capable of voluntary action, because they can refrain from a particular action, either by pursuing their goals in a different way or by forsaking them, at least temporarily.[8] By the same token, in a particular situation they are capable of either heeding or disregarding a difference. That is to say, they are capable of distinguishing objects of types *F* and *G* in one situation, and of ignoring the difference in another. It would be wrong to ascribe such classification to butterflies. But it seems equally wrong to deny classification to chimpanzees capable of selecting or making tools in advance of attempting a task. For these creatures deliberately distinguish between different kinds of objects (e.g., different types of stones) in some situations, but may disregard the difference in others, or if they are not in the mood (see BYRNE 1995, pp150, 187–9, 225, ch. 7; TOMASSELLO/CALL 1997, pp36, 78). And here the question of whether a discrimination has been learned is important. An unconditioned reflex cannot be the exercise of an ability, but a learned pattern of response can, because it is not necessarily automatic.

This response *may* confine non-linguistic concept possession in the normative sense to infants and the great apes. However, it does not make concept possession dependant on language possession, but on discriminatory behavior that is sufficiently complex and flexible to be subject to normative assessment. Once more, the lingualist could not reply that the appearance of flexible behavior among animals is deceptive, without inviting the same challenge concerning humans.

The idea that cognition in animals is a measure of the complexity and flexibility of their behavior is widespread (e.g., TOMASELLO/CALL 1997, pp7–12; DUPRÉ 1996). DAVIDSON's point that the ability to recognize and correct mistakes distinguishes classification from discrimination has also been taken on board in recent contributions (ALLEN 1999; STEPHAN 1999, p87). I want to suggest that these two points are connected through two requirements: classification must be rule-guided, and rule-guided behavior must be intentional.

However, these requirements may yet prove to be the undoing of animal concepts.[9] For it is possible to argue as follows. Even if animals are capable of acting *voluntarily*, in the sense of doing otherwise, and of acting *intentionally*, in the sense of acting for a purpose, they are incapable of acting intentionally in the stronger sense of *acting for a reason*. We explain the behavior of animals by reference to reasons, (e.g., "the dog runs to the oak tree because he wants to catch the cat"). But in doing so we indicate only what their purposes or goals are, not *how they have reasoned*, i.e., what their *justification* is for acting as they do. For, as RUNDLE (1997, ch. 4) has argued, that would presuppose that they are in principle capable of stating such reasons. If this is correct, we can rule out animal concepts by ruling out animal reasoning. Animals might discriminate for a purpose (e.g., to reap certain rewards), but they cannot *reason*. Even though their discriminations may be voluntary, they do not follow rules: they do not distinguish *F*s from non-*F*s for the reason that *F*s possess certain distinctive features.

But is this line of argument correct? Can a creature only act for a reason if it is capable of communicating this reason? Take a chimpanzee that has learned to use different tools in the pursuit of *dorylus* ants and *macrotermes* termites. It is plausible to maintain that its reason or justification for matching tool and prey is that they possess certain features. This impression is strengthened by the fact that chimpanzees display non-linguistic forms of behavior that go together with the correction of error among humans, such as hesitation, displeasure, discarding one type of tool in favor of another, etc. More generally, it is far from obvious that animals are incapable of reasoning. Chimpanzees seem to do just that in their construction and employment of tools in advance of feedback from the task itself.

Finally, consider the story of CHRYSIPPUS' hunting-dog (see SORABJI 1994). In chasing a prey of which it has lost the scent, this dog reaches a crossroads; it sniffs down the first path, then sniffs down the second path, then it immediately follows the third *without* sniffing. In the case of dogs, perhaps such behavior could only be a rigid conditioned reflex. But I can see no reason for denying that this is an intelligible form of behavior for a non-linguistic creature capable of voluntary action. And if it is, what is wrong with the explanation that the behavior evinces a disjunctive inference ("*p* or *q* or *r*; neither *p* nor *q*; ergo *r*")? We might grant that there is a difficulty in describing such a creature as silently consulting a principle. But as Ryle has convincingly argued, even the intelligent performances of humans are rarely accompanied by conscious consultations of this kind.

However, a related problem has been pointed out to me by Anthony KENNY. Although humans need not actually verbalize their reasoning, they are capable of doing so. In the absence of this capacity, the question arises of what in an animal's behavior could correspond to the *ergo* of linguistic reasoning. This point is unanswerable with respect to creatures like dogs. But in the case of chimpanzees there *can* be an analogue to our *ergo*, however thin. In the context of encountering and pondering a problem, certain gestures and grimaces, followed by renewed activity, can be interpreted as marking the point when the shilling dropped. Even if this is an anthropomorphic interpretation in the case of chimpanzees, we can easily imagine a non-linguistic hominoid whose facial expressions and gestures are so close to ours as to make such a description inevitable. Furthermore, even without the *ergo*, as regards context (problem solving), demeanor (e.g., head scratching) and result (problem solution), the deliberations of chimpanzees are close enough to those of humans to qualify as instrumental reasoning.

In my view, therefore, there is no compelling case for claiming that animals cannot possess concepts. Moreover, even if there is, it does not suffice to deny them thoughts, because of the possibility

of holodoxastic belief of perceptual kind. Lingualists are right to insist that "concept-formation is not a way station between mere dispositions ... and judgements" (DAVIDSON 1997, p25). Concepts and judgements remain on a par. A chimpanzee capable of classifying things into sticks and knifes, and hence of concepts like "stick", can also believe that the object it confronts is a knife, or wish that it were a stick. My point is rather that it has yet to be shown either that concept-formation and judgement require *linguistic* judgement or that holodoxastic belief cannot be a way station between mere dispositions and judgement involving concepts.

7. The holistic nature of thought

There is one last argument against animal concepts and, by implication, against the possibility of conceptual thought among animals. It is that the concepts which feature in the beliefs we commonly ascribe to animals require certain general beliefs with which we cannot credit them. This is part of a wider qualm, namely that attributing thoughts to animals is incompatible with "the intrinsically holistic character of the propositional attitudes", the alleged fact that "to have one is to have a full complement" (DAVIDSON 1985, p473; see also STICH 1979). Since at least some members of that complement are definitely beyond their pale, animals cannot even have the simple beliefs commonly ascribed to them.

Elsewhere I have argued that this general line of argument rests on holistic principles which are either too strong for the lingualist case, because they preclude plausible cases of human thought, or too weak, because they allow for some forms of animal thought (GLOCK 2000, sct. 4).This leaves more specific qualms about the holistic costs involved in crediting animals with beliefs involving specific concepts. That argument does not threaten the possibility of holodoxastic belief (sct. 4). To assess its consequences for conceptual belief, one should distinguish between two types of general beliefs, namely conceptual and empirical (without denying that some of the examples are borderline cases, e.g., that trees are growing things that need soil and water).

DAVIDSON insists that beliefs presuppose certain general empirical beliefs on account of their constituent concepts. This is implausible for many of his examples (see 1985, p475; 1984, p200). It would seem that one can believe that the cat went up the oak tree without knowing that trees burn, or that one can believe that a cloud is passing before the sun without knowing that clouds are made of water vapor. Such a radical conceptual holism also creates a general problem, namely that any empirical discovery amounts to a conceptual change, with the possible consequence that scientific theories talk about different things as soon as some of their empirical claims are at odds (FODOR/LEPORE 1992, ch. 1).

DAVIDSON is on firmer ground when he invokes general beliefs which are clearly conceptual, e.g., that cats are animals or continuing physical objects that move in certain ways (1999, pp8–9). Animals cannot recognize the conceptual status of such beliefs, nor do they need to in order to have thoughts. But as DAVIDSON himself admits, they can "generalize" in the sense of reacting similarly to similar stimuli (e.g., 1985, p480). By this token, they can also in principle distinguish not just between, e.g., mice and cats, but between animals, plants and inanimate objects. The notion of a continuing physical object is more problematic, since it is in many respects a result of philosophical reflection that exceeds the requirements of ordinary human thought (see STRAWSON 1992, ch. 2). But if it is spelled out in a pedestrian way, it is clear that chimpanzees can learn to distinguish, for example, between physical objects on the one hand, mirror reflections or TV images on the other. Finally, we must remember that a creature could have concepts without having *our* concepts.

This means that there is no holistic argument against animal concepts over and above the arguments discussed in section 5. More generally, holism does not provide a compelling objection against the possibility of animal thought. This is no reason to abandon holism altogether. There are plausible holistic principles which exclude the possibility of a creature having just a single belief. In any event, the complexity required for conceptual belief is incompatible with a behavioral repertoire capable of exhibiting just a single belief. But these reflections do not establish that the web of which any belief must be part need extend as far as the web of sophisticated human thought. There may be larger and smaller webs. What kind of web is required may depend on the belief and the creature concerned. From the fact that an animal lacks our web of beliefs and our concepts, it does not follow that it has no beliefs and no concepts (see BEKOFF/JAMIESON 1991, pp19–20; DUPRÉ 1996, p332; DEGRAZIA 1996, pp154–8; ALLEN 1999, p39).

8. Conclusion

Where does all this leave the lingualist thesis that the capacity for thought requires the capacity for language, especially on account of concepts? As regards intensionality (sct. 3), the problem is not that a non-linguistic analogue (recognizing the same objects through different features) is ruled out in principle, but that the range of distinctions that can be displayed in non-linguistic behavior is restricted. The same message emerges from my moderate holism, according to which thoughts come in larger or smaller packages (sct. 7). Packages that include beliefs manifestable only in linguistic behavior are the preserve of language-users. Again, rejecting the building-block conception of thoughts and concepts makes room for employing intentional verbs like "believes", "wants" or "intends" to animals, without necessarily imputing concepts (sct. 4). Finally, whether or not animals can have concepts of a simple kind depends not on their having a language, but to the extent to which their discriminations are rule-governed and hence intentional (scts. 5–6).

To this extent, our discussion reinforces the intermediate position. At the same time, they also support DAVIDSON's suggestion that animals do not just have thoughts of a simpler kind, their having these thoughts *amounts* to something simpler. In the case of animals, there is at most an analogue of intensionality. In so far as thought-ascriptions to animals are holodoxastic, they are not only restricted to thoughts about perceptible features of the environment, they also lack conceptual connections which apply in the human case: we cannot infer from the fact that the dog thinks that *x* is *F* that the dog grasps the concept *F*. Furthermore, even if animals can have concepts, these are not just confined to concepts of a (roughly speaking) perceptual kind. Animals are also incapable of satisfying one of the two criteria which we standardly use in attributing concepts to humans. They may be able to *apply* principles of classification, but not to *explain* them. In fact, the two restrictions are linked. A chimpanzee may discriminate between its keeper and other humans just as deliberately as it does between red and black ants. But we are more inclined to ascribe to it the concept of redness than the concept of a keeper, because there is so much more to explain with respect to the latter. Finally, even those thoughts animals might be credited with stand in a smaller logical space. They lack the kind of context which characterizes sophisticated linguistic thought.

Accordingly, attributions of simple thoughts to animals are neither intensional, nor conceptual, nor holistic in the way thought-attributions to humans are. However, the best analogy is not the anthropomorphic explanation of missiles, but one DAVIDSON has suggested in discussion. Attributing thoughts to animals is like using numerals for the purpose of labeling members of a football team. Although natural numbers stand in complex relations of order and numerical difference, these relations are ignored in this context. What matters here is not the numerical difference between two numbers, nor even which one is greater, but only that no two numbers should be used for the same player.

The analogy is illuminating. Thought attributions to animals employ a rich conceptual apparatus to an area in which many of the logical connections which constitute that apparatus do not apply. But it breaks down in one important respect. Attributing thoughts to animals is not simply an impoverished application of a rich technique. For that richer technique evolves around a central core of cases in which creatures believe, know or desire things on account of their wants and perceptual capacities. These biological basics of belief are shared by humans and animals. At the same time, when we move from this core area in the direction of conceptual thought, we also move *in the direction* of linguistic thought. The features which non-linguistic creatures must possess to be capable of conceptual thought—intentionality, complexity, flexibility—correspond to those features by which theorists from DESCARTES to CHOMSKY have distinguished language from more basic systems of communication. In this respect, at least, our reflections tend to confirm rather than to negate the connection between thought and language.[10]

Author's address

Hans-Johann Glock, Department of Philosophy, The University of Reading, Reading RG6 6AA, United Kingdom.
Email: h.j.glock@reading.ac.uk

Notes

1 For a defense of a third person perspective on animal thought, see DUPRÉ (1996). For a critique of the explanation of animal behavior by reference to complex calculations, see SEARLE (1997), and for a judicious assessment of cognitive representationalism BECKERMANN (1999), ch. 10.

2 As far as I can tell, this holds true even for authors wedded to a stronger notion of mental representation, e.g., in ALLEN/HAUSER (1996, pp54–5).

3 These qualms are not due to a failure to heed ALLEN's distinction between social and individual concepts (1999, p35). Even a concept that is not actually shared between several individuals must be capable of being shared, and the extent to which individuals share a concept can be determined independently of establishing patterns of neural activity in the brain.

4 DAVIDSON's work also features other arguments in favor of lingualism, including the argument that having a belief requires having the concept of a belief and that the latter requires language because of the need for "triangulation". I discuss these arguments in GLOCK (2000), but leave them aside here, since in my view they are less revealing about the connection between thoughts, concepts and language.

5 For primates, see TOMASSELLO/CALL (1997), chs. 4–5), and for the even more surprising achievements of pigeons, see HERNSTEIN/LOVELAND/CABLE (1976).

6 The distinction is explained in detail by BAKER/HACKER (1985, pp154–8).

7 For this distinction, see KENNY (1975). In ordinary parlance, dispositions include character traits which are half-way between dispositions in this technical sense and abilities: they are neither realized automatically nor simply subject to the will.

8 This will be obvious to the parents of any pre-linguistic two-year old. As regards the great apes, see GOODALL (1986), MENZEL (1974), BYRNE (1995), chs. 8–9.

9 Even if this requirement is met in the case of certain animals, it still remains to be shown that they really correct mistakes, rather than simply alter their behavior. And if they do, does this require that these animals have second-order beliefs? For a discussion of these questions, see GLOCK (2000), sct. 5. The role of rules for animal behavior also plays a role in HENDRICHS (1999).

10 For advice and comments on related material, I should like to thank Ansgar BECKERMANN, Peter HACKER, John HYMAN, Sir Anthony KENNY, John PRESTON, Eike von SAVIGNY, Stuart SHANKER and Michael TOMASSELLO. This paper was written during tenure of a research fellowship from the Alexander VON HUMBOLDT Foundation, for which I am grateful.

References

Allen, C. (1999) Animal Concepts Revisited: The Use of Self-Monitoring as an Empirical Approach. Erkenntnis 51: 33–40.

Allen, C./Bekoff, M. (1996) Intentionality, Social Play and Definition. In: Bekoff, M./Jamieson, D. (eds) Readings in Animal Psychology. MIT Press: Cambridge MA, pp. 229–239.

Allen, C./Hauser, M. (1996) Concept Attribution in Nonhuman Animals: Theoretical and Methodological Problems in Ascribing Complex Mental Processes. In: Bekoff, M./ Jamieson, D. (eds) Readings in Animal Psychology. MIT Press: Cambridge MA, pp. 47–62.

Baker, G. P./Hacker, P. M. S. (1985) Rules, Grammar and Necessity. Blackwell: Oxford.

Beckermann, A. (1999) Analytische Einführung in die Philosophie des Geistes. de Gruyter: Berlin, New York.

Bekoff, M./Jamieson, D. (1991) Reflective Ethology, Applied Philosophy, and the Moral Status of Animals. Perspectives in Ethology 9: 1–47.

Bennett, J. (1976) Linguistic Behavior. Cambridge University Press: Cambridge.

Byrne, R. (1995) The Thinking Ape Oxford University Press: Oxford.

Carruthers, P. (1989) Brute Experience. Journal of Philosophy 86: 258–269.

Chater, N./Heyes, C. M. (1994) Animal Concepts: Content and Discontent. Mind and Language 9: 209–246.

Crisp, R. (1996) Evolution and Psychological Unity. In: Bekoff, M./Jamieson, D. (eds) Readings in Animal Psychology. MIT Press: Cambridge MA, pp. 309–321.

Davidson, D. (1984) Inquiries into Truth and Interpretation. Oxford University Press: Oxford.

Davidson, D. (1985) Rational Animals. In: LePore, E./ McLaughlin, B. (eds) Actions and Events. Blackwell: Oxford, pp. 473–480.

Davidson, D. (1994) Davidson, Donald. In: Guttenplan, S. (ed) A Companion to the Philosophy of Mind. Blackwell: Oxford, pp. 231–236.

Davidson, D. (1997) Seeing Through Language. In: Preston, J. (ed) Thought and Language. Cambridge University Press: Cambridge, pp. 15–27.

Davidson, D. (1999) The Emergence of Thought. Erkenntnis 51: 7–17.

DeGrazia, D. (1996) Taking Animals Seriously. Cambridge University Press: Cambridge.

Dummett, M. (1993) Origins of Analytical Philosophy. Duckworth: London.

Dupré, J. (1996) The Mental Lives of Nonhuman Animals. In: Bekoff, M./Jamieson, D. (eds) Readings in Animal Psychology. MIT Press: Cambridge MA, pp. 323–336.

Fisher, J. A. (1996) The Myth of Anthropomorphism. In: Bekoff, M./Jamieson, D. (eds) Readings in Animal Psychology. MIT Press: Cambridge MA, pp. 3–16.

Fodor, J. (1975) The Language of Thought. Crowell: New York.

Fodor, J./LePore, E. (1992) Holism: a Shopper's Guide. Blackwell: Oxford.

Glock, H. J. (1997) Philosophy, Thought and Language. In: Preston, J. (ed) Thought and Language. Cambridge University Press: Cambridge, pp. 151–169.

Glock, H. J. (1997a) Truth without People? Philosophy 72: 85–104.

Glock, H. J. (2000) Animals, Thoughts and Concepts. Synthese 119.

Gould, S. J./Eldredge, N. (1977) Punctuated Equilibria: The Tempo and Mode of Evolution Reconsidered. Paleobiology 3: 115–51.

Goodall, J. (1986) The Chimpanzees of Gombe: Patterns of

Behavior. Harvard University Press: Cambridge MA.

Hendrichs, H. (1999) Different Roots of Human Intentionality in Mammalian Mentality. Erkenntnis 51: 145–164.

Jamieson, D. (1998) Animal language and thought. In: Craig, E. (ed) The Routledge Encyclopedia of Philosophy. Routledge: London.

Kenny, A. J. P. (1975) Will, Freedom and Power. Blackwell: Oxford.

Kenny, A. J. P. (1989) The Metaphysics of Mind. Oxford University Press: Oxford.

Malcolm, N. (1972–3) Thoughtless Brutes. Proceedings and Addresses of the American Philosophical Society 46.

Menzel, E. W. (1974) A Group of Chimpanzees in a 1-acre Field. In: Schrier, A. M./Stollnizt, F. (eds) Behavior of Nonhuman Primates. Academic Press: New York, pp. 83–153.

Pinker, S. (1994) The Language Instinct. Penguin: Middlesex.

Price, H. H. (1953) Thinking and Experience. Hutchinson: London.

Rundle, B. (1997) Mind in Action. Oxford University Press: Oxford.

Savage-Rumbaugh, S. (1986) Ape Language: From Conditioned Response to Symbol. Oxford University Press: Oxford.

Savage-Rumbaugh, S./Lewin, R. (1994) Kanzi. Doubleday: London.

Savage-Rumbaugh, S./Shanker, S./Taylor, T. (1998) Apes, Language and the Human Mind. Oxford University Press: Oxford.

Savigny, E. von (1995) Doubts Concerning Hopeless Dogs. In: Hintikka, J./Puhl, K. (eds) The British Tradition in 20th Century Philosophy. Hölder-Pichler-Tempsky: Vienna, pp. 207–12.

Searle, J. (1997) The Explanation of Cognition. In: Preston, J. (ed) Thought and Language. Cambridge University Press: Cambridge, pp. 103–126.

Seyfarth, R./Cheney, D. (1996) Inside the Mind of a Monkey. In: Bekoff, M./Jamieson, D. (eds) Readings in Animal Psychology. MIT Press: Cambridge MA, pp. 337–343

Sorabji, R. (1994) Animal Minds and Human Morals. Routledge: London.

Stephan, A. (1999) Are Animals Capable of Concepts? Erkenntnis 51: 79–92.

Stich, S. (1979) Do Animals Have Beliefs? Australasian Journal of Philosophy 57: 15–28.

Strawson, P. F. (1992) Analysis and Metaphysics. Oxford University Press: Oxford.

Tomassello, M./Call, J. (1997) Primate Cognition. Oxford University Press: Oxford.

Wittgenstein, L. (1967) Philosophical Investigations. Blackwell: Oxford.

Rolf Lachmann

Animal Perception as Value Perception

Abstract

Understanding animal consciousness demands a sophisticated methodology. One approach takes evidence steming from human consciousness as its starting point. Here again various approaches are possible. This paper claims the fruitfulness of a phenomenological investigation and illustrates it concerning perceptual consciousness. The phenomenological investigations of Alfred North WHITEHEAD, Arnold GEHLEN and Maurice MERLEAU-PONTY converge in their demonstration of the primitive layer of consciousness as pragmatic value perception. The hypothetical extension of human perception to animal perception is supported by Susanne K. LANGER's interpretation of animal consciousness which gives additional weight to this interpretation by pointing to evidence steming from experiments on human perception.

Key words

Pre-objective knowledege, phenomenological method, primitive layer of perception.

THE ASSUMPTION THAT animals are conscious beings is widely accepted. However, many also hold that their mode of consciousness differs from our human consciousness. The question arises from these two assumptions of how to characterize animal consciousness. Human consciousness has turned out to be a very tricky subject and difficult to grasp. Thus a large part of twentieth century psychology has tried to evade any reference to consciousness altogether in order to become an objective science. While understanding our own consciousness is difficult enough, animal consciousness poses an even larger problem. We cannot even rely on direct "knowledge by acquaintance" as we can in the case of our own conscious life.

An adequate understanding of animal consciousness can only be reached by a combination of evidence and research results from various approaches. In this paper I will depict one approach to animal consciousness and illustrate it by developing a hypothesis on animal perception. This approach assumes that human consciousness still includes traces of animal consciousness which are, however, not open to immediate inspection but have to be discovered in the course of a systematic investigation. Two possible forms of systematic investigation are the study of experimental results originating in human psychological research and the phenomenological method developed by philosophers in the first half of this century. On the basis of first hints and suggestions gained through these two avenues of approach, hypotheses can be formulated which, in turn, have to be verified with available biological knowledge such as on the behavioral capability or deficiency in animals. The final outcome is a "reflective equilibrium": a plausible and highly corroborated interpretation. In the long run, such investigations have to be restricted to particular species. Nevertheless, it seems to be possible to develop plausible hypotheses on animal perception and, until such time as neurophysiology has developed more substantial approaches, these are the best interpretations we can accomplish.

Restricting this paper on animal perceptual consciousness has the advantage that here in both heuristic approaches one does not have to start at the very beginning. There are numerous experimental results regarding human perception. There are also penetrating phenomenological investigations of human perceptual consciousness. In this paper I will first give an account of various interpretations of human perceptual consciousness which are either explicitly phenomenological or can at least be understood to be such. In outlining the interpretations of Alfred North WHITEHEAD, Arnold GEHLEN and Maurice MERLEAU-PONTY I will draw attention to the primitive layer of perceptual consciousness as understood in these interpretations. Despite slight shifts in emphasis, their interpretations lead to a convergent view which serves as a plausible candidate for an interpretation of animal perception. I will then turn to Susanne K. LANGER's interpretation of ani-

mal perception. In her interpretation LANGER relies a great deal on findings which stem from experiments on human perception. Indeed, the leading idea of this paper is derived from LANGER's theory and intends to give her interpretation additional support. The goal of this paper is both to clarify the methodological position and to form a hypothesis on animal perception.

1. Phenomenological investigations of human perception

Phenomenological investigation is essentially a methodologically guided attempt to an unprejudiced description. The major sources of prejudices against which phenomenology directs its objections are the traditional interpretations of philosophical theories, the theoretical accounts of science and the prejudices of common sense. The central goal of phenomenology is a fresh, comprehensive and unbiased look at its particular subject. It tries to ignore the numerous interpretations already given and steps back, for example, from the ordinary understanding of human perception. Max SCHELER, one of the founders of the "phenomenological movement"[1], criticized both the philosophical and psychological theories of perception of his time in that they never ask "What *is in fact given* in our perceptual consciousness?" but only "What *can be given* in our perception?" The answer to the latter question was, of course, that only what is provided by our sense organs can be given. Hence, man receives various visual, acoustic or kinesthetical impressions and, by combining them, organizes his knowledge. This view, however, is in sharp contradiction to perceptual consciousness as we experience it. Phenomenology protests against any account of human perception which ignores the factual complexities and appearances of our perceptual life (SCHELER 1973, p55). The shared opinion of authors who try to take a new and more penetrating view of human perception is that the one-sidedness and the difficulties of many modern philosophical problems derive from a highly artificial view of human perception. Therefore, a new look at the real ingredients and structures of human perception is needed.

Despite the fact that WHITEHEAD and GEHLEN did not consider themselves primarily as phenomenologists, their contributions can be regarded as phenomenological investigations. In combination with MERLEAU-PONTY's phenomenology of perception the shared view emerges that human perception is never an "isolated given" but always replete with references and complexities. All three philosophers draw attention to the essential relatedness of perception and regard this as the basis for symbolic references. In the following I will ignore this aspect of their interpretations and focus on their depiction of the primitive layer of human perception.

2. Alfred North Whitehead: The emotional basis of perception

WHITEHEAD's position that perceptual consciousness includes much more than our sense perceptions is central to his interpretation of human perceptual consciousness. Over and above our clear and distinct perceptions of form, color etc. our perceptual consciousness also includes numerous indefinite and vague impressions. In order to clarify the different perceptual layers WHITEHEAD distinguishes between two forms of perception: perceptions in the mode of "presentational immediacy" and perceptions in the mode of "causal efficacy".

Perceptions in the mode of presentational immediacy are the range of distinct perceptions in the focus of our waking consciousness when we see colors and forms or have smell, taste, acoustic or touch impressions (see WHITEHEAD 1985, p25). They are what is ordinarily understood as perception. These perceptions appear in our waking consciousness as "directly given" or "immediately present". Therefore WHITEHEAD calls them perceptions in the mode of "presentational immediacy". They are relatively clear and distinct. Here we are confronted with a complex variety of distinct perceptions, which are, to a large extent, controllable at will by closing our eyes, ears or nose. Moreover, these perceptions are relatively insignificant for our existence. Loss of eyesight does, of course, impair our ability to orient ourselves in our surroundings. It does not, however, threaten our physical existence. WHITEHEAD characterizes perceptions in the mode of presentational immediacy as "vivid, precise, and barren" (WHITEHEAD 1985, p23). They present a complex richness of distinct perceptions but are without intrinsic relevance. They are in themselves nothing but a spectacular "adventitious show".

But there is another perceptual sphere. WHITEHEAD refers to its content as perceptions in the mode of "causal efficacy". In this case we feel forces acting upon us, sense a causal influence. Our feeling of causal influence is not in the focus of our conscious attention but it is always present. We feel that each

concrete moment is influenced and shaped by the contents of its immediate predecessor. The pain we had a moment ago impresses itself on each succeeding moment and remains effective until it fades away or is ruled out by opposing forces. In every actual moment we feel that the pain has been forced upon us. Here we feel the causal insistence of the past and its determining force. WHITEHEAD draws attention to the fact that every instant is a continuation and reproduction of its predecessors in the immediate past: "Our primitive perception is that of 'conformation' vaguely..." (WHITEHEAD 1985, p43). In contrast to the "vivid, precise, and barren" presentations of forms and colors, perceptions in the mode of causal efficacy are vague, massive and important. They are "...heavy with the contact of the things gone by, which lay their grip on our immediate selves." (WHITEHEAD 1985, p44). In this perceptual mode we have the vague but insistent feeling of forces acting upon us. Hence, they are felt with great emotional weight. Moreover, we cannot control or ignore these perceptions in the same way as we can control perceptions in the mode of presentational immediacy. Perceptions in the mode of causal efficacy create a continuous background of feeling at the base of our consciousness. These perceptions are closely connected with our bodily feelings (WHITEHEAD 1968, pp152–153).

It may appear odd to call these feelings of causal efficacy perceptions since they are not perceptions in the sense of being derived from our sense organs. It might seem more adequate to speak of an "feeling of efficacy" or "experience of efficacy". Indeed, WHITEHEAD describes this perceptual mode as "nonsensuous perception" (WHITEHEAD 1967, p180). Nevertheless, it must be regarded as a form of perception because it is not a self-invented product but a reception of factors which are external to each particular moment in question.

WHITEHEAD adds another distinguishing feature in the perceptual modes. Whereas perceptions in the mode of presentational immediacy are simply "given" without any reference to a past or a future, perceptions in the mode of causal efficacy are full of moving and directing force. They are vectorial in character. At the root of our consciousness there is a vague but constant feeling of movement, activity, expansion towards, retreating from, and passage.

Hence, perceptions in the mode of causal efficacy are overpowering, inescapable and motivating. In contrast to the "adventitious show", the superficial and secondary product of complex subtleties, the directing force of perceptions in the mode of causal efficacy is seen in our emotions of anger, hate, fear, love or attraction. "The basis of experience is emotional. Stated more generally, the basic fact is the rise of an affective tone originating from things whose relevance is given." (WHITEHEAD 1967, p176). At the basis of our consciousness lies a vectorial and directing layer. The intrinsic emptiness and barrenness of our clear perceptions contrasts with the deep emotional significance of our causal feelings.

In order to act successfully we have to take the surrounding forces into consideration. Therefore we use visual and acoustic perceptions in order to anticipate causal forces.[2] This is the elementary form of symbolic reference: the use of distinct perceptions to indicate surrounding forces.

3. Arnold Gehlen: Perceptual enrichment through practical behavior

This last aspect of practical orientation, a combination of distinct presentations and the feeling of forces is also a central theme of Arnold GEHLEN's interpretation of human perceptual consciousness. GEHLEN develops his interpretation as part of his question concerning the biological "special place" of man. The central idea of his theory is the demonstration that man is biologically maladjusted. In his organic constitution man is a "deficient being". This is particularly obvious in man's almost total loss of instinctive reaction, a fact that also has far-reaching consequences for his perceptual consciousness. Whereas for animals perceptions function essentially as signs, e.g., as triggers for genetically determined instinctive reactions, they have no biologically prepared function for man. Because of man's loss of instinctive reaction perceptions have no biologically prepared behavioral significance. This leads to "man's unusual receptivity to perceptions" (see GEHLEN 1988, p27). GEHLEN interprets the human situation as "flooded with stimulation" in "an abundance of impressions" (GEHLEN 1988, p28), to which the infant is exposed after birth. "Since he cannot rely upon instincts for understanding his environment" (GEHLEN 1988, p28)[3] the child has to be relieved of the burdensome mass of perceptions.

In order for man to be relieved of the burden of sensory impressions, learning becomes imperative. Contrary to animals man has to build up his life independently and shape it step by step. The essential task is to enrich his perceptions with meaning. Since perceptions have no biologically determined "meaning", the perceptual enrichment with mean-

ing has to be reached by active manipulation of the surrounding situations.[4] GEHLEN gives a detailed account of the endless manipulation and "working through" of things by children. In their development a considerable amount of time is spent on touching, moving, seeing, manipulating and deforming everything within their reach in an attempt to experience all there is to know about their properties, their uses, and their effects. These activities are not motivated by any direct practical necessity. "In general, these efforts consist of the 'movements' that occupy man's childhood, movements through which man gradually experiences the world around him. In this process, objects are seen, touched, moved, handled in communicative movements of interaction ... As a result, the world is 'worked through' with an eye toward its potential availability and usefulness to man. In succession, objects are experienced by man and then set aside; the objects are thereby unwittingly endowed with a high degree of symbolism such that, eventually, the eye alone (an effortless sense) can take them in and quickly assess their potential usefulness and value." (GEHLEN 1988, *p*32). The result of this extended manipulative behavior is an enrichment of our distinct perceptions with pragmatic "values". He illustrates his interpretation by the perception of a teacup. In our ordinary perception we see the teacup extremely selectively. For example, we ignore the variation of its color by the distribution of light and shade, its ornamentation and other esthetic features. On the other hand, without closer examination we immediately grasp its pragmatically relevant features such as its weight, material structure, stability and handling possibilities. The pragmatically relevant aspects of surrounding things are the focus of our ordinary perceptual attention. The teacup exists for us primarily as a condensation of our previous handling experiences, as a complex of various "visual indications of movement" (GEHLEN 1988, p55 and p160).

Thus, in the course of our manipulating experiences, our perceptions become "symbolic", they are loaded with "meaning". At first glance, the perceptual constellation informs us about the pragmatic aspects of a thing or situation. "We orient ourselves in the sensory world through certain minimal optical, acoustical, tactile, etc. symbols. From a biological perspective, this is highly expedient: it saves us from getting caught up in the profuse abundance of things; furthermore, the greatest possible degree of receptivity to stimuli in an organism is not the purpose of perception. Our visual perceptions give us 'symbols'—of outcomes and reactions we can expect, of resistance we might encounter. With the aid of these symbols, we can initiate and regulate movements, even before success of failure has occurred." (GEHLEN 1988, p158). In the course of this development our perceptual "contact" with surrounding things is reduced to a minimum. Our perceptual consciousness is relieved of the burden of too many functionally meaningless sense perceptions.

4. Maurice Merleau-Ponty: The biological formation of perception

GEHLEN's view that our daily perception is saturated with "pragmatic values" reinforces WHITEHEAD's position that perceptions in the mode of presentational immediacy function as directing our awareness to the causally relevant aspects of our environment. MERLEAU-PONTY's phenomenological investigation of our perception clarifies another aspect. Similar to WHITEHEAD, MERLEAU-PONTY's main theme is his opposition to the empiricist psychology of association. The assumption that our perceptual consciousness consists of clear, distinct and isolated sense perceptions contrasts sharply with the real nature of our perceptual consciousness. MERLEAU-PONTY claims that the assumption of an isolated perception of distinct features is a myth. We never have a completely isolated perception. On the contrary, even the most simple perception, say of a red spot, includes structures and relations. One elementary structure is that we see the spot against a background, or the spot has an outline, a border, internal shading etc. We spontaneously judge all the points as either belonging to the spot or to the surrounding space. The elementary embeddedness of the spot in a background and all other structurings have the quality of a "phenomenal field". "The perceptual 'something' is always in the middle of something else, it always forms part of a 'field'." (MERLEAU-PONTY 1962, p4).

Despite the elementary relatedness of any perception it is far from true that every perceptual structuring is completely distinct and determined. On the contrary, most elements of our perceptual consciousness remain vague and indeterminate. Our perceptual consciousness is not a range of clear and distinct perceptions but is largely amorphous and ambivalent. "The perceived, by its nature, admits of the ambiguous, the shifting and is shaped by its context." (MERLEAU-PONTY 1962, p11) MERLEAU-PONTY calls this vaguely structured perceptual consciousness a "pre-objective realm" (MERLEAU-PONTY 1962,

p12). He illustrates his interpretation by referring to the experience of vague localization. Normally, when a particular part of our skin is touched, we can locate it spatially even if we cannot see the touching. We have a spatial impression. We can also feel the distance between various stimulated spots. Nevertheless, these impressions often remain vague and imprecise. This becomes evident when we are asked to point our finger to the spot. Our finger often has to move around that area of the skin in order to identify the exact spot. "*A vaguely located* spot is a contradictory phenomenon and reveals a pre-objective space where there *is* extension, since several points on the body touched together are not confused by the subject but as yet no univocal position, because no spatial framework, persists from one perception to another." (MERLEAU-PONTY 1962, p50).[5] Although our perceptual experience is more or less structured it remains obvious that it is not objectively determined and clearly given. MERLEAU-PONTY calls this layer of consciousness a "non-positing consciousness" (MERLEAU-PONTY 1962, p49).

MERLEAU-PONTY describes the process of reaching a distinct and objective identification of qualities or relations as a "passage from the indeterminate to the determinate" (MERLEAU-PONTY 1962, p24) or as a "passage from indistinctness to clarity" (MERLEAU-PONTY 1962, p27). This transition to an explicit apprehension is exceptional in our everyday life. Here we mostly remain in the pragmatically sufficient sphere of a vaguely structured perceptual consciousness. Only in rare moments our perceptual consciousness has the character of an explicit apprehension of the distinct features of surrounding things.

This position is corroborated by an earlier analysis of Max SCHELER. SCHELER talks of the "value perception", which is active before the higher development of "form perception" takes place. He claims that the more primitive value perception determines the range of what we then perceive distinctly, if we reach the phase of distinct perception at all. Anything that does not lie in the direction of our value perception is neglected and ruled out for perception. He illustrates this "suppressing" force by referring to a common experience. "Thus, when occupied in important affairs, for instance, we may feel a 'pull' (perhaps caused by the facial expression of another person) toward a specific direction in our environment [Umwelt]; but we do *not* follow it, and so this *picture-content* of what is striven for does not come about." (SCHELER 1973, p34). The actual focus of attention suppresses the explicit perception of any elements which do not fit into the scheme of our primary orientation.

MERLEAU-PONTY characterizes the pre-objective structuring as an endowment of meaning. This again can be illustrated by the perception of the red spot. We are strongly inclined to assume the background continuing behind the spot despite the fact that we cannot see that. We do not usually see the background as being interrupted by the spot. The same is true of the fact that our perceptions have expressive qualities. The red spot is felt to be "warm", "cold", "attracting" or "repulsing". These examples show that our perception is full of aspects and assumptions which are not validated by the perceptions themselves. MERLEAU-PONTY calls these aspects which are "read in" to our perceptions "meaning" (MERLEAU-PONTY 1962, p4). In fact our inclination to meaning is the very basis of our perception.

It would be false to regard the pre-objective phase of our perceptual consciousness as disturbing. Quite the contrary, it is at this layer that the direction which determines what we are going to perceive more closely in the first place and how the perceptions are synthesized is fixed. The structuring process does not start with our distinct perceptions but is the very basis of having particular perceptions at all. It is set up in the pre-objective phase of perception. Every explicit perception is "already endowed with a meaning" (MERLEAU-PONTY 1962, p9), which means that every explicit perception is motivated by a disposition to a certain meaningful constellation. "The elementary event is already invested with meaning, and the higher function will bring into being only a more integrated mode of existence or a more valid adaptation, by using and sublimating the subordinate operations." (MERLEAU-PONTY 1962, p10).

MERLEAU-PONTY goes a step further in characterizing the primary orientation towards meaning. The pre-objective orientation, which determines what we perceive and in which way, is closely connected with our bodily constitution. The meaning that we endow our perceptions with depends on the "biological meaning of the situation" (MERLEAU-PONTY 1962, p11). Every perception is penetrated by bodily feelings. This is the reason why we see a wheel lying on the ground differently from a wheel carrying a heavy burden. MERLEAU-PONTY calls the bodily relatedness of our perceptions their "vital value" (MERLEAU-PONTY 1962, p52). At other times he refers to the effects of the elementary imposition of meaning as the "affective value or biological meaning" (MER-

LEAU-PONTY 1962, p123) of perceptions. He thereby intends to make clear that our perceptions are directly related to our emotions and our motor activity, and to our entire bodily constitution. This is due to the fact that our body is not an inert and passive thing but a "momentum of existence". "Existence", a central concept of MERLEAU-PONTY, means the organization of our perceptions by the predispositions of our body. "When we say that an animal *exists*, that it *has* a world, or that it *belongs* to a world, we do not mean that it has a perception or objective consciousness of that world. The situation which unleashes instinctive operations is not entirely articulate and determinate, its total meaning is not possessed, as is adequately shown by the mistakes and the blindness of instinct. It presents only a practical significance; it asks for only bodily recognition; it is experienced as an 'open' situation, and 'requires' the animal's movements, just as the first notes of a melody require a certain kind of resolution, without its being known in itself..." (MERLEAU-PONTY 1962, p78).

In conclusion one can say that perception of distinct qualities (perceptions in the mode of presentational immediacy) is not elementary but a late development and, in fact, exceptional. It is an elaboration of a pre-objective and vague perception in which the higher phases of perception are already structured and directed. This structuring is basically due to our bodily constitution. Hence perception is essentially a pragmatic value perception.

5. Results of the phenomenological investigations

Although originating in different philosophical contexts, the interpretations of human perceptual consciousness outlined above are mutually corroborative. WHITEHEAD's analysis draws attention to the fact that besides the perceptions in the mode of presentational immediacy there are other perceptions which have a directing and motivating power. This can be found in GEHLEN in the context of the ontogenetical development. Finally, MERLEAU-PONTY argues that distinct perceptions do not simply enter our consciousness but are the outcome of earlier processes of structuring which stem from our bodily constitution. MERLEAU-PONTY's theme is the intermediate sphere between the perceptions in the mode of presentational immediacy and perceptions in the mode of causal efficacy. The given depiction gives sufficient evidence for the following account:

1. Our perceptual consciousness is much more encompassing than the sum of our distinct perceptions derived from our sense organs.
2. The perceptual layer at the base of our distinct perceptions is vague and highly indeterminate but it is not totally unstructured.
3. Primitive perception is of great massiveness and emotional force. It motivates the direction of what we then perceive distinctly.
4. The elementary layer of perception has a pragmatic and behavioral significance.
5. The perception of pragmatic meaning has the character of perceiving the pragmatic "values" of things and situations.

The outlined positions are interpretations of human perceptual consciousness.[6] Putting them in the context of the leading question of this essay one can ask (1) whether it can be assumed that the layer of value perception exists in animal perception as well. (2) Moreover, if one is inclined to answer in the affirmative one can go further and ask how much weight this form of perception has in animal perception. One extreme position is that animal perception remains entirely within the bounds of value perception and does not extend to any distinct perception. However, more modest interpretations are also possible. An adequate answer cannot be given without a comprehensive investigation of the evidence from biological research.

One such investigation based both on a survey of ethological data and on a consideration of results from experiments on human perception was carried out by Susanne K. LANGER in the context of her naturalistic philosophy of mind. LANGER undertook an interpretation of general principles of animal perception and animal social behavior in order to fix the dividing line between animal and man. She reaches the position that the elementary form of human perceptual consciousness extends back to animal perception. Indeed, although LANGER does not refer to the authors considered above or any other philosophical theory, they were of major importance for her philosophy.[7]

6. Susanne K. Langer's interpretation of animal perception

In her *Mind: An Essay on Human Feeling*, which aims at a "biologically" based philosophy of mind, LANGER discusses the dividing line between man and animal.[8] Before the field of "cognitive ethology" had emerged LANGER objected to the behavior-

istic prohibition of investigating animal and human consciousness. Despite the complexities and difficulties of interpreting the subjective form of feeling, a comprehensive view of animal intellect and human thinking has to include an analysis of consciousness.

According to LANGER animal perception has to be understood as depending fully on physiological and behavioral conditions. The frame of what an animal perceives is always directly controlled by the intraorganic motivation i.e., by hormonal cycles. Both seasonally changing sensibilities and the direct physiological situation (e.g., hunger) sensitize animals for specific sensory stimuli. Hunger sensitizes perception for everything that can be eaten and the seasonally changing motivations activate dispositions to specific behavior such as nest-building, mating, breeding or migration. "The currently prevailing chemical balance is a general condition underlying the play of perceptions, and as these are essentially emotive perceptions of momentary values the animal's total awareness is governed by the inward pattern of sensitization." (LANGER 1972, p128). This entails a constantly changing evaluation of the environment.

LANGER's central hypothesis here is that the perceptual consciousness of animals is completely integrated in animal behavior. She speaks of the "inherence" of "sensory acts in larger behavioral acts" (LANGER 1972, p59). For animals, perceptions depend directly on current behavior. This total integration in action is illustrated by space perception. Most animals are not able to assess the structure of a location just by looking at it. Carrying a cat around a location to make it familiar with the area will not teach it anything. For the cat, acquaintance is a matter of moving and behaving in the location. The location exists only as a behavioral space, not as abstract knowledge.

Animals do not perceive their environment in its individual appearance of forms and qualities. Perceptions function only in their pragmatic relevance as "values" for action.[9] A cat that is sensitized to chasing behavior chases anything that runs or flutters. The individual perceptual qualities are comparatively insignificant. What counts is its pragmatic value for particular behavioral action. Thus, LANGER maintains that: "...the primary characteristics, which animals see are values, and all the qualities of form, color, shape, sound, warmth, and even smell, by which we would naturally expect them to recognize things, enter into their perceptual acts only as they enter into their overt behavior as values for action." (LANGER 1972, p55). Hence, animals do not perceive distinct qualities in their individual appearance but only in their value character for actual behavior. LANGER, too, uses the value concept in order to characterize the pragmatic structure of animal perception.[10]

LANGER refers to the observations of Hans VOLKELT for further evidence on the structure of animal perception. In his famous experiments with spiders VOLKELT showed that they obviously did not recognize the same thing when it appeared in different behavioral situations. If a fly was put into the web in front of the spider, the spider performed as usual, paralyzing and ensnaring its prey. However, if the fly was put into the rear of the lair the spider showed a difference in behavior and retreated from the fly. So what the spider perceived was evidently not one and the same fly appearing in different situations. VOLKELT concluded that the spider "perceived" totally different situations the elements of which have different values for action. Felix KRÜGER called this relational dependence of perceived elements the "complex-quality".[11] This implies that for animals perceptions have no stabile and invariant meaning. What perception means to them is always dependent on their actual behavior. The value of a perception is its role in a larger complex of factors determined by behavioral orientation. No aspect has an independent meaning. Any minor change in the constellation can lead to a total shift of meaning. LANGER writes: "...animal perception might be normally a matter of locating situations for action, in which a center of highest value draws the agent's interest; that center (for us, the 'object') presumably has sensory properties which the animal recognizes without conceiving them descriptively, i.e., without distinguishing them as shapes, colors, surface feelings or even characteristic smells..." (LANGER 1972, p116). Animal perceptual consciousness is restricted essentially to the pragmatic layer of perception which is at the root of distinct perception. However, the perceptual process originating in that layer is interrupted before it reaches the higher elaboration of distinct qualities as takes place in human perception.

LANGER gives additional weight to her interpretation of animal perception by referring to findings from experiments on human perception. Here, it is a well-established fact that conscious perception follows a phase of unconscious evaluation. Understanding words, for instance, is influenced by a premonition of their positive, negative or inappropriate value. Motorists on the highway recognize more quickly the message "keep right" than contextually

inappropriate words. Other experiments add to the evidence that the value of a perception is recognized even before it is distinctly perceived. "The significance of these findings in human psychology for our necessarily speculative judgments of animal feeling, perception and motivation lies in the objective demonstration that value may be adumbrated before perception of forms is complete; indeed, the expectant, covert anticipation of the full percept appears to be an emotively tinged process, missed in our ordinary introspective analysis because of its minuteness and transitory character." (LANGER 1972, p115). LANGER suggests that this phase of preconscious perception is probably much more dominant and perhaps even more elaborate in animal perception because contrary to man it is not a transitory phase for animals. Perhaps animals never enter the phase of distinct presentational perception at all. They may remain in the more primitive phase of value perception since the more highly elaborated phase is constantly interrupted by the fact that animals only perceive in the framework of a constantly changing behavioral situation. Perception is permanently fixed to a constantly shifting behavioral constellation. This hypothetical interpretation does not suggest that animals do not have very differentiated perceptions. On the contrary, in many respects animal perception is superior to human perception. However, what is meant here is that for animals perception does not arrive at full conscious apprehension of distinct qualities. Animals perceive highly sensitively but do not enter a conscious perception of the qualities of form and color in their individual characteristics (see also GIBSON 1973, pp308–309).

This interpretation of animal perception also suggests that animals are capable of generalization and behavioral flexibility. Since situations never recur in exactly the same way, successful behavior depends on their ability to recognize different situations as "equivalent" to each other. Flight behavior can be brought to an end successfully in a variety of actions of the same pragmatic value: climbing a tree, jumping into the water or running into a hole. Hence, successful behavior may not be so much the outcome of complex and explicit thinking, comparison and judgment but of the spontaneous recognition of the pragmatic values of a behavioral situation.[12]

LANGER's interpretation of animal perception in terms of primitive phases of human perception has far-reaching consequences. If animals do not have distinct perceptions then one also has to deny that they have images, conceptions or a memory in the same sense as humans do. Past and future situations cannot be represented in terms of having images of those situations, as in the case of humans. Then, memory is only possible for animals in the sense of a behavioral familiarity with situations. Past experiences are integrated in the behavioral structure. Memory is nothing but behavioral reproduction—without any accompanying imaginary consciousness.[13]

7. Results

The depicted interpretations served both to (1) clarify a methodological position on how to reach hypotheses on animal consciousness and (2) a particular hypothesis concerning animal perception. If human consciousness includes traces of animal consciousness (which is plausible from the evolutionary constellation (then one can reach a hypothesis about animal consciousness by means of experimental and phenomenological investigations. The most plausible candidates for such hypotheses are the more primitive aspects still found to be operating in human consciousness. This can serve as a starting point for the second step: a corroboration or falsification of the hypothesis by checking it with biological evidence and findings.

The methodological position has been illustrated by dealing with the particular aspect of animal perception. A phenomenological investigation as well as experimental findings suggest that the recognition of value takes place before there is any distinct perception of qualities. It has been suggested that this hypothesis is in accordance with the available evidence on animal behavior and their intellectual abilities. Even if this interpretation is only the beginning and still has to be specified with regard to a particular animal species, it does provide a general staring point for further research with the hypothesis that animal perception is essentially value perception.

Author's address

Rolf Jürgen Lachmann, Lehrstuhl für Philosophische Anthropologie und Kulturphilosophie, UL 6, Raum 2049, Humboldt-Universität Berlin, Unter den Linden 6, D-10099 Berlin, Germany.
Email: rolf.lachmann@t-online.de

Notes

1 For a comprehensive historical account see SPIEGELBERG (1984).

2 WHITEHEAD illustrates this with an example of traffic behavior. We do not wait in front of the traffic lights in a contemplative attitude enjoying the various colors and tones. Instead we look at them in order to get information concerning the forces that determine our behavioral environment and are relevant to the success of our intentions.

3 GEHLEN also stresses a particular human "excessive nature of impulses" which has to be processed. See GEHLEN (1988, p49ff).

4 GEHLEN also gives an account of animal perception which is different in structure from human perception. He stresses that animals have perceptions in their role as indications and triggers of instinctive reactions. See GEHLEN (1988, pp137–157).

5 This ambivalence is typical for the pre-objective perception. See also MERLEAU-PONTY's characterization of this process as a "still 'empty' but already determined intention." MERLEAU-PONTY (1962, p28).

6 An exception is WHITEHEAD who does not restrict his analysis to human perception and includes animal perception as well.

7 This is particularly true of WHITEHEAD and GEHLEN. There are no indications, however, whether LANGER knew MERLEAU-PONTY. But again, phenomenological philosophy played an important part in her thought. Therefore, the sketched positions can be regarded as a "philosophical background" for the development of her position on animal perception.

8 LANGER's theory is not restricted to animal perception. She also discusses the different forms of social behavior (suggestion versus imitation), the contagion of feeling (empathy versus sympathy) and of mutual exchange (communion versus communication) in animal and man.

9 LANGER's interpretation is influenced by GIBSON (1973) who also refers to an elementary form of perception as the perception of pragmatic values.

10 The value concept as it is used here does not include moral values, which according to LANGER are restricted to the human sphere.

11 The concept "Komplexqualität" was introduced by F. KRÜGER and was used in VOLKELT (1914, p84).

12 LANGER's position is corroborated by D. MCFARLAND's position that animal behavior "…is directed, not by goal-representation, but in a holistic manner which takes into account all (relevant) aspects of the animal's internal state, and of the (perceived) external situation." (MCFARLAND 1989, p49).

13 In this regard Langer is in sharp contradiction to the leading assumption of cognitive ethologists. See e.g., GRIFFIN (1991, p5). LANGER's thesis that animals have no explicit memory in terms of a conscious recollection of past events is in accordance with Henri BERGSON's distinction between two forms of memory. BERGSON regards the elementary form of memory as a motor habit or a system of movements. This is different from an explicit memory in terms of an imaginary remembrance. See BERGSON (1991, pp68–80).

References

Bergson, H. (1991) Materie und Gedächtnis. Meiner: Hamburg (Originally appeared in 1896).

Gehlen, A. (1988) Man. His Nature and Place in the World. Columbia University Press: New York (Originally appeared in 1940).

Gibson, J. J. (1973) Die Wahrnehmung der visuellen Welt. Beltz Verlag: Weinheim (Originally appeared in 1950).

Griffin, D. R. (1991) Progress Toward a Cognitive Ethology. In: Ristau, C. A. (ed) Cognitive Ethology. Erlbaum: Hillsdale.

Langer, S. K. (1967) Mind: An Essay on Human Feeling, vol. I. Johns Hopkins Press: Baltimore.

Langer, S. K. (1972) Mind: An Essay on Human Feeling, vol. II. Johns Hopkins Press: Baltimore.

McFarland, D. (1989) Goals, No-Goals and Own Goals. In: Montefiore, A./Noble, D. (eds) Goals, No-Goals and Own Goals. Unwin Hyman: London, pp. 39–57.

Merleau-Ponty, M. (1962) Phenomenology of Perception. Routledge & Kegan Paul: London (Originally appeared in 1945).

Spiegelberg, H. (1984) The Phenomenological Movement. A Historical Introduction. Martinus Nijhoff: The Hague (Originally appeared in 1960).

Scheler, M. (1973) Formalism in Ethics and Non-Formal Ethics of Values. A New Attempt toward the Foundation of an Ethical Personalism. Northwestern University Press: Evanston (Originally appeared in 1913/1916).

Whitehead, A. N. (1985) Symbolism. Its Meaning and Effect. Fordham University Press: New York (Originally appeared in 1927).

Whitehead, A. N. (1967) Adventures of Ideas. The Free Press: New York (Originally appeared in 1933).

Whitehead, A. N. (1968) Modes of Thought. The Free Press: New York (Originally appeared in 1938).

Volkelt, H. (1914) Über die Vorstellungen der Tiere. Engelmann: Leipzig.

Hans Lenk

Humans as Metasymbolic and Superinterpreting Beings

He who understands the apes, does more for metaphysics than LOCKE.
(Charles DARWIN[1])

Abstract

CASSIRER characterized "man" (i.e., the human being) as the symbolic being (animal symbolicum). This is analyzed and criticized according to new research with wild living and trained primates with regard to using tools and establishing local cultures as well as utilizing symbolic gestures and sign languages. The human being therefore cannot be characterized any more just as the symbolic animal. Instead, it can be demarcated from primates by its capability of ascending and transcending to higher meta-levels of interpretation, cognition and language. It is rather the meta-interpreting being (animal meta-symbolicum) and superinterpreting being (trans- and super-interpreting). Yet, this characterization has to be integrated within a rather pluralistic (non-monolithic) practical anthropology including results of many other anthropological disciplines.

Key words

Animal symbolicum, CASSIRER, animal meta-symbolicum, symbol usage, the meta- or supra-interpreting being, toolmaking, primate behavior (research), chimpanzees (Washoe), bonobos (Kanzi), gorillas (Koko).

It WAS THE PHILOSOPHER Ernst CASSIRER who characterized "man", the human being as the "symbolic animal" (*animal symbolicum*), i.e., the being which construes a world of symbols, orients itself in it and has to weave a net of symbols or symbolic net (1944), even living in a new dimension of reality, in "a symbolic universe" (1944, p221). (S)he cannot confront reality directly, but (s)he has to interstice this artificial medium between him or her and reality being basically—by nature—dependent on *symbolic* means and capacities of representation (i.e., on *culture*).

Already in antiquity, POSEIDONIUS stressed the idea that "man" is by nature necessarily and indispensably dependent on culture (REINHARDT 1953; POEHLMANN 1970). POSEIDONIUS would as probably the first author use the expression 'second nature' as a central concept of anthropology and of his doctrine of the origin of culture. HERDER used this concept later on in his theses of humans as beings characterized by deficiencies, insufficiencies and incompleteness (by the way a contention already implicitly seen by PLATO) which had been used for the development of language and culture. In modern philosophical anthropology, Helmuth PLESSNER (1975, p301) developed the thesis of "the second nature", i.e., culture, as highlighting the constitutive "eccentricity" and homelessness of humans by the dependence on this kind of second nature and the original "homelessness of man" making cultural development necessary. "Man" would even be "the apostate, defector of nature, the troublemaker, attention-seeker, achieving being" with all respective consequences and phenomena of a "tendency of self-aggrandizements of life in terms of the power instinct" (PLESSNER 1975, p320). The consequence of the basic interpretation of the human as the constitutively eccentric being (locating itself besides its own center) which fundamentally has to distance itself from itself only secondarily working towards the unity of life under drafts and anticipations, which has to plan and to "lead" its life: "man as an eccentrically organized being he has only to make himself that what he already is", by not merely executing or performing his life according to the "*law of natural artificiality*", in an "*artificial* form of existence", but also by "leading" his life (PLESSNER 1975, p319). GEHLEN not only took up the thesis of the "deficient being" ("Mängelwesen") from HERDER, but also this thesis of the "second nature of man" from PLESSNER: "*exactly at that locus*, where for an animal the 'environment' ('Umwelt') figures, stands as regards man the

cultural world ('Kulturwelt'), i.e., the sector of nature superwhelmed by him and creatively altered by him to become supporting means of life" (GEHLEN 1962, p38). GEHLEN also develops this insight within the framework of his philosophical doctrine of institutions interpreting the function and the central nature of institutions in the respective relieving import and in the functions of mediating as well as constituting and constructing artificial world structures and norms, values etc. Culture is understood as second nature. The "background" or primary nature in GEHLEN takes up the role as a quasi KANTIAN *a priori* behind the only knowledgeable and available "second nature" shaped by cultural versions of the world (i.e., by world representations, *or world versions*). Cultural means and sorts of representation (media of grasping and representing) are of course *symbolic*: only by symbolic representation we can and may grasp, constitute and structure our world versions; the second nature is necessarily a "symbolic universe", it is spanned by CASSIRER's "symbol net" ("Symbolnetz"). The cultural world is the symbolic or symbolically shaped world (version).

CASSIRER cites the traditional sign-theory after PEIRCE, MORRIS (1938) and S. K. LANGER (1942) according to which philosophy all symbols are not only vicarious substitutes for objects as a traditional interpretation of PEIRCE's triadic sign theory, but a "vehicle for the grasping of objects" in general serving to grasp, symbolize and interpret a thing or situation, beyond the mere behavioristic reaction or response: it is the "conception", not the things, what symbols directly 'mean' (LANGER 1982, pp60f)[2]. In this approach CASSIRER's theses are anticipated—at least in some sense. CASSIRER however uses this insight for the foundation of an enlarged symbolistic philosophical methodology by characterizing man as the only being which by "a *symbolic imagination and intelligence*" (1944, p33): therefore "we instead of defining man as an *animal rationale*, we should define him as an *animal symbolicum*" (1944, p26). He relates his symbolistic theory also to Charles MORRIS's elaborating and differentiating PEIRCE's semiotics or sign-theoretic approach (1944, pp31ff)[3]. Intrinsically, characteristic for the principle of this "symbolic" is among others "*universality*", the general "validity and general applicability", "the magic world, the Open Sesame!" which allows an entrance to the specifically human world, "to the world of human culture" and "by architectural structure" (1944, p36), the "general symbolic *function*" of symbols and language as well as multisidedness and variability: it is also only a highly "complex system of symbols" which renders rational and relational thinking, human grasping, reflection and culture possible, since "without a complex system of symbols relational thought cannot arise at all, much less reach its full development" (1944, p38).

Moreover, CASSIRER refers to HERDER's theory of "reflexive thinking" with regard to the origin of human language and culture. All these fundamental capacities and explanatory factors depend on the ability and function of using symbols so masterly and intriguingly analyzed by CASSIRER in his main work *The Philosophy of Symbolic Forms* (1923). Any grasping and "objectification", even representations of nature states are only to be covered by "constructive process(es)", dependent on "symbolic constructs" (CASSIRER 1944, pp160, 62). Even "all classification systems are artificial. Nature as such does contain only individual, multifarious phenomena. By abstracting and summarizing as well as conceiving ('zusammenfassen') these phenomena under concepts of sets and general laws, we do not describe natural facts. "Every system is a work of art" ('Kunstwerk')—a result of conscious creative activity" (1944, p209). "Theoretical" work always means also "constructive work": "spontaneity and productivity is the very center of all human activities. It is man's highest power and it designates at the same time the natural boundary of our human world. In language, in religion, in art, in science, man can do no more than to build up his own universe—a *symbolic universe*, that enables him to understand and interpret, to articulate and organize, to synthesize and universalize his human experience" (1944, pp220f, my italics). Therefore, it is indeed the human and according to CASSIRER (s)he "alone" (1944, p33) who is the symbolic being, the "*animal symbolicum*" (1944, p26). This symbolic being needs a "symbolic universe", into which it has to perform its projections, in which it would "live" (in a secondary meaning) and by which "the symbolic net" is extended (1944, p25).

CASSIRER (cf. e.g., 1944, pp28, 30ff. a. e.) wanted to substantiate and support his philosophic anthropology of the symbolic being by a comparison with the most modern behavior research with animals, especially chimpanzees (Yerkes) attributing to these primates the ability to use signs to utilize "designators" but indeed not the capability of a symbolic interpretation (with universality, functionality, variability and general language functions).

This certainly was plausible at the time (1944) when CASSIRER wrote that, but this does not hold any more nowadays, for in the meantime research on

primates is much more advanced. For instance, one taught chimpanzees like Washoe and gorillas like Koko to communicate (the latter one with more than 1000 "words", "concepts" or "meaningful gestures") in American Sign Language (A. and E. GARDNER and Francine PATTERSON). These primates are able by using a functional quasi symbolic combination of meaningful signs and sentence parts to combine simple sentences up to six words or so or even add and subtract small numbers (e.g., Sheba, after S. BOYSON). There occurs a sort of generalizations, generalizations and self-reflective utterances (designation of the sign-talking chimpanzee as a subject herself or himself by a symbolically represented expression). In addition, besides the gestures language plastic chips or symbolic notations on the computer display in a computer game are used in some kind of combination game like "Scrabble" with chimpanzees and bonobos as RUMBAUGH and SAVAGE-RUMBAUGH successfully experimented with the chimpanzees Austin and Sherman as well as with the real primate genius, the bonobo Kanzi, and his sister Panbenisha (SAVAGE-RUMBAUGH et al. 1998). Washoe, Koko and Kanzi not only occasionally played the fool, but also made a fool of the respective interviewing partner or experimenter. For instance, Koko, the gorilla, communicated with American Sign Language the wrong answer "red" of a white handkerchief, laughing and necking the experimenter PATTERSON. The latter one was annoyed and wanted to give up already. Koko then picked all of a sudden a very tiny red thread from the handkerchief and signalled triumphally laughing "Red, red, red!".

Chimpanzees and bonobos are not only able to use and even produce tools (e.g., to put together bamboo sticks to get at bananas in the distance as W. KOEHLER had already earlier substantiated in his famous primate experiments), but they are also able creatively to apply tools to a respective situation *type* and single situations. They are also capable of using in a limited sense at least means of representation to *transfer* and in a sense *generalize* them referred to *other* situations. They are able consciously to make a fool of somebody (see above), to betray or deceive partners and rivals (DE WAAL 1983, 1990; SOMMER 1992). They even developed an elementary "culture" by establishing a certain tradition of learned and discovered uses of instruments and tools. Famous is the macaque "girl" Imo in Japan which had discovered that potatoes would taste better after they had been washed in salt water. Imo's washing activity soon spread out over the whole colony of the respective group of macaques and even to a neighboring one. BOESCH (1991) and BOESCH/BOESCH (1990; 1991) observed chimpanzee mothers carrying instruments as, e.g., a stone which they used to open nuts (19 types of tool uses and 6 ways of making tools were recorded with chimps in Taï National Park, Ivory Coast). The opening was taught to the youngsters which had for quite a while really to learn and practice a certain kind of skill to be able to crack a hard nut. The mother occasionally hid such a tool and carried it along—defending and/or reclaiming it even as some kind of "private property"—when migrating to another feeding place. This seems to be local "culture"—because there was also a spreading of this utilization of tools carrying in some sense symbolic functional meaning (including a certain limited generalization). Whereas in Senegal wild chimpanzees also used sticks and branches to dig for termites (Wrangham), respective competitive groups of chimpanzees in Uganda did not develop this "cultural" utilization of a sophisticated tool, although they had also termite heaps available. SUGIYAMA (1981) in Cameroon even observed chimpanzees *using tools to make tools* with several functions: like using a sharp cutting stone or so to make a point for boring at one end of the twig and to hammer with a stone the other end to form a certain kind of brush to retrieve the termites by putting this end into the hole bored before (cf. BECKER 1993, p102). Therefore, primates not only use tools, but they also understand a certain kind of general function of tools and they even use tools to make tools, thereby falsifying FRANKLIN's thesis and definition of man as the (only) toolmaking animal.

Primates are therefore able to generalize and again individualize functions of learned utilizations of tools and also to specify individualized meanings of symbols and gestures by establishing little local-cultural traditions etc. They even discover or invent new artifacts or activities which in a limited sense are spread within the local colony or even neighboring groups. That is to say, the utilization of tools and symbols is not only characteristic for man, but also in a limited sense to be found in wild living primates. (To be sure, primates like chimpanzees or bonobos in zoos or experimental situations are much more extensively taught tools and symbols, but this is certainly an artificially arranged situation, which is not so telling here in the first place, although it drastically may refine and extend the use of symbols and tools extraordinarily (see FOUTS-MILLS 1997; SAVAGE-RUMBAUGH et al. 1994; 1998). By the way, some analyses of animal languages like the one of dolphins studied by LILLY, PRYOR and HERMAN or pairwise in-

dividualized speech or singing recognition patterns with birds seem to substantiate the use of communication gestures and symbols also with other species (see e.g., GRIFFIN 1984).

That means that higher animals like primates and some others engender signs which they can in a limited sense and with respect to classified situations use in a symbolic way or even generalize in some (though rather restricted) sense. They use, to a certain degree, symbols. Therefore, they are also within some limits "symbolic beings" or in a at least capable to understand sign combinations as symbols or symbolic way and to use them more or less systematically for the control of behavior and reaction. This is especially true for the highest primates like chimpanzees, gorillas and the bonobos as has been extensively substantiated and confirmed by observation of wild living primate groups or experiments with trained primates (see e.g., SOMMER 1992; DE WAAL 1983, 1998;GRIFFIN 1984; FOUTS-MILLS 1998; SAVAGE-RUMBAUGH 1998; for tool use cf. BECKER 1993). The activities imply certain symbolic representations for grasping their functional "meaning" and manipulating as well as controlling behavior—even including self-reference, reflexivity, and (sign) reciprocity (at least with chimpanzees, bonobos and orangutans)[4, 5]. That is to say, CASSIRER's restriction of the use of symbols and the capacity to establish and generalize as well as individualize symbolic articulations, representations and networks as being unique for humans and his attempt to characterize "man" by this symbolic ability or function turns out to be too rash, if not wrong. This thesis has to be modified.

Higher animals like primates indeed use signs and objects in a symbolic or ritual manner and they may also in a limited sense perform transfers or generalizations and some kind of inferential discrimination as well as simple logical inferences like exclusions or even contrapositions, but they certainly do *not* again analyze these symbols or symbolically "understood" functioning signs as *objects* of a symbolic analysis on a higher level (metalevel). This is particularly true for the gestures and signs as well as activities which are not genetically fixed but "culturally" developed, so to speak. They do not make the symbols and the utilization of them including the rules of symbol use and symbolic functions again an object of a higher (meta)-level symbolic representation. They do not analyze and symbolize the application of signs *per se* again. This seems to be a characteristic trait for the human being that (s)he is able to not only apply symbols on a higher metalevel but also to designate, "mean" and interpret the utilization and function of symbols in a higher-level representative form and analysis. Only man or woman is the being who is able to ascend to higher symbolic levels in a rather unlimited manner. To my mind, the human being is the only being who may—without restriction—establish levels and metalevels of symbols and signs to symbolize and refer to lower-level symbols and represented objects (again in the form of a kind of object representation).(S)he is able to perform a sort of "symbolic ascent" to abstract and go to higher-levels, to vary, again "objectify", i.e., symbolically and abstractly designate lower-level activities and symbolizations in the form of new, higher-level symbolization processes and activities. Procedures of checking, evaluating, controlling and planning action in the sense of an anticipatory and situation transgressing "anticipation" towards future behavioral and action patterns are necessarily represented or incorporated by symbols in processes of symbolization (symbolic representation) within cultural activities of structuring and interpreting norms and signs. To make it short: the human being is not only the sign using symbolic being inventing, fashioning, varying and utilizing symbols and representing by symbols, but the being which projects, establishes and changes symbols as the object of higher-level symbolization processes, i.e., interprets these construed artificial "objects" on a metalevel. Man is not only the being who uses symbolic forms, but the one reflecting and projecting as well as varying symbols about and over symbolic signs and symbols as well as symbolization processes. The human being therefore is not only the *animal symbolicum* but the *animal meta-symbolicum*, being the being which not only interprets, after NIETZSCHE, but who in turn interprets its interpretations and interpretation processes as well as the respective symbolizations and representations as higher-level and abstract "objects" and again analyzes these, attaches meanings to these etc. This creature called human is the being capable of not only creating and using symbols and metasymbols as well as interpretations and metainterpretations—(s)he is the *animal super(inter-)pretans et symbola symbolisque creans*. (S)he transgresses the realm of mere sign utilization and symbolic functions by trans- and superinterpreting, symbolically transcending these forms and usages in a reflexive manner. (S)he is the being potentially able to transgress *any* levels of representation towards ever-accumulating higher metalevels varying, projecting, reflecting or rejecting "meaning", schematizing and structuring as well as ordering—in short: interpreting—representations of the (mostly proximate) next-lower level. *The human being is the*

being of meta-levels. (S)he does not only go beyond certain sets and extensions of sets within one and the same level of designation and denotation but potentially and tendentiously is able to transcend beyond different meta-levels of language by higher-level symbolization and interpretation. If we refer to the transgressing of sets within one and the same level by the Latin syllable 'trans' and to the ascending to higher-levels by 'super' or 'supra' respectively, we can call the human being the *trans-interpreting and*/or *super- or supra-interpreting* being, respectively, or for short: *the meta-interpreting being* (see LENK 1995; 1998, pp38ff). The human being does not only interpret within one level of interpretation and sign utilization, (s)he is not only the interpreting being restricted to just one symbolic plane, but the level-transgressing interpreting being and by that in turn really the reflecting being. Abstract reflections are only possible if you are able to transcend the actual level at hand, if you transgress the levels by going meta-symbolic or superinterpreting. Therefore, it is most plausible to talk of the human being as the "*animal metasymbolicum*" (a sort of extension of CASSIRER's terminology) or as the *superinterpreting being* (in extension of NIETZSCHE's conception of the interpreting being). (We could conceive of superinterpreting as including trans- and supra-interpreting, i.e., not only to another set or denoting symbol within the same level of denotation, but also to higher-level representations.)

Meta-levels and even methodological interpretations or analyses of reflective and interpretative levels (so to speak meta-levels of interpretations) can well be coordinated with or even integrated within this approach. The interpretative levels of interpretation using higher-level symbols for representation is certainly open with respect to the ever-transgressing and ascending leveling of symbolizations upwards. One may again in a cumulative combination so to speak incorporate these higher-level interpretations in the talk of the "interpreting being" and stay with NIETZSCHE's terminology, but then one has explicitly to add that the capacity of interpretation in humans is not restricted to just one level of symbolization. NIETZSCHE and CASSIRER certainly would have underlined this capacity and would have included the higher-level symbolic functions and utilizations in their approaches. However, to avoid terminological misunderstandings and difficulties of level interpretation etc., it seems to be better to talk of "animal metasymbolicum or the transinterpreting and suprainterpreting being" or "superinterpreting" (combining trans- and supra-interpreting) being in general.

Herewith, it is certainly not hypothesized or explicitly contended that this anthropological feature would be the *only* one: earlier (LENK 1983; 1986; 1989) I tried to develop a practice-oriented philosophical anthropology of rather pluralistic provenience, based not on just one unique and only anthropological factor, but a multifactorial and multifunctional philosophical anthropology taking into consideration many results of empirical anthropological disciplines. A philosophical anthropology cannot any more merely highlight one unique trait in a monolithical manner or just rely on constitutive self-production and artificial self-identification of the human being in a traditional idealistical–"subjective" or transcendent or a reified transcendental understanding, but it has to take into account many and different functions of the so-called "second nature" or "symbol world" (including within the symbolic net higher- and lower-level symbols in a rather architectonic order of meta-level character). The capability of transcending and ascending to ever higher-levels and meta-levels of interpretation may be a unique sign of delineation and demarcation of the human being from primates well-suited to demarcate the higher(-level) human culture as a realm of meta-symbols and meta-interpretations, but this does not stand out as the only and unique function for the constitution, development, and foundation of our "second nature" (i.e., "culture"). This feature of meta-interpretation is to be sure central and indispensable, essential and unique for man, but it is certainly not the only characteristic feature of man.

To summarize again: One may understand and demarcate, or distinguish, man from higher animals—including especially primates—by identifying and characterizing the human being as the *animal metasymbolicum* or as the meta-representative (meta-representing) or meta-interpreting or trans- and supra-interpreting ("super-interpreting") being. It or (s)he is the meta-level being of representations and interpretations—as well as the being of meta-cognitions and meta-actions—in short: *the meta-level being*.

Author's address

Hans Lenk, Universität Karlsruhe Institut für Philosophie, Geb.Nr. 20.12; Bau II am Schloß, Postfach 6980, D-76128 Karlsruhe, Germany. Email: ed05@rz.uni-karlsruhe.de

Notes

1 *Notebook*, my re-translation.

2 LANGER replaced the traditional semiotic triadic theory of signs by a tetradic one in which beside the symbol, the subject and the object indeed also "conception" figures as an important instance (1982, p64). She distinguished strictly between "sign meaning", and "symbol meaning" the latter one rather referring to "connotations", namely to an expression of graspings of meaning instead of denotations or objects referred to. Indeed, according to her there is no "real denotation" whatsoever in the symbolic process of grasping, but always only "connotation-mediated meaning" (1982, pp65f).

3 CASSIRER (1944, p32) conceives, after MORRIS, of "signals" as "operators, symbols are 'designators'". Signals, even when understood and used as such, have nevertheless a sort of physical or substantial being; symbols have only a functional value".

4 For the function of "property" and the "principle of property" among longtail macaques and chimpanzees see KUMMER (1982).

5 We know that chimpanzees and orangutans (as well as bonobos) are indeed able to recognize themselves on-time in front of a mirror. They even use a mirror in a goal-oriented manner, "by intention" so to speak, as an aid to clean, check and inspect hidden body parts (after GALLUP in STRAUSS/GOTHAL 1991/92). However, gorillas do not recognize themselves individually in a mirror image.
Lately, however, experiments with video movies moderated this interpretation: Only in *directly* concomitant replays of videos chimpanzees could recognize themselves—not if the replaying was postponed by even a relatively short time (cf. Frankfurter Allgemeine Zeitung, April 7, 1999). Like with human children (are only) under the age of three it seems that chimps probably lack a sense of self-continuity or of (autobiographical) self-recognition within a movie representation over time: They apparently cannot individually reidentify themselves on the movie-screen if the video is replayed a while after it had been taken whereas they seemed to be able skillfully to do the reidentification on-time with the filming and simultaneous replaying on a screen. The experimenter D. J. POVINELLI (in FERRARI 1998) concludes that the chimpanzees would not infer "that's me" but possibly only "that's like me". Again, the upshot seems to be (like SAVAGE-RUMBAUGH's result) that these (adult) primates (chimpanzees and bonobos) are stuck at a developmental stage compared to a human child at 2.5 or 3 years.

Bibliography

Becker, P.-R. (1993) Werkzeuggebrauch im Tierreich. Hirzel: Stuttgart.

Boesch, C. (1991) Teaching among wild chimpanzees. Animal Behaviour 41: 530–532.

Boesch, C./Boesch, H. (1990) Tool Use and Tool Making in Wild Chimpanzees. Folia Primatol 54: 86–99.

Boesch, C./Boesch, H. (1991) Dim Forest, Bright Chimps. Natural History 9, Vol. 100: 50–56.

Byrne, R. (1995) The Thinking Ape. Evolutionary Origins of Intelligence. Oxford University Press: Oxford, New York.

Cassirer, E. (1923) Philosophy of Symbolic Forms. 3 vols. Yale University Press: New Haven (quoted after German translation, Philosophie der symbolischen Formen 1956, 2nd ed, 1977, 7th ed (II), 1964, 4th ed (III)).

Cassirer, E. (1944) Essay on Man. Yale University Press: New Haven. (2nd German translation: Frankfurt 1990).

Cheney, D. L./Seyfarth, R. M. (1990) How Monkeys see the World. Inside the Mind of Another Species. Chicago University Press: Chicago, London.

De Waal, F. (1983) Unsere haarigen Vettern. Neueste Erfahrungen mit Schimpansen. Harnack: München.

De Waal, F. (1990) Peacemaking among Primates. Harvard University Press: Cambridge MA.

De Waal, F. (1995) Die Bonobos und ihre weiblich bestimmte Gemeinschaft. In: Spektrum der Wissenschaft, Mai: 76–83.

De Waal, F. (1998) Die Bonobos. Birkhäuser: Basel.

Ferrari, M. (ed) (1998) Self-awareness. Guilford: New York.

Fouts, R./Mills, S. T. (1997) Next of Kin: What Chimpanzees Have Taught Me About Who We Are. Morrow: New York.

Gehlen, A. (1962) Der Mensch: Seine Natur und seine Stellung in der Welt. Athenäum Bonn (Originally published 1940).

Griffin, G. R. (1984) Animal Thinking. Harvard University Press: Cambridge MA.

Kummer, H. (1982) Social knowledge in free-ranging primates. In: Griffin D. (ed) Animal Mind—Human Mind. Springer: Berlin, pp. 113–130.

Langer, S. K. (1942) Philosophy in an New Key. A Study in the Symbolism of Reason, Rite and Art. Harvard University Press: Cambridge MA (quoted from 19823 ed.).

Lenk, H. (1983) Wie philosophisch ist die Anthropologie? In: Frey, G./Zelger, J. (eds) Der Mensch und die Wissenschaft vom Menschen. Bd. 1. Solaris: Innsbruck, pp. 145–187 (reprint in Lenk 1986, pp. 131ff, and partly in Lenk 1987, pp. 112ff).

Lenk, H. (1986) Zwischen Wissenschaftstheorie und Sozialwissenschaft. Suhrkamp: Frankfurt.

Lenk, H. (1987) Zwischen Sozialpsychologie und Sozialphilosophie. Suhrkamp: Frankfurt.

Lenk, H. (1989) Anthropologie nach vorne: Bausteine der philosophischen Anthropologie als einer Interpretations- und Konstruktdisziplin. In: Bonk, S./Lazzari, A. (eds) Ideen zu einer integralen Anthropologie. Festschrift Mácha. Munich, pp. 89–100.

Morris, C. W. (1938) Foundations of the Theory of Signs. University of Chicago Press: Chicago.

Plessner, H. (1928) Die Stufen des Organischen und der Mensch. Berlin, de Gruyter: New York. (quoted from the 1975 ed).

Poehlmann, E. (1970) Der Mensch—das Mängelwesen? Archiv für Kulturgeschichte 52: 297ff.

Reinhardt, K. (1953) Poseidonius. In: Real-Enzyklopädie der Altertumswissenschaften, vol. 22, pp. 19ff.

Savage-Rumbaugh, S./Lewin, R. (1994) Kanzi: The Ape at the Brink of the Human Mind. Wiley: New York.

Savage-Rumbaugh, E. S./Shanker, S. G./Taylor, T. J. (1998) Apes, Language and the Human Mind. Oxford University Press: New York, Oxford.

Sommer, V. (1992) Lob der Lüge. Täuschung und Selbstbetrug bei Tier und Mensch. Beck: Munich.

Strauss, J./Gothals, G. R. (1991/92) The Self. Springer: Heidelberg, New York.

Sugiyama, Y. (1981) The Brush-Stick of Chimpanzees found in South-West Cameroon and their Cultural Characteristics. Primates 26 (quoted after Becker 1993).

Theodore R. Schatzki

Humanism on the Ropes?

Abstract

This essay examines what of humanism remains when one incisive "posthumanist" critique of the uniqueness of human beings is digested. This critique, issuing from a stream of contemporary thought that can be labeled "social multiplicity theory," proliferates agency while also dissolving the human variety. Part one describes the social multiplicity theories of Michel FOUCAULT, Gilles DELEUZE and Félix GUATTARI, and Bruno LATOUR and Michel CALLON. Parts two and three examine two of their critiques of humanism. One denies the uniqueness of humans as agents by analyzing actors as anything to which doing is attributed. The second analyzes human agents themselves as multiplicities of subagents. I contend that the linguistic multiplication of agency does not challenge the distinctiveness of the human sort, and that the claim about subagencies does not undermine the integrity of human agency. The conclusion upholds the viability of a humanism sensitive to the insights of multiplicity theory.

Key words

Humanism, agency, human agency, types of agents, doing, social multiplicity.

The two of us wrote Anti-Oedipus *together. Since each of us was several, there was already quite a crowd.* (DELEUZE/GUATTARI 1987, p3)

THIS MANNER OF DISCOURSE is familiarly plausible. At least since FREUD it is common stock that conscious agents do not fully master the whys and wherefores of their activity. Some aspects of the psyche escape self-monitoring and can even oppose consciously hewn designs. Many thinkers today avow, as a result, a multiplication of the agencies that bear on human conduct.

At the same time, the sentences, however humorous, do not seem quite plausible. Anyone who has tried knows the tribulations of coauthorship. If two can be trouble, a crowd poses a real problem. If there actually had been a crowd, *Anti-Oedipus* probably would not have been written. So there really *wasn't* a crowd. And we don't think twice before thanking just two individuals for their efforts.

Yet, would it be right to assume that the only agency responsible for the book is the conscious intentional agency of its two named authors? Many contemporary writers assert that myriad other entities are actors, among them nonhumans, parts of humans, and collections of humans. Denying this, it is said, is anthropocentric. Worse, it is humanism, that roundly denounced modern ideology that treats human beings as something unique and special.

There is something, as indicated, to this injunction. Today we know too much about people to leave the free agent of Western lore unimpeached as either a definitive portrait of human being or the unique locus of the do-ing called "agency". But does abandoning the majesty and uniqueness of humans entail the leveling of entities and the evisceration of a distinctly human manner of shaping the world? Can there be a chastened and modest humanism that retains a distinctive place for humans in the cosmos instead of centering the cosmos around human beings?

The current essay urges a positive answer. It interrogates three mutually-reinforcing contemporary "posthumanist" accounts of agency, which share above all one crucial characteristic: They are social theoretical harbingers and practitioners of an incisive metaphysics of multiplicity that discerns dispersion where sameness was once assumed. The ways, consequently, they multiply agency and dissolve the human variety are part and parcel of much larger and portentous pictures. As a result, the subtext of this essay is, whether an ontology of multiplicity tolerates an updated and clearer-sighted humanism.

1. Social ontological multiplicity

One significant development in 20th-century philosophy has been a challenge to the age-old domination of sameness and unity. In a variety of guises,

philosophers have questioned whether the entities with which we deal, the worlds in which we live, and we ourselves including our actions, thoughts, and concepts are marked by well-defined and stable essences, substances, meanings, and space-times. Such long-standing Greek figures as essence and accident, species and genre, the one and the many, and form and matter have suffered withering assaults; assaults challenging, *inter alia*, their reductionist penchant, the well-definedness and stability of the entities they invoke, their adequacy to the diversity of things, and their capacities to capture becoming and to do justice to the particulars subsumed under them. Perhaps the most renowned and influential instance of such critique is Ludwig WITTGENSTEIN's conception of family resemblances: Instances of any natural language concept that is not introduced through explicit definition fail to share any specific property—a unifying essence—and are instead linked by an array of crisscrossing similarities that yield overall "family resemblances" among them. Where traditional philosophy sees sameness and continuity, WITTGENSTEIN sees multiplicity. Instances of a concept form a *dispersion*, a distribution of similar but ultimately disparate phenomena that do not as a group share anything in common.

In the 1960s and 70s, such thinkers as Jacques DERRIDA, Michel FOUCAULT, and Thomas KUHN offered analyses of concepts and meaning parallel to or supportive of WITTGENSTEIN's. In the meantime, moreover, a growing line of thought has expanded the notion of dispersion into a potent social ontology. FOUCAULT's *The Archaeology of Knowledge* was ground-breaking in this regard. The book's leading question runs, What is the being of a discourse? In pursuit of an answer, FOUCAULT begins by provisionally accepting such conventionally designated discourses as those of natural history, political economy, and medicine. He rejects, however, the phenomena traditionally countenanced as unifying and demarcating these discourses and instead pursues a radical strategy. Much as HUSSERL in his epoché places the reality of the external world at bay, FOUCAULT suspends the assumption that discourses are unified wholes. He dissolves their alleged commonalities and continuities, erases the boundaries that seemly demarcate them, and resolves discourses as a set into a field of naked discursive events: a space of statements *per se* devoid of relations, principles, or contexts to connect, group, or explicate them.

"Once these immediate forms of continuity are suspended, an entire field is set free… [It] is made up of the totality of all effective statements … in their dispersion as events and in the occurrence that is proper to them. Before approaching, with any degree of certainty, a science, or novels, or political speeches, or the *oeuvre* of an author, or even a single book, the material with which one is dealing is, in its raw, neutral state, a population of events in the space of discourse in general" (FOUCAULT 1972, pp26–27).

On this gambit, the being of a discourse lies in whatever connects dispersed statements either into the discourses from which FOUCAULT begins or into hitherto unrecognized unities. FOUCAULT explains that what he once thought effected unification cannot do so, namely, relations among or the objects, concepts, theories (etc.) referred to and utilized in statements. For the relations, objects, and concepts tied to the statements of any conventionally recognized discourse form irreducible dispersions. Just as on WITTGENSTEIN's analysis the instances of a concept do not share a common property, so, too, are discourses not demarcated by common objects and concepts. FOUCAULT (1972, p37) then wonders whether these dispersions themselves might hold the key to a solution:

"Hence the idea of describing these dispersions themselves; of discovering whether, between these elements, … one cannot discern a regularity: an order in their successive appearance, correlations in their simultaneity, assignable positions in a common space, a reciprocal functioning, linked and hierarchized transformations."

The being of a discourse is tied to regularities in its objects and concepts; more precisely, it is tied to the rules, defined by FOUCAULT as the conditions to which dispersions submit, that govern these regularities. A discourse, consequently, is a multiplicity of statements whose objects, concepts, relations etc. form rule-governed dispersions. It is, in short, a constellation of dispersions: of statements, objects, concepts, theories, relations, subject positions, and rules.

A second landmark in the development of what can be called "social multiplicity theory" is DELEUZE and GUATTARI's *A Thousand Plateaus*. This work proclaims multiplicity the central ontological concept. It is also the key phenomenon opaque to "arborescent" thought, that dominant form of thought that comprehends a set of phenomena by relating it to one or more common and often underlying sources, essences, elements, or reference points (DELEUZE/GUATTARI 1987, p4) In misconstruing dispersion by reconstructing it as a ramification of a common essence or source, arborescent thinking reduces it to

variation of the same. For DELEUZE and GUATTARI's "rhizomatic" thought, by contrast, multiplicity is a substantive, indeed THE substantive; it is not an adjective of unity.

"…this substantive … was created precisely in order to escape dialectics, to succeed in conceiving the multiple in the pure state, to cease treating it as a numerical fragment of a lost Unity or Totality or as the organic element of a Unity or Totality yet to come…" (DELEUZE/GUATTARI 1987, p32)

Instead of illusorily seeking the unity of dispersion, rhizomatic thinking maps its dimensions. Unities exist for it only as either the singularity of words that designate dispersed phenomena (à la WITTGENSTEIN) or the singularity of self-conscious subjects who stand over against the world:

"The notion of unity appears only when there is a power takeover in the multiplicity by the signifier or a corresponding subjectification proceeding… Unity always operates in an empty dimension supplementary to that of the system considered (overcoding)." (DELEUZE/GUATTARI 1987, p8)

In contrast to wholes or totalities, whose unity is an additament to the things unified, multiplicities have no dimensions over and beyond their lines (components); their components fully fill, or occupy, their dimensions by way of constituting them. Multiplicities, consequently, cannot be treated as emanations, versions, or determinations of some common point, source, or entity (overcoding). They are acentered and irreducible distributions.

This conception of multiplicity informs DELEUZE and GUATTARI's general metaphysics. A central feature of reality on their view is pure process-becoming multiplicity made up of continuums of intensities, conjunctions of flows, and emissions of "particles–signs" (in their terms: the plane of consistency, or the body without organs). Out of this plenum coalesce various strata, on each of which these dynamic dispersions receive determination in a "double articulation" of contents and expressions. On the anthropomorphic stratum, contents are things done and expressions things said. On any stratum, moreover, the articulation of contents and expressions is effected in and through multiple assemblages, each of which (1) combines mutually presupposing contents and expressions, (2) occupies a territory, and (3) contains "lines of flight," which are phenomena already fleeing and thus breaking down the assemblage. On the anthropomorphic stratum, assemblages are social formations: configurations of (a) power regimes (organizations of activities and bodies–things) and (b) regimes of enunciation (organizations of statements) that occupy a territory and embrace individuals, groups, or things that flee its orders and thereby threaten it with disintegration. Any social formation is thus a multiplicity of multiplicities: of contents, expressions, vanishing points, and lines of segmentation that define territoriality.

A third key conception of social multiplicity is the theory of the actor-network principally developed by Michel CALLON and Bruno LATOUR. This theory contends that the primary medium of a person's life is a multileveled and interconnected array of networks, each an association of heterogeneous types of phenomena. A student, for instance, is embedded in networks that embrace professors, classrooms, examinations, bicycles, apartments, pets, desires, hairstyles, sleepless nights, fast food, peer pressure, and the like. The network array of this sort in which a student is suspended defines his or her situations, identities, allies, and possibilities. Furthermore, such traditional social theoretic phenomena as power, organizations, and society are mere side effects of networks and do not contribute to the dynamic reality of network building and dissolution. Processes peculiar to networks are instead what is responsible for their evolution, for example, association and black boxing (treating networks as units).

Networks comprise two primary sorts of multiplicity. First, the types of material out of which they are composed are heterogeneous. In opposition to practically all social thought, actor-network theory refuses to identify basic elements of social life. Just as multiplicities for DELEUZE and GUATTARI, in occupying all their dimensions, proscribe reductive overcoding of their components, so, too, is for actor-network theory the multiplicity of network materials a positive reality irreducible to a list of primary stuff. Taking a vacation at a distant lake, for instance, can involve a network combining computers, alloys, jet engines, research departments, market studies, advertisements, welcoming hostesses, natives, fishing boats, luggage, and so on interminably (see CALLON 1991, p139). Attempts to reduce such burgeoning lists to categorical dimensions falsify that as which entities enter and function in networks.

Actor-network theory further maintains that each entity in a network is itself a multiplicity that is viewed as a unit by other network individuals and entities. A computer, a research department, an advertisement, and a welcoming hostess, for instance, are all multiplicities that are taken to be single entities by the users, travellers, readers, bosses, and employees who network with them. Only on such oc-

casions as breakdown, deliberate scrutiny, or change in interest do the multiplicities involved come to attention and thereby become network components in their own right. According to actor-network theory, consequently, networks embrace heterogeneous materials that are themselves networks of entities.

FOUCAULT, DELEUZE and GUATTARI, and LATOUR and CALLON all claim that dispersion is central to the social, and further concur that social phenomena are *configurations of heterogeneous multiplicities*. Refusing to reduce variation and difference to sameness, they insist that social formations can be grasped only via maps of their heterogeneous constitutive multiplicities, not by reference to origins, single basic phenomena, or common reference points. In arguing thus they echo WITTGENSTEIN's contention that understanding a concept requires an overview of the use of words, not, as traditionally supposed, a grasp of essence.

Now, social thought has long deemed agency central to social affairs. Since multiplicity, for the above theorists, marks everything the social comprises, agency enters its province where it is proliferated in two basic ways. Following FOUCAULT's conception of the dispersion of the subject into multiple discourse positions, one line of thought, found above all in actor-network theory, cleaves the occurrence and nature of agency to the multiplicity of attributions thereof. A second line of thought, found both in DELEUZE and GUATTARI and actor-network theory and again presaged in FOUCAULT's treatment of the subject, analyzes agents themselves as multiplicities. Both lines, by scattering agency into either the vagaries of discourse or the heterogeneities of networks, license and encourage the attribution of agency to nonhuman entities, not just living organisms but artifacts and inorganic things as well. Combined, moreover, they transform all entities into potential agents, while also tying human activity to the determination and support of subhuman and nonhuman thing–agents. Herein lies the central challenge of multiplicity thinking to those humanisms that trumpet the uniqueness of humans as doers, effectors, and instigators.

2. Discourse multiplicity

The discourse multiplicity approach to agency take its lead from FOUCAULT's conception of discourse positions. In *The Archaeology of Knowledge*, FOUCAULT counters the once dominant idea, that subjects are the source of discourse, by dispersing them among subject positions that are furnished in discourses. (For present purposes, think of these positions as both roles and occupations such as mother and banker as well as ways of being such as listener and speaker.) Subjects are "dispersed" among such positions in that their erstwhile self-sameness is replaced by a multitude of discontinuous positions. Part of the import of the notion of a discourse position, moreover, is that what is entailed in occupying one is specified by the discourse providing it. When someone, for instance, occupies the position "author" that is found in certain modern discourses, that the name of the author performs a classificatory function vis–à–vis texts and that to be the author of a text is to own it, to have a "deep motive" for producing it, or to be the source of the expression manifested in it—these matters are specified by the discourses concerned and do not result from anything a particular author is or does (see FOUCAULT 1984).

FOUCAULT's analysis is usually taken to epitomize contemporary "decenterings" or "fragmentations" of the subject. It is important, consequently, to point out that it fragments only one dimension of the subject, the discursive. Consider the following sentence: "Thus conceived, discourse is not the majestically unfolding manifestation of a thinking, knowing, speaking subject, but ... a totality, in which the dispersion of the subject and his discontinuity with himself may be determined." (FOUCAULT 1972, p55) Despite appearances, this sentence does not aver that the subject is *nothing but* a discontinuous pastiche of positions. For the subject involved, the subject spoken of throughout *The Archaeology*, is a living, bodied, and acting person who is never simply whatever disparate positions it assumes when speaking, writing, listening, watching and so on. Evidence for this interpretation is found, *inter alia*, in the fact that throughout the book, without adequate explication, FOUCAULT calls discourses "discursive practices". It is only on pages 208–9 that the reason for this expression becomes clear: A discourse's statements are things people do, i.e., actions. So the rules that govern a discourse's dispersed objects, concepts, positions etc. are in fact rules that govern discursive activity, although they are "not so much limitations imposed on the initiative of subjects as the field in which that initiation is articulated, ... rules that it puts into operation, ... relations that provide it with a support... [Mine] is an attempt to reveal discursive practices in their complexity and density; to show that to speak is to do something." (FOUCAULT 1972, p209; cf. p200)

Subject positions, in short, are statuses people assume in their capacity as discoursing actors, as producers and consumers of statements, a capacity that embraces perception. Since they are actors, however, they are always more than discoursers alone. As detailed in FOUCAULT's subsequent writings, and as already suggested by the acknowledgment in *The Archaeology* that the discoursing subject is a perceiver, people are also (1) performers of nondiscursive actions and (2) bodily creatures possessing and possessed by physiology, anatomy, sensations, and pleasures. It is true that these later writings analyze body and nondiscursive activity as socially molded phenomena, just as *The Archaeology* analyzes discourse positions as similarly shaped. The social entities involved, however, are no longer discourses but biodisciplinary apparatuses, which comprise nondiscursive activities and architectures in addition to discourses (see, e.g., FOUCAULT 1980).

Actor-network theory applies discursive dispersion to agency. It treats "actor" (or "agent") as a discursive status, though not so much one that discoursers occupy as one that is attributed to entities discursively. Bruno LATOUR, for example, analyzes actors as actants, where an actant, roughly speaking, is anything said (in a story) to do something. Anything that as narrated has effects, brings something about, impacts this or that, or makes a difference in some way, counts as an actor. LATOUR (1991, p122) more picturesquely defines an actor as a stabilized "list of answers to trials [of strength]": a stabilized concatenation of performances, occurring in different agonistic states of affairs (each embracing multiple actions), that is attached to a name and thereby attributed to a substance.

Meanwhile, Michel CALLON (1991, p140) writes that an actor is anything able to associate texts, humans, nonhumans, and money in a network. People obviously qualify. In her work, for instance, a nuclear power plant technician might tie together technical manuals, fellow employees, graphite rods, control panels, poker winnings, and decontamination chambers. The graphite rods, however, might associate such entities, too, by thwarting prescribed procedures, cracking control panel housings, drawing technicians into the control room, causing one of them to drop her winnings, and forcing workers into decontamination chambers. For CALLON, who or what the agents are in a given network is a matter of who or what is credited with associating entities. Consequently, what is taken as associator (actor) or associated in a given network is relative to who describes the network together with the discursive conventions he or she cites in doing so.

For actor-network theory, consequently, the ordering of networks around agents and nonagents is an effect of imputations of these statuses. On this view, both the character of agency and what qualifies as an agent or nonagent depends on the imputations involved. For John LAW, an actor-network theorist under the influence of FOUCAULT, moreover, such imputations come organized in the form of discourses. This entails that what qualifies as an agent derives from discourse. For example, whereas in some discourses events result from indomitable structural forces, in others even the same events arise from the efforts of individuals, or the struggles of collectives. In LAW's (1994, p74) words, "…it is possible to impute several *modes of ordering* to the talk and actions of managers. And I'm saying that people are written into them in varying degree… So I am saying that agents are effects which are generated by such modes of ordering. The subject has been decentred."

Construing an agent as anything to which do-ing is attributed obviously abets the idea that nonhuman entities are agents. Graphite rods, computers, alloys, power plants, black holes, tornados, expert systems, people, dogs, scallops, viruses, jungles, you name it—it qualifies as an agent if there exists a discourse in which it is said to do something.

In one regard, I believe, this position is unassailable. Doing *is* the core feature of agency. Indeed, agency is roughly equivalent to doing. To say that Y is attributable to the agency of X is to say that X either did Y or did something that directly effected Y. If Y is an action, the agent is whatever did it, that is, carried it out; if Y is a change in the world, such as a shovel's change of location or the collapse of a toy house, the agent is whatever did something that directly brought it about, for example, the child who moved the shovel or the fan that blew on the house.[1] So all actor-network theory does is correctly note that doing is attributed to many sorts of entity. However, this theory plays cavalierly with distinctions in which *types* of do-ing (agency) are appropriately attributed to different sorts of entity, despite its predilection for multiplicity.

Compare, for instance, attributions of agency to a person, a dog, and a tornado. When we say of a person that she did something, say, cut down a tree, more than likely we are saying that she did so intentionally (i.e., aimed at a result brought about). Many of the actions she performs intentionally, moreover, will be carried out deliberately and as a result of plan-

ning.[2] And in many cases where she does something unintentionally (e.g., exhausts the family), her doing so will be a product of things she did intentionally and maybe also as a result of planning (e.g., relentlessly searching for just the right ax). By contrast, when we say of a dog that it did something (e.g., asked for more grapes), we only sometimes mean that it did so intentionally. Dogs often act without intention, and more or less nothing they do is done either deliberately or as a result of planning. Saying of the tornado, finally, that it flattened a seven mile long path does not entail intentionality, let alone deliberation and planing. It implies only that a meteorological phenomenon caused by violently colliding air masses had certain physical effects. Agency in this case is physical causality and nothing more. If doubts linger about the significance of these differences, consider the very different measures that are taken against marauding people, dogs, and tornados, or the very different techniques that must be employed to "enroll" these different entities into networks.

These points, I assume, are obvious. Yet, actor-network theorists insist upon ascribing the paradigmatic human type of do-ing, intentional agency, to a wider variety of nonhuman entities than is standard practice. At one point in his famous article about the scallops of St. Brieuc Bay, for instance, CALLON (1986, pp219–220) describes them as "dissidents". A trio of scientists had concluded that scallop larvae would anchor themselves to collectors immersed in the Bay. Many larvae, however, fail to do so, thus "betraying" their "enrollment" into networks of scientists, fishermen, tides, parasites, local politicians, and so on as previously "negotiated" with the scientists. Descriptions such as this, however jarring, raise a familiar and important issue: To which nonhuman entities is intentionality *correctly* ascribed? Today we know that animals boast varying degrees of intentionality and that machines might soon qualify as intentional beings. But scallops?

Unfortunately, actor-network theorists nowhere argue that certain entities hitherto denied purpose and will, for example, scallops and tornados, have been wronged. Without offering arguments germane to particular cases, they simply extend a way of talking paradigmatically applied to human beings to creatures immensely different from the paradigm case. (Notice that my complaint is not that actor-network theory anthropomorphizes, nor that anthropomorphisms are illicit.) As justification, CALLON/LATOUR (1992) instead claim that expansive extension of the language of intentionality is required to overcome an absolutist distinction between "the language of things in themselves" and "the language of scientists among themselves," that is, a hard and fast distinction between nonhumans and humans. Indeed, they do not really mean, they write, to grant intentionality to things. The point of extending the language is methodological—to force recognition of the interminglings of humans and nonhumans and the contributions nonhumans make to the networks of which humans are part. CALLON/LATOUR (1992, p354) add that they are only trying to develop a "symmetric metalanguage" for the description of humans and nonhumans and cannot be held responsible if no "unbiased" vocabulary exists.

An unbiased, symmetrical vocabulary does exist, however: the language of doing, applicable without prejudice to humans and nonhumans alike. If, moreover, the issue is one of recognizing the contributions of nonhumans to the networks of which humans are part, the vocabulary of doing is sufficient. There is no need to add to this simple common vocabulary further terms that apply paradigmatically to humans. In fact, doing this is likely to undercut the goal of sorting out the contributions that different entities make to networks. Using a vocabulary that connotes human ways of being obfuscates the overwhelming likelihood that entities act "categorically" differently, that is, that the natures of their doings and contributions vary; and getting these different contributions right will involve differential applications of terms, indeed, probably some of the differences leveled by LATOUR and CALLON's overly thick symmetrical metalanguage. In addition, justifying the extension to nonhumans of terms paradigmatically applied to humans on the grounds that this overcomes absolutist distinctions between humans and nonhumans ignores that any such distinction is already blurred in extant practices, for instance, by the attribution of intentionality to chimpanzees and dogs. Attributing intentionality to scallops in particular and nonhumans in general is unnecessary for the purpose of counteracting theories that consign intentionality to humans alone.

In short, it is one thing to say on linguistic or other grounds that graphite rods, computers, scallops, chimpanzees, and humans are all agents, i.e., do-ers. It is quite another to attribute intentionality to them. Certain distinctions among entities must be respected; and not every word used for humans should be applied to nonhumans indiscriminately—or at all (cf. CALLON/LATOUR 1992, p353).

To approach this point from a slightly different angle, consider Steve WOOLGAR's attempt to defend

the attribution of activity and mentality to machines. In response to the objection that such talk is just metaphorical, WOOLGAR (1991, p91) remarks that the interesting question is what entitles us to attribute intentionality to humans in the first place. The answer is that nothing "entitles" us. The language of mentality is today paradigmatically a discourse about human beings, whose extension to nonhumans reflects the degrees to which the activities and constructions of these entities resemble those of humans.[3] The more creatures approximate us, the easier and more comprehensive the attribution to them of actions and mental conditions. This means, incidentally, that the distinction between humans and nonhumans defended here is not an a priori one. It is based, on the one hand, on analyses of the human form of life which theoreticians find themselves already carrying on once they attain the sophistication to analyze it and, on the other, on observed differences between the character of that form of life and the lives of other species. Consequently, even though conceptions of human kinship to other entities vary over sociohistorical space–time, and even though "primitive" peoples attributed intentionality to a wider variety of phenomena than modern Westerners do, WOOLGAR, in the contemporary scientific context in which he writes, illegitimately shifts the burden of proof. Detractors of actor-network's profligacy need not defend the applicability of the language of mentality/activity to humans—it is already ubiquitously applied to them, and it is here that it acquires its meaning. Rather, actor-network theory must argue for the extension of this language to nonhumans on a case by case basis; it is not enough that this extension crisscrosses certain misguided absolutist ontological divides.

3. Network multiplicity

Actor-network theory does not simply fracture agency into diverse occurrences of doing. It also treats agents themselves as heterogeneous networks. Agency, consequently, is an unstable and possibly disunified feature of whole networks that depends on the operations of their parts.

Before considering these claims, consider the parallel views in DELEUZE and GUATTARI. (The following combines ideas found in DELEUZE/GUATTARI 1983; 1987.) For these theorists, an individual entity (organism, person, machine, social formation) is a multiplicity of molecular assemblages. A human being, for instance, is a multitude of physiological, neurological, genetic, hormonal, cognitive, and conative assemblages. More expansively, individual entitites such as human beings, organisms, and social formations are "molar" phenomena that arise from the segmental organization of molecular multiplicities. The psyche, personality, and self-consciousness that help mark a human being, for instance, arise from the segmental organization of physiological, neurological, genetic, hormonal, cognitive, and conative assemblages. A human being, consequently, is an integral organized molar assemblage of molecular multiplicities.

"It will be noted that names … function as common nouns ensuring the unification of an aggregate they subsume. The proper name can be nothing more than an extreme case of the common noun, containing its already domesticated multiplicity within itself and linking it to a being or object posited as unique." (DELEUZE/GUATTARI 1987, p27)

The molecular, it might be added, is the home of desire; its components are "originally" structured as a manifold of "desiring machines". Individuals, moreover, are molar desiring machines that invest their physical, organic, and social environments. Disclosing this picture's psychoanalytic provenance, DELEUZE/GUATTARI (1983, p289) call the energy of desiring machines "libido" and label the theory of "the machinic connections of a molecular order, their distribution into maps of intensity on the giant molecule of the body without organs, and the statistical accumulations that form and select the larger aggregates" the "theory of schizophrenia". They also tie this psychoanalytic conception to contemporary cell microbiology in treating DNA-coded proteins as the "unconscious" components of molecular desiring machines (see DELEUZE/GUATTARI 1983, p290).

DELEUZE and GUATTARI never focus on agency *per se*. Given their (1987, pp256, 260) definitions of latitude and longitude as, roughly, the particle aggregates belonging to an entity and the affects (in SPINOZA's sense) of which that entity is capable, it makes sense to construe the agency of a human being (or of any other entity)—like his or her psyche, self-consciousness, and personality—as a molar phenomenon. Accordingly, stable and unified individual agency rests on the successful and persisting organization of micromultiplicities, the ordering of a "city" of molecular desiring machines. Agency, however, is a property not just of individuals, but also of the micro desiring machines whose combined action gives rise to individuals' agency. Indeed, molar agency is at once something people

share with organisms, machines, and social formations and an epiphenomenon dependent on the organization of the molecular agencies that supply these molar entities with motion.

Actor-network theorists treat entities, similarly, as networks. As in DELEUZE and GUATTARI, they are networks to which unity is ascribed. A human being, for instance, is a network of neurons, muscles, memories, skills, preferences, hormones etc. taken as a unit. (Actor-network theory calls this telescoping "black boxing" or "punctualization".) An actor, moreover, is any network to which both unity and doing are ascribed: Taking X to be an actor is apprehending the network that X is as a unit and crediting that unit with doing something. "[A]n agent is a spokesperson, a figurehead, or a more or less opaque 'black box' which stands for, conceals, defines, holds in place, mobilizes, and draws on, a set of juxtaposed bits and pieces." (LAW 1994, p101) LATOUR, recall, defines actors as the substances to which stabilized concatenations of action are ascribed. We now see that the "substances" involved are really networks taken as units. John LAW (1991) offers a variation of this theme when he writes that any network sufficiently stable to generate power effects is an actor.[4]

In this context, consider CALLON's (1991, p141) description of nuclear power plant. This hybrid of "graphite rods, turbines, atoms, operators, control boards, flashing lights, concrete slabs, and engineers [converts] files, bills, fuel, water, skills, and budget lines into electrons transported to consumers, taxes paid to local councils, and waste products—which in turn lead to the formation of groups of angry environmentalists."

If one groups together such actions as transporting electrons, paying taxes, and producing waste products, the actor that performs them, i.e., the substance to which they are attached, is the power plant. The power plant, however, is a network of graphite rods, turbines, operators, and the like. So the actor that performs these actions is a network treated as a unit. If, by contrast, one groups together such actions as containing emergencies and working shifts or heating up and melting down, the actors involved are the operators and graphite rods respectively. Each of these agents is also, of course, a network taken as a unit. Consequently, an ascription of agency, as in DELEUZE and GUATTARI, is an instantaneous apprehension of multiplicity. By considering different congeries of action, moreover, agency can be seated in any component of a network, as well as in the network as a whole.

Agents are not just networks. They are also "effects" of networks. An individual cannot do anything without the support of the network composing it (organs, memories, skills, preferences, and artificial limbs in the case of human beings; machinery, operators, fuel rods, and so on in the case of the power plant). Indeed, an individual's activity is generated by the network composing it. Agency is *also* dependent, however, on the wider network(s) in which agents are embedded (chairs, computers, students, and administrations in the case of professors; customers, suppliers, operating companies, and subsoil geologies in the case of the plant). So the networks of which agency is a unity-effect come in two flavors: compositional and embedding. And the components of networks of both sorts are themselves (potential) agents that are generated by the further networks whose unities they are. Incidentally, one might wonder where this decomposition bottoms out (or the contextualization tops out). Only LATOUR tackles the issue philosophically. Reality, he (1988, chapt. 1) claims, is a NIETZSCHEAN plenum of incommensurable and indivisible, acting and acted on forces, the number and identity of which can be determined only after they take the measure of one another in "trials of strength". According to his account, consequently, actor-network decomposition terminates somewhere, though its endpoint is a shifting one subject to redetermination in further trials and never an immutable fundament.

I wish to describe one further proposition resulting from the analysis of agents as networks. CALLON and LAW suggest that features of the networks that underwrite (human) agents determine the types of actor they are. In, for instance, networks whose human components largely (1) share definitions of people, things, and situations and (2) observe the same rules in both attributing identity and agency and black boxing generally, people are actors with precise objectives and instruments, perfect but limited information, few choices, and no disagreements. When definitions, attributions, and black boxing diverge, by contrast, people are actors who pursue strategies, negotiations, revisable projects, and varied aims. Intermediate cases between these extremes compel the application of other conceptual apparatuses to human activity, for instance, procedural rationality or game theory. Hence, "…there is no theory or model of the actor, even in the plural. The actor has variable geometry and is indissociable from the networks that define it and that it, along with others, helps to define". (CALLON 1991, p154) This means, I believe, that which properties beyond doing are ascribed to some-

thing *qua* agent depends on the networks within which it is one.

Much of these accounts, I believe, is unproblematic. Agents, to begin with, *are* networks to which action is ascribed. Human beings, for instance, *are* molar aggregates of molecular assemblages, networks of organs, preferences, memories etc. to which unity and doing are ascribed. Agency, moreover, *is* a unity-effect generated by these networks. Just as human consciousness depends (presumably) on brain operations, self-conscious intentional agency arises from the intermeshed operations of multiple bodily systems. Agency also indeed rests on embedding networks. Human agency, for instance, depends on the networks of people, organisms, artifacts, and things that populate human worlds. And, yes, components of compositional and embedding networks are themselves (potential) agents that are generated by the further networks they black box: Pets, computers, bodily systems, and the like do things, thereby qualify as agents, and upon closer inspection turn out to be networks of entities, the components of which are likewise agents. In short, when actor-network theory is treated as an account of the causal material composition of agents and the causal conditions of agency, it is relatively uncontroversial (or at least unexotic).

Multiplicity theorists write, however, as if their conceptions of constitutive multiplicities captured a deep truth that subverts human agency. Actor-network's picture of multileveled networks connected via association and black boxing, for instance, asks the reader to treat humans, parts of humans, and nonhumans alike. Much of the fanfare attending the "discovery" that people are assemblages or networks arises from the thesis that human agents are composed of and dependent upon nonhuman and "subhuman" actors. As discussed, however, the thesis that these other–than–human entities are actors holds only at a high level of generality: Actor-network theory resembles an obedience trainer who remarks that because any dog is a dog, all dogs are alike. Once the thinness of the claimed equivalence is appreciated, initially provocative formulations become more familiar. For instance, the proposition that human agents are (effects of) networks of further agents resolves into such claims as (1) that human actors are composed of active physical subsystems that maintain causal relations among themselves and with the environment outside the skin, and (2) that what people are capable of doing depends in part on the people, organisms, things, and artifacts around them. Human actors, consequently, are indeed assemblages-networks. But this realization does not undermine human agency and should disconcert only those who metaphysically distance will and its motivational context from material (and social) reality.

In its haste to level humans and nonhumans, actor-network theory also neglects to sort out significant differences between compositional and embedding networks. The agency effected by networks, for example, is located differently depending on the network involved. *Qua* effect of compositional networks it is attributed to the networks themselves as units, whereas *qua* effect of embedding networks it is ascribed to components thereof. These two sorts of network also maintain different functional relations to activity.

Of a compositional network it makes sense to say that an agent is both a network and an effect thereof. An actor is its compositional network because anything is that of which it is composed. An agent is also an effect of its compositional network since its capacity to act as a single entity depends on the cooperation of its components. An actor is not, however, its embedding networks: a professor is not her chair, computer, secretaries, and administration, just as a power plant is not its operating company, nuclear fuel suppliers, and subsoil geologies. To deny this is to deny that entities have boundaries, ultimately, that there are distinct entities at all. Something's agency, furthermore, is not *generated* by its embedding network as it is by its compositional one. The doings of a professor or power plant are not generated or brought about by networks of chairs, secretaries, operating companies, and subsoil geologies: A professor can act (hopefully) sans her chair and secretary, just as a power plant can act absent its operating company (it is now employee owned).

Agency is an "effect" of embedding networks in at least three other ways. First, agency *presupposes* certain general types of embedding network, for example, arrangements of physical things, since without them there wouldn't be anything to act on and the networks that compose agents couldn't function or survive. Second, embedding networks circumscribe agency (they are not alone here). Without his computer the professor might not be able to finish next Tuesday's presentation, but he still can bemoan losing the ability to write legible longhand. Similarly, without its graphite rod suppliers the plant won't be able to supply its customers with electricity, but it might still lobby the government to ease safety restrictions on the production of nuclear fuel. Agency, therefore, is an effect of embedding networks, sec-

ond, because networks *enable* and *constrain* (better: *prefigure*) what agents do. An important form this effectuation takes is the enforcement of norms pertaining to the whats and hows of action. A third way embedding networks effect agency is by people *imputing* agency (doing) to other network elements (cf. section 2).

Not only has actor-network theory neglected to sort out these differences; the further thesis that networks determine the character of agency overextends its insights. As noted, CALLON and LAW maintain that which features other than intentional doing mark human agency depends on the character of the networks as part of which people proceed. Whereas in some situations actors must be credited with precise objectives, perfect but limited information, and few choices, in others they must be imputed strategy, negotiation, and varied aims. Variations of this sort, however, do not entail that the nature of human agency is variable and that a "model" or "theory" of the actor is impossible. Precise versus varied objectives, perfect versus imperfect information, and few or many choices are different values of specific parameters of a single (broadly speaking rational choice) model of activity, not the constants of different ones. Moreover, the presence or absence of strategies and negotiations results from the specific values these parameters assume in sets of individuals in particular interactional circumstances. Variability in these matters means neither that what a human actor is varies across networks nor that a single model of agency is utopian.

I write this realizing that no extant model is sufficiently capacious or realistic to capture all occasions and aspects of human agency. This failure arises, however, from the patently reconstructive nature of many models and/or the simplification inherent to them all. It does not confirm the alleged dependency of the character of human agency on embedding contexts, even though circumstantial features of agency clearly are so conditioned. Although, in addition, no current model covers intentional along with all forms of nonintentional agency, a future model might prove successful here.

At the same time, fundamental differences among the agencies of humans, dogs, scallops, power stations, and tornados are tied, I presume, to the different types of network *composing* these entities. In this regard, the nature of agency is indeed an effect of networks, and a single model of agency is implausible. Notice that these claims align WITTGENSTEINIAN conceptual dispersion vis–à–vis agency with the disparate multiplicities that compose different types of agent.

4. Humanism saved?

"Humanism" has no precise meaning. Generally speaking, it is a cultural stance that arose in Europe during the 1500s and 1600s to celebrate Man as thinker, creator, and actor. This exaltation of the human coincided with growing sentiment that humans are not essentially beholden to God, the chain of being, or the implacable order of the universe. They can create and take responsibility for conceptual and material structures whose splendor honors their capacities and power. A pair of centuries passed before more militant atheisms proclaimed Man the equal or vanquisher of God as well as the unique source of meaning, truth, and being.

Theories of social multiplicity damage humanism in this general sense scarcely. They do not, for instance, deny that human beings are agents, a gainsaying that would indeed hollow out any humanism. Their claim that human agency is an effect of multiplicities at bottom means that agency arises from the co-operation of the multiplicities that compose a person and is enabled and constrained by those embedding her. This claim does not debunk agency; rather, it conceptualizes its character and provenance scientifically. Like the neurophysiological explanation of human activity familiar since at least the end of the 18th-century, it challenges humanism only on unnecessarily reductive or deterministic readings.

Theories of multiplicity might also be thought to decapitate humanism by fragmenting the subject in the manner FOUCAULT allegedly epitomizes. As discussed, however, his dispersion of subjects among discourse positions does not exhaust their being. Indeed, part of what remains is the performance of nondiscursive actions. Many theorists who develop FOUCAULT's line press, not just discursive action, but its nondiscursive kin too, into the fragmented template of subject positions or their ilk (e.g., BUTLER 1993). In their hands, however, subject positions are mere scripts or bundles of norms (whose specificity and malleability varies for different theorists). Such accounts emphasize the social formation of agency, not its dissolution.

Of course, if the humanist subject is take to be the ethereal free agent of Western lore, humanism indeed suffers at the hands of multiplicity theory. Obituaries of this agent, however, first appeared long ago. If, consequently, multiplicity theorists are taking aim at this subject, they, like too many others, are simply re-sounding its death knoll in the face of dwindling efforts to save it.[5]

The main challenge multiplicity theory poses to humanism arises from its proliferation of agency. In theory, of course, the idea that animals, machines, and power plants are actors need not trouble anyone intent on celebrating the thinking, making, and doing of humans. Humanism need not proclaim people the only agents. There is no denying, however, that it has tended in this direction and thereby rendered itself vulnerable to intellectual developments that multiply agency. Still, the proliferation of agency is compatible with the distinctive and unique richness of the human variety—and this is distinction enough to satisfy many humanists. So the real challenge to humanism cannot arise from the insights that people are multiplicities and many (potentially all) nonhumans agents.

It is instead posed by the possibility that entities other than humans are the type of actor they are. Perhaps dolphins act intentionally, deliberately, and on the basis of plans. Maybe machines will confront and confound us one day soon with such behavior. Maybe some day prosthetic and implant technology will produce exotic human–machine hybrids who display it. And perhaps one day anywhere along the way we will be contacted by extraterrestrials who have been acting thusly all along. Modesty precludes abjudicating whether humans alone possess the self-conscious, intentional, deliberative, and planning agency they display. At the same time, we must not overlook that as far as we do know today humans alone, and not even all of them, possess it fully.

To go one step further: Even if the above possibilities materialize, we might not want to consign humanism to the past. We might still continue to view humans as special, on the grounds that they are the entities overwhelmingly responsible for the quality of life they themselves enjoy or suffer. This fact, increasingly the case throughout history, shows no sign of abating in the future. It is human agency that primarily fashions the worlds people live in and that must be confronted in making those worlds better places. (Notice that this is a causal, not an ethical point.)

I don't know whether this "saves" humanism. It certainly leaves room for a clairvoyant variant, one cognizant that nonhumans are agents, that humans are multiplicities whose agency rises therefrom, and that humans may not be the sole creatures in the cosmos capable of self-conscious, intentional, deliberate, planning activity.[6]

Author's address

Theodore R. Schatzki, Department of Philosophy, University of Kentucky, Lexington, KY 40506-0027 USA.
Email: schatzki@pop.uky.edu

Notes

1 This sort of definition, I acknowledge without further comment, faces subtle dilemmas concerning omissions, the distinction between doing and happening, and the proper delimitation of causal chains. See, for instance, BRAND (1971) and DAVIDSON (1980a, 1980b).

2 For a discussion of different ways of doing something, see AUSTIN (1970).

3 This way of thinking derives from WITTGENSTEIN (1958).

4 The relevant effects are stockpiles of, degrees of discretion concerning, and the particular circumstantial character of both "power-to" and "power-over".

5 Jean-Paul SARTRE may be metaphysical freedom's last great defender. In his eyes (1966), *pace* what I wrote earlier, a human agent is no more her compositional network than she is her embedding ones—for the subject cannot be constituted by anything in the world. In later works, of course, SARTRE (e.g., 1968) tried to square free action with sociohistorical contextualization.

6 This essay was completed while a visiting scholar at the Max Planck Institute for the History of Science in Berlin. I would like to thank the Institute and the director of Abteilung Drei, Hans-Jörg RHEINBERGER, for their support.

References

Austin, J. (1970) Three Ways of Spilling Ink. In: Austin, J. Philosophical Papers, 2nd edition. Oxford University Press: Oxford, pp. 272–287.

Brand, M. (1971) The Language of Not Doing. American Philosophical Quarterly 8 (1): 45–54.

Butler, J. (1993) Bodies that Matter. Routledge: New York.

Callon, M. (1986) Some elements of a sociology of translation: domestication of the scallops and the fishermen of St Brieuc Bay. In: Law, J. (ed) Power, Action and Belief. Routledge and Kegan Paul: London, pp. 196–233.

Callon, M. (1991) Techno-economic networks and irreversibility. In: Law, J. (ed) A Sociology of Monsters: Essays on Power, Technology, and Domination. Routledge: London, pp. 132–161.

Callon, M./Latour, B. (1992) Don't Throw the Baby out with the Bath School! A Reply to Collins and Yearly. In: Pickering, A. (ed) Science as Practice and Culture. University of Chicago Press: Chicago, pp. 343–368.

Davidson, D. (1980a) Agency. In: Davidson, D. (ed) Essays on Actions and Events. Oxford University Press: Oxford, pp. 83–102.

Davidson, D. (1980b) Intending. In: Davidson, D. (ed) Essays

on Actions and Events. Oxford University Press: Oxford, pp. 43–62.

Deleuze, G./Guattari, F. (1983) Anti-Oedipus. (Translated by R. Hurley, M. Seem, and H. Lane). University of Minnesota Press: Minneapolis.

Deleuze, G./Guattari, F. (1987) A Thousand Plateaus. (Translated by Brian Massumi). University of Minnesota Press: Minneapolis.

Foucault, M. (1972) The Archaeology of Knowledge. (Translated by A.M. Sheridan Smith). Harper and Row: New York.

Foucault, M. (1980) The Confession of the Flesh. In: Gordon, C. (ed) Power/Knowledge. Pantheon: New York, pp. 194–228.

Foucault, M. (1984) What is an Author? (Translated by J. Harari). In: Rabinow, P. (ed) The Foucault Reader. Pantheon: New York, pp. 101–120.

Latour, B. (1988) Irreductions. In: Latour, B. The Pasteurization of France. (Translated by A. Sheridan and J. Law). Harvard University Press: Cambridge, part 2.

Latour, B. (1991) Technology is society made durable. In: Law, J. (ed) A Sociology of Monsters: Essays on Power, Technology, and Domination. Routledge: London, pp. 103–131.

Law, J. (1991) Power, discretion, and strategy. In: Law, J. (ed) A Sociology of Monsters: Essays on Power, Technology, and Domination. Routledge: London, pp. 165–191.

Law, J. (1994) Organizing Modernity. Blackwell: Oxford.

Sartre, J.-P. (1966) Being and Nothingness. (Translated by H. Barnes). Pocket Books: New York.

Sartre, J.-P. (1968) Search for a Method. (Translated by H. Barnes). Vintage: New York.

Wittgenstein, L. (1958) Philosophical Investigations. (Translated by G. Anscombe). Macmillan: New York.

Woolgar, S. (1991) Configuring the User: the case of usability trials. In: Law, J. (ed) A Sociology of Monsters: Essays on Power, Technology, and Domination. Routledge: London, pp. 57–99.