Animal Sentience

A Study of Paradigmatic Arguments in Cognitive Science

Manuel Bremer

Yet, although animals easily communicate to us, by voice or bodily movement, their natural impulses of anger, fear, hunger and so on, it has never been observed that any brute animal reached the stage of using real speech, that is to say, of indicating by word or sign something pertaining to pure thought and not to natural impulse. Such speech is the only certain sign of thought hidden in a body. All men use it, however stupid or insane they may be, and though they may lack tongue and organs of voice; but no animals do. Consequently it can be taken as a real specific difference between man and dumb animals. For brevity's sake I here omit the other reasons for denying thought to animals. Please note that I am speaking of thought, and not of life or sensation. I do not deny life to animals, since I regard it as consisting simply in the heat of the heart; and I do not deny sensation, in so far as it depends on a bodily organ.

(Rene Decartes, Letter to More, February 5th 1649)

Content

Preface		1
§1	Paradigm Case Arguments	9
§2	Reflective Equilibrium (I)	10
§3	Not Simply Top-Down	11
§4	The Neurophysiological Paradigm	17
§5	The Behavioural Paradigm	18
§6	Appeal to a Certainty Intuition as Paradigm	19
§7	The Evolutionary Paradigm (I)	22
§8	Dualism and Sentience	23
§9	The Non-Conscious Capacity Paradigm	25
§10	The Pain Paradigm	26
§11	The Evolutionary Paradigm (II)	29
§12	Animal Attitudes (I)	31
§13	Pain as Functionalist Paradigm	33
§14	Primates as Special Case	34
§15	The Simplicity Paradigm	36
§16	The Fitness Paradigm	38
§17	Animal Concepts (I)	39
§18	Animal Attitudes (II)	41
§19	The Language Paradigm (I)	43
§20	The ‘Theory of Mind’ Paradigm	45
§21	The Simulation Paradigm	49
§22	Reflective Equilibrium (II)	50
§23	The Anthropomorphic Fallacy	52

§24 Heterophenomenology 54
§25 A New Anthropomorphism 55
§26 Sentience and Animal Welfare 59
§27 Subjectless Awareness? 68
§28 The Language Paradigm (II) 72
§29 Animal LOT (I) 73
§30 Shared LOT Types? 74
§31 Animal LOT (II) 80
§32 The Gricean Paradox 81
§33 The Language Paradigm (III) 82
§34 Human Cognitive Architecture 83
§35 The Versatility Paradigm 88
§36 Versatility in Behaviour 89
§37 Animal LOT (III) 91
§38 The Priority of Behavioural Complexity 92
§39 Varieties of Sentience 93
§40 Human Dual LOT Coding 93
§41 Animal Attitudes (III) 96
§42 Goal/Registration-Explanations 97
§43 Animal Concepts (II) 99
§44 Conditions of Sentience 102

Epilogue 105

Bibliography 107

Preface

Humans are unique among living beings in having language, culture – and consciousness, all in the proper (narrow) sense of the words. Extending the sense of "language" (to cover animal signalling), "culture" (to cover animal tool use) and "consciousness" (to cover animal mentality) may be motivated by moral concern for animals, but hinders a scientific understanding of animal minds, so far as animal species have minds. The book deals with the distinction between consciousness and sentience. It raises the question how to investigate sentience. In so doing the focus is on paradigmatic arguments concerning their distinction or continuity in cognitive science.

Cognitive Science (CS) is typically understood as the inter- and transdisciplinary field of science targeting cognitive capacities and even the mind in general. It contains parts of empirical cognitive psychology, theoretical philosophy, cognitive neuroscience, theoretical computer science, artificial intelligence, cognitive linguistics, cognitive ethology, and maybe more. Philosophy of Mind has become part of CS. Nowadays a theory of (parts of) the mind has to be informed of the state of the art in CS and to be continuous to theories in CS. Large parts of traditional epistemology have been dissolved and integrated into CS.

Additionally there is the Philosophy *of* Cognitive Science, a branch of the philosophy of science focusing on the conceptual and methodological foundations of CS. It aims at elucidating fundamental concepts and developing models which could cover the (types of) results of CS.

The Philosophy of Animal Minds partially overlaps with theory development in CS, in relating models of animal minds and their capacities to theses of the Philosophy of Mind, and partially contributes to the Philosophy of Cognitive Science, especially with meta-scientific reflections on Cognitive Ethology.

Paradigmatic arguments are arguments that re-occur in CS literature and might be employed across several sub-fields. The philosophy of CS tries to define such paradigm/paradigmatic arguments and to evaluate their cogency in their respective uses.

This will be done at the beginning of text. The text then collects and reflects on some paradigm arguments in the fields of animal cognition and sentience (a.k.a. 'animal consciousness' or 'animal minds' by many in CS). Later on in the text, the focus lays on *how* to conceive of sentience in animals.

This book is *not* an introduction or text book covering consciousness or animal minds. There are superb collections on animal cognition and cognitive ethology (Bekoff/Jamieson 1996; Bekoff/Allen/Burghardt 2002), and on the philosophy of animal minds (Andrews/Beck 2019; Lurz 2000), and shelves of general treatises on animal minds (*inter alia*: Bekoff/Allen 1997, Budiansky 1998, Dawkins 1993, DeGrazia 1996, Griffin 1992, Hauser 2000, Humphrey 2022, Pena-Guzman 2022), the journal *Animal Sentience*, as well as shelves of treatises on specific species (*inter alia*: Balcombe 2016, Chittka 2022, Geoffrey-Smith 2016, Hayes 2008, Mertrinho-Truswell 2022, Premack/Premack 1983, Schustermann et al. 1986, Young 2005), which often – and controversially – ascribe massive cognitive capacities to these species.

This book is neither a treatise *in* cognitive ethology or empirical cognitive science. This is a study in the Philosophy of Cognitive Science. It analyses paradigm arguments of CS. The book contains theses belonging to the Philosophy of Mind and the Philosophy of Animal Minds, as a mere meta-scientific investigation typically seems not feasible and may stay too abstract. It supplements the existing literature in (i) focussing on the issue of reflective equilibrium between several paradigmatic approaches and their paradigmatic arguments with respect to animal minds in CS. Concerning theory building it focusses (ii) on the central issue of language in relation to thought, and the theories of *Languages of Thought* in humans and animals.

Researchers sometimes deride this type of *philosophy book* or philosophical comments on empirical research – especially if the author critiques their theories and

views (say on the broad abilities of invertebrates, fish or some other species). They join physicists in proclaiming "philosophy is dead" (Hawking/Mlodinow 2011: 13). The rationale of such philosophical studies, nonetheless, rests on their meta-scientific perspective, which evaluates standards and types of arguments. Philosophers (mostly) lack the qualifications to criticize empirical research and experimental set ups or the quality of field studies. Debates in CS, however, involve not just empirical research, but model building, drawing conceptual distinctions and putting forth theoretical arguments. Here philosophers of science are experts. Ideally the philosopher of (empirical) cognitive science keeps a distance from the controversies of the field researchers and focusses instead on the strength of methodological and argumentative standards. Philosophy in the tradition of Logical Empiricism also considers the advantages and disadvantages of different linguistic frameworks for some form of inquiry (cf. Bremer 2025: 8-19).[1]

The range of the debate about sentience extents from a complete denial of sentience to animals (e.g. Macphail 1998) to an ascription of sentience to insects (e.g. Chittka 2022). From a theoretical point of view the extreme form of human exceptionalism turns out to be more coherent than many may have expected. From the point of view of our manifest image of the world the denial of sentience at least to grown mammals and some birds seems outrageous. That a cow feels the breeze and experiences the colours of the farm's backyard seems a *certainty* on par with "I have not been born 5 minutes ago". Although I endorse strong versions of human exceptionalism (like Chomsky 2016), I also embrace the certainty with respect to the sentience of grown mammals and some birds. The reoccurring centre of the problem here will be *how* to say *what* about *which* animal's sentience.

The exposition sometimes returns to issues raised earlier in an attempt to zoom in on the problems from a slightly different ankle. The first half of the text

[1] I have been told by a field researcher better to leave the study of animal sentience to the experts, and then he went on to deliver a poetry reading the narrative of which (in complete sentences) was told from the perspective of a monkey! This kind of unbridled anthropomorphism – common in popular science books and media – does a disservice to the proper understanding of animal minds. I am not claiming that field researchers better leave the philosophy of cognitive science to the experts, because they could not do it. But somebody has to do it.

focusses on paradigmatic arguments and the different perspectives on the question of sentience and animal minds. Theoretical arguments dealing with language and *Language(s) of Thought* in the second half of the text motivate returning to some issues (like animal concepts and animal attitudes) mentioned before.

This is neither a book on animal ethics. A longer paragraph in the second half of the book, however, puts forth an ethical statement on our obligations against animals. The question of sentience connects to questions of moral consideration of animals, both directly (in utilitarian ethics) and indirectly (in anthropocentric ethics). Of course, a broader ethical justification needs to be given for a proper animal ethics (cf. Bekoff/Meaney 1998; Beauchamp/Frey 2011), but this paragraph only supplements the main text in acknowledging that the issue of animal welfare cannot be neglected here. Moral concern for animals need not and should not entail exaggerating the mental abilities of animals. This book endorses human exceptionalism. Humans uniquely take the moral stance, even with respect to members of other species. We can decide morally to err on the side of caution instead of adhering strictly to our best current scientific theories.

Well known theories and concepts in cognitive science (broadly taken) are not explained, but simply referred to (e.g., the Language of Thought, Computational Theory of Mind, Bayesian Networks, Turing Machines, Gricean Intentionalism, Theory of Mind etc. etc.). The book engages the debate about sentience from the perspective of the philosophy of mind, and the philosophy of cognitive science. The prospected reader is at least a graduate student of these fields or of cognitive science, and so has (at least) some (vague) background with respect to the fundamental concepts and theories.

Terms and usage cannot be separated from theory development in CS. They are interconnected with other theories or – especially in philosophy – metaphysical world views, and are often even ethically or politically loaded. Therefore, preliminary outlines of usage and conventions are helpful and often needed:

1. “Human” refers to our species, typically to a healthy adult member if not indicated otherwise.

2. “Animal” refers to all living things not being plants, fungi or humans. The term “non-human animal”, although correctly pointing to the biology of humans and the evolutionary origin of the human species, is misleading and loaded, as it suggests a communality that in the cognitive realm is at issue. Within the realm of animals as well there are obviously huge and important cognitive distinctions.

3. “Consciousness” refers to the way humans experience themselves and the world, and conduct their lives. The features of consciousness have to be set out, but it essentially includes rationality and selfhood. Animals lack these. “Human consciousness”, therefore, in this usage of “consciousness” is a pleonasm, but useful to stress the distinction to animal states. “Animal consciousness” using an adjectival modification on “consciousness” could be used for the mental life of animals, but too misleadingly indicates a sufficient similarity to the human case to speak of “consciousness”. There are some animals typically to be taken to be close in their faculties to humans (primates, ravens, and dolphins). With respect to them (only) one may consider the issue of ‘animal consciousness’ as precursor to human consciousness with some relevant similarities. Whether there is any animal consciousness of such a type has to be shown and argued for.

4. “Sentience” refers on the one hand to the faculty of having *experiences* of one’s body and the environment. Sentience is what, if one abstracts from the influence of rationality and selfhood, it seems animals of some species (presumably warm-blooded vertebrates) share with humans. One crucial question of CS is which animals are sentient and to what extent. Human experience cannot be isolated from consciousness to a mode which will not be affected by our linguistic abilities and self-related thoughts, but we may start with the idea that sentience constitutes, so to say, the lower part of consciousness, which the human capacities are added and build on, and which might be to a sufficient degree of similarity be shared with some

animals. Sentience involves *phenomenality*. Sentient animals perceive their environment and themselves from a *perspective* depending on their location.[2]

Further Usage:

- "algorithm": a substrate neutral, effective procedure which is finite (in its sub-computations) in input, consulted states and output (given it stops), by the *Church/Turing Thesis* executable by a TM
- "concept": in case of the LOT a *symbol type*, otherwise: a body of beliefs associated with a word (like: the concept of 'experience')
- "convention": mutual knowledge of a contingent rule (Lewis' [1969] concept of 'convention')
- "the I/Ego": the agent/subject of a conscious mental event (Husserl's [1950] 'ego', James' [1890] 'agent and observer'), with distinct functional aspects
- "information that …": referential content of an eternal sentence … (objective information like in [Bremer/Cohnitz 2004])
- "knowledge": true belief
- "language": a production rules governed compositional system of discrete symbols
- "lower animal": a physiologically less complex animal, remote from humans in the non-linear evolutionary tree; not a pejorative term
- "meaning": what is in a lexical entry for a word of a natural language: an association of a word with a (LOT) concept, syntactic and phonological features, containing (analytic) links to other words, and associated (i.e. not meaning constitutive) knowledge of criteria of word application
- "mental state/event": a state/event of the mind, conscious or not

[2] 'Point of view' is already misleading as it connotes subjectivity and animals have no self-concept in the proper sense. They are not in the proper sense 'subjects of a life' as Regan (2004: 50-52) claims. [We come to this.]

- “module”: an encapsulated part of the mind, either a *representational* module (containing specific concepts and rules) or a *computational* module (computing mental sub-routines and faculties)
- “natural language”: a spoken language with conventionally shared word meanings
- “percept”: a phenomenal unit in perception (like the sight of a house or the sound of a bell)
- “phenomenal”: the experiential quality of some representations
- “representations”: primarily vehicles of computation, which need not ‘represent’ (referentially) something outside the mind, but often do
- “the self”: the concept and narrative the Ego develops about itself (Dennett’s [1991] ‘centre of narrative gravity’)
- “situation”: an objective part of reality consisting of objects and their properties/relations (like non-abstract situations in Situation Semantics [Barwise/Perry 1983])

Abbreviations:

- AI (Artificial Intelligence, ‘symbol based’, like in Logic Programming)
- BL and DL (belief like, desire like states, [introduced in the text])
- CS (Cognitive Science)
- CTM (Computational Theory of Mind)
- FOT (First Order Theories of consciousness)
- GG (Generative Grammar)
- GRE (Goal/Registration-Explanation [introduced in the text])
- GWS (Global Workspace, like in Baars’ [1992, 2021] GWS-model of consciousness)
- HOT (Higher Order Theories of consciousness)
- LAD (Language Acquisition Device, like in Generative Grammar)
- LOT (Language of Thought, ‘Mentalese’)
- ML (Machine Learning)
- NN (Neural Networks, mostly meant: Recurrent or Bayesian NNs)

- PRAM (Programmable Random-Access Machine, exemplified by a standard computing machine like a desktop computer)
- RTM (Representational Theory of Mind)
- TOM ('Theory of Mind' as cognitive capacity)
- TM and BTM (Turing Machine as standard of computation, and Bounded TM, being bounded in storage)
- vNA (von Neumann-Architecture, as in standard computing machines)

Small Greek letters are variables for unspecified sentences (or on occasion for sub-parts of sentences).

Single quotes are used as 'scare quotes' for concepts or theoretically loaded concepts. Double quotes are used for quotes from texts or quotation of terms.

§1 *Paradigm Case Arguments*

Paradigm case arguments are 'paradigmatic' in two ways: (1a) they appeal to paradigms of research and (empirical) investigation of the science in question, and (2) they are often used. Paradigms in logic are meta-logical proofs of soundness or completeness, in sub-atomic physics use of particle accelerators or cloud chambers, in animal learning theory training in the Skinner Box. Paradigmatic arguments can also more abstractly focus on (1b) a paradigm approach instead of focusing on specific empirical set ups in research. Paradigm approaches in this sense in cognitive science are evolutionary functional explanations, learning theory, or analyses in computational complexity theory. Even more abstract paradigmatic arguments refer to (1c) methodological paradigms of meta-theoretical evaluation (like explanatory power or simplicity).

Both conditions (1) and (2) should be met for an argument to be paradigmatic.

The most commonly used arguments – in fact only rhetorical questions – in public debates about animal minds are:

1. How do you know that animals are not conscious?
2. How do I know that you are conscious?

Both are not paradigm arguments in the sense above, but just bad philosophy. We should get rid of them at the very beginning.

(ad 1) In general the *burden of proof* in science rests with those who put forth existence claims (be it for a unicorn or a sub-atomic particle). Otherwise it is simpler to have a theory and picture of reality where these supposed things are absent. Absence of evidence is evidence of absence, unless the absence of evidence can be scientifically explained. So, we need positive evidence for animal consciousness and – starting from our convictions about sentience in mammals and some birds – positive evidence for animal sentience *beyond* mammals and some birds.

(ad 2) Human share a mutual knowledge about their respective possession of conscious states. They talk about them. And they can talk about them in a natural language, because psychological vocabulary in learned in situations of triangulation. The teacher introduces an expression for a conscious state observing the verbal and non-verbal behaviour of the learner, who thereby associates the term

with the quality of her conscious state, and further learns to attribute this type of state to others. Psychological vocabulary and patterns of usage found our mutual knowledge of our possession of consciousness.[3] Animals are not members of our linguistic community, therefore we have no mutual understanding of their supposed sentience.

§2 *Reflective Equilibrium (I)*

Which is the science to study animal minds? There is no single of the empirical sciences that covers all ingredients of a theory of (animal) minds. We need behavioural concepts as well as neurophysiological evidence. We need evolutionary considerations as well as simulation. CS is defined as such an interdisciplinary research programme (cf. Green et al. 1996). The treatment of some topic has to reach a *reflective equilibrium* (cf. Tersman 1993) between our intuitions and a phenomenological approach, philosophical conceptual analysis and various empirical approaches and model building. Reflective equilibrium means in *this* context that we have to reach a *coherent model* which incorporates as much of our intuitions concerning sentience and integrates at the same time the findings of the different co-operating sciences. In doing this none of the approaches is favoured. There can be various trade-offs in case of conflict.

Similar trade-offs appear in *inference to the best explanation* (cf. Lipton 1991). Better explanations are more likely than others. Mechanistic or algorithmic explanations are more precise than others. Explanations extending our understanding of the phenomenon in relation to others are better than others. These qualities, however, need not coincide in a competition of explanations, thus the need for overall coherence in our explanatory patterns.

Investigating sentience seems to be especially difficult since one can ask whether our usual concepts of human cognition should be applied to animals or whether our phenomenology can be used at all as a heuristic device. On the other hand, the treatment of sentience might be a test case of various trade-offs and checks between, say, philosophical definitions of mental terms as to be applied to

[3] There is an extensive literature about this in the philosophy of mind and language, and I will not further discuss the issue here, cf. (Davidson 2001), (Strawson 1959: 87-114) among others.

animals, neurophysiology, our reflected intuitions and ethological model building based on a computational theory of animal minds. Concerning animal behaviour the question reoccurs what should be taken as the *best* explanation of the behaviour. For example, the intuitive intentional description of a bug is given up as unwarranted anthropomorphic given that the bugs behaviour can be simulated by a little robot, which certainly is no intentional system. The neurophysiological guideline to look for human like neurophysiological structure excludes non-vertebrates as candidate for sentience but is disregarded with respect to cephalopods since they exhibit intelligent behaviour (e.g. in a maze). Our mental terms as applied to humans and tied to the human phenomenology set the agenda for looking for animal cognitive abilities. There is, however, a first stumbling block on that road: starting from a theory of human consciousness is a dead end for exploring animal minds.

§3 *Not Simply Top-Down*

Turing recognised that the question "Can machines think?" will be decided immediately if one were to start from the then existing understanding of the two terms. If 'thinking' is defined as a genuine human faculty and 'machines' as something like a steam engine the question is answered by definition. Something like that applies here. Consciousness as we know it from the human case has a highly complex structure. Especially, it can be argued that consciousness in the human case is identical to or not separatable from self-awareness.

In a theory of consciousness we start from a phenomenally given: our flow of consciousness. A theory can and must appeal to this given. At some point phenomenological observations are stated. The interpretation of these observations and the descriptive categories may be controversial and topic of a debate, but these observations enjoy supremacy over more theoretical models of consciousness. A mere theoretical model of consciousness elucidates some conception of how consciousness may be structured or even how consciousness comes about. Such a model has to incorporate the phenomenological description of consciousness. It carries the burden of proof in case of criticizing a phenomenological description. Phenomenology might secure *explananda* by establishing wildly shared description of being conscious. At the level of the phenomenally given the

request for further arguments hits rock bottom. Descriptions are debatable and may be improved upon, but then criticism has to offer a better description, not a conceivable theoretical model.

Traditional theories of consciousness point out correctly that in consciousness I am *aware* of what *I* am doing, i.e. of me as the thinking agent, although typically the focus is on the content state of affairs not on my thinking, consciousness is transparent in coping with the world. This transparency, however, does not mean that consciousness is anonymous or that one – who? – only later discovers that there is a subject of conscious life. The question who is undergoing the experience has always been answered.

The agent is non-focussed (non-explicit) content in many states. ('implicit' may be misleading because of the notion of 'implicit knowledge', which is not conscious at all.) Other states bring the thinking agent *into focus*. This focussing is *not* the introduction of higher order thought. It is a shift of focus. Kant captures this with the 'I think' that *can* accompany all thought (i.e. one *can* focus on oneself), but it need not, as typically consciousness is not *focussed* on its agent. That the agent is not focussed does not mean that the agent is absent from conscious experience, it means that in transparent intentionality the agent is given 'only' as background aspect of consciousness, which can be brought out of the background by shifting focus.

In contrast to shift in focus higher order belief turns the agent of consciousness into an intentional object or part of the objectual content of a conscious state. In this state the agent of consciousness is part of the objectual content of a conscious state and at the same time is conscious of himself as *having* this higher order state, i.e. knowing that the content is about himself. [4]

Thus

(mere) representation ≠ consciousness

higher order thought ≠ consciousness

[4] The thinking agent in its role as thinking agent has no specific gender; pronouns can be used at will.

Most conscious acts are straightforward (i.e. focussed) on the world, not the act or agent of the act. Consciousness is thus mostly transparent. A *shift of focus* on the agent of the conscious act is not a reflexive act (not generating a state of higher order).[5] 'Higher order' respectively 'reflexion' is a notion of structural embedding or iteration. Shift of focus is a different operation. Phenomenological descriptions of self-consciousness try to verbalize what we experience under this shift of focus. To describe higher order states differs from that. Higher order states are typically involved in deliberation or belief updating. They posit conscious acts/states as intentional objects.

Building a theory of consciousness involves reflection as higher order state, the observational phenomenological basis of such a theory may not, or only when the theory turns to make higher order states its topic.

Consciousness is monolithic in the sense that one can have and be conscious of higher order states (embedding states of the same or different kind), but one cannot have *a second* consciousness within one's consciousness.

In

(C) I am conscious of me believing the pen is black.

the "I am conscious of" represents the *occurrent* process of thought with the thinking agent. This agent is represented as "I" since it is to be identified with the "me" in the *content* of consciousness. The 'me' has a phenomenal quality as representing the agent of thinking. It constitutes the 1st person perspective; in nowadays parlance one might say: 'the what it is like of what it is like'. In it the agent of consciousness is *conscious* agent.[6] The expression "I am conscious *of* me ___" articulates that the thinking agent (the subject) is aware of itself as involved in

[5] Husserl at times speaks of an aspect of consciousness as 'conscious but not recognized'. Recognition in this sense is not reflexion. Sartre distinguishes between 'pure reflection', which shifts focus from 'impure reflection', which makes consciousness an object of thought [see next note].

[6] In Sartre's *Being and Nothingness* notation the conscious agent is conscious "~~of~~ itself", i.e. immediately, not in an act of reflexion (i.e. a higher order state, or in Sartre's terminology 'improper reflexivity'). One may with Sartre distinguish threefold: phenomenal subjectivity, which is present at itself, 'pure reflection' (shift of focus to an explicit conscious agent), and higher order beliefs ('impure reflection').

this or that act of the thinking agent; it does not articulate itself as just another *object* – not as "I am conscious *that* I ____". It articulates itself as relating to itself as the agent of those acts.

Every consciousness has self-access be it focussed or not. There is no 'ego-less' or 'agent-less' consciousness. In straightforward conscious states the subject is not in focus. In a limited sense one can thus say that 'the given is subject-less', a formulation which stresses the perceptual objective contact to reality, but is misleading in the philosophy of mind. Every conscious act involves the 1st person perspective in which the thinking agent is present.

The activity as activity (something being non-static) cannot be *exhausted* by a representation (something static). Self-access possesses the quality of a lived experiential/phenomenal process.[7]

There is no way around this difficulty of the unity of self-access and activity by appealing to reflexion, as reflexion involves the distinction between reflecting (active) and reflected conscious agent (object of the reflexion). This may be a unique difficulty characterizing the thinking agent. Nonetheless one must try to work out some form of expressing self-access, because without some way of expressing it, the thinking agent becomes ineffable, and the theory of consciousness shorter than it already is. This self-access and -awareness is a moment/part of the conscious act. As such it can be made the object of a description and a theory. One cannot, however, make it the sole object of a conscious act, inasmuch as it is always *accompanying* the consciousness of something. Thinking about self-

[7] Again in Sartre's terminology: the agent is always 'exstatic'. Fichte in later versions of his *Wissenschaftslehre* (1794) tried to capture this difficulty with the image of the Ego being 'an activity with an inset eye'. Natorp in *Einleitung in die Psychologie nach kritischer Methode* (1888) declines such images and puts the Ego *completely* beyond representational content, and thus beyond being represented, apart from the minimal representation "the x that does the representing". This does not seem to match well with phenomenology as we experience ourselves *as* agents of thoughts. Natorp does not decline that in every conscious act the relation to the Ego is phenomenally given, he declines that this relation can be made the sole *object* of the occurrent thought. He emphasizes in this way the *unity* of the conscious act as not consisting of a consciousness of the object and a *second* consciousness of the conscious act. One can, however, understand Natorp's thesis in the vein of Sartre's claim that self-consciousness is void of *worldly content*, and thus in this sense contentless.

access – presumably in natural language – makes self-access the object of thought, but in this thought the occurrent self-access of the agent of this thought is not objectified (apart from being one of the species the thought is about), but exstatic (alive).

Even if the acting Ego is always something exstatic (and thus in a sense is always escaping being completely objectified) this does not preclude developing a theory of this process and its structure. Compare: a theory of time, which is a theory of a process – at least in the 'tensed' view – which itself takes time to express and comprehend, but nonetheless covers time and temporal entities (like expressing the theory itself).

Self-consciousness being *sui generis* (i.e. not a subject-object relation, not a reflexive relation, not an identification with an already somehow known self or 'I/Ego') can hardly be expressed in our common language of intentional states and propositional judgements. Attempts to point to its *sui generis* status have to use otherwise *aporetic phrases* like "immediate self-awareness" (sounding like an awareness *of* a self, which nonetheless should be immediate), or "non-relational/non-thetic self-consciousness" (which nonetheless is awareness the thinking agent has *of* himself). Self-awareness has features that highlight that notwithstanding its non-relational and immediate quality moments of it (like awareness, agency, subjectivity) can be distinguished and have to be present in an unbreakable unity.

One may say that immediate self-awareness is 'ego-less' highlighting that this awareness should not be modelled on the subject-object scheme, for all the traditional reasons based on different regress arguments. In that way of talking any object 'ego' is transcendent with respect to immediate self-awareness. Nonetheless does immediate self-awareness include knowing *of* the thinking agent in the way that the idea that somebody else might have these thoughts does never arise. It is preferable, therefore, to speak of the thinking agent as 'Ego', and of immediate self-awareness as a state in which the Ego is *immediately* (not mediated as an 'object') *present to itself*. This very state seems *sui generis* as it cannot be modelled on the form of ordinary intentional acts – leaving moods, to be another problem, to the side for the moment. Its extraordinary nature lies at the heart of many convolutions in theories of self-consciousness and many of their *cul-de-*

sacs. It being *sui generis* stands in the way of explaining it within a theory of consciousness. It seems to be beyond theory. One can, however, describe representational forms of consciousness in which occur some special representations and representational combinations, like "I am conscious of me ___", sometimes – only slightly sarcastic – conceived as an 'I-symbol' the tokening of which results in the presence of immediate self-awareness. Talking of an 'I-symbol' also elucidates that self-consciousness *does not create itself*. The tokening of the 'I-symbol' gives rise to a conscious state, but the symbol does not do the tokening. In a person as cognitive system certain symbols are tokened into complex representations, some of which are the representations underlying or giving rise to consciousness (cf. Bremer 2006a).

Human consciousness cannot be separated from self-awareness and inner speech. Once you look at this complexity which has to be preserved whatever theory or approach you favour, I take it to be absurd to ascribe this structure to even highly developed mammals.

We cannot put ourselves in a state of mind which corresponds to some kind of sentience below human consciousness. Whether there is something like that cannot be decided phenomenologically. We should at least assume that there could be something like that to leave the question of animal sentience open. But methodologically we have to keep in mind that it would misfire to list the features of human consciousness and then set out to find these in animals. We should start from animal cognition in the diversity of cognitive/mental faculties. Human cognition contains – related to consciousness – faculties like forming beliefs and desires, applying concepts, using language, modelling the action of others etc. If animals have *something like* these faculties they might have something like consciousness, namely sentience. So, it might be more promising to investigate animal cognition bottom-up then top-down (starting with 'consciousness').

In general, it might turn out a fruitful attempt to start with human cognitive faculty *x* and see whether animals have *something like x*. From this *something like* way of rendering things an appropriate terminology of (animal) ethology can take off (cf. Bremer 2007, 2008a) This *bottom-up* approach may concern itself first with the so called 'cognitive mind' (i.e. faculties which need not be accompanied by consciousness).

Whereas with respect to cognition a bottom-up approach seems preferable there is some rationale for a top-down approach with respect to neurophysiology. Given empirical knowledge about neural correlates in the human case one can either look for similar structures in some animal species or, in case of absence of these structures, look for other structures that may (their role embedded in a complete evolutionary account) be reasonably be assumed to play a similar role. In these cases the argument from structures of similar function to similarity in phenomenality will, however, be even weaker than in the cases of similar structures.

§4 *The Neurophysiological Paradigm*

One paradigm case argument is the argument from similar brains. A lot of research in neuroscience depends on the transferability of observations on animal brains which diverged evolutionary from human brains millions of years ago. Still neuroscience assumes that individual structures (like the visual system) are similar enough to learn about human vision. Mice models are used for human neurodegenerative diseases, but these models need not involve ascription of sentience. Animals are also ascribed experiences of some type because they have a brain structure that resembles a structure in human brains, where activity in that part of the brain is correlated with experiences of that type. For us and mammals: "The brainstem structures are highly similar, the thalamic organization is highly similar, the structures that regulate emotions are highly similar, as is the cortical organization.", which with other evidence on sleep suggests "that during the waking state there are also fundamental similarities in our experiences" (Churchland 2013: 249-50). So, given their brain structure, at least mammals should be sentient.

On the one hand this sounds convincing as (i) one should assume that same causes in a similar environment have same effects – or at least effects similar enough, loosely speaking, of course, and (ii) human evolution supports the idea of gradual build up of human cognition, sentience, and – maybe – finally consciousness.

On the other hand these ideas expressed thus have two obvious shortcomings: (i) the argument by similarity need not be transitive, otherwise we have slugs with human capacities because they are similar to an animal which is similar to an

animal, and so on; maybe the relevance of similarity stops rather early (say with primates); (ii) the very idea that consciousness or sentience come in degrees defies imagination: What should it mean to be partially sentient? Should we imagine a diming down like with an electric lamp? The steps to sentience and then to consciousness are qualitative steps, which means there is an abrupt change between two ways of life. A minor change in human brain evolution could have made all the difference.

The use of mice in neuroscience may be reliable with respect to information processing and basic forms of learning, but tells us little, if anything, about sentience. Studying a brain structural element in abstraction from overall cognitive architecture and its resemblance to the human case may yield many physiological insights, but has no relevance to issues of awareness. Moving to monkeys and apes can tell us more, but inasmuch raises moral concern [cf. §26].

So, arguments of the brain similarity paradigm have some force, but are not knock-down arguments for respective animal capacities. They have to be weighed against other arguments concerning the presence or absence of the capacity in question.

§5 *The Behavioural Paradigm*

Another paradigmatic argument type is the attribution of some form of sentience or cognitive capacity as explanations of animal behaviour. This type of argument infers the capacity from the observation of behaviour as the presence of the capacity would make the behaviour both feasible and appropriate to the animal's needs.

Such arguments are plausible as *inference to the best explanation* and as part of a coherent outline of an animal's mental life. (Often they have a part of the 'similar to human behaviour' paradigmatic argument.)

Such arguments are controversial inasmuch as the behaviour in question may be explained without invoking the hinted at capacity. Sparse explanation should be preferred in general, and especially in an evolutionary context, where adaptations are sufficient to keep an animal alive but rarely overshoot the needed capacity, although such cases may have occurred. Further on, the ascribed capacity has to

fit in into the overall observed behaviour: a capacity that would be immensely helpful in another type of situation faced regularly by the animal where the animal does *not* show a corresponding versatility and competence may be missing in both situations. Anecdotical evidence has been put forth to ascribe advanced cognitive capacities (like mind reading) to animals which are otherwise incapable to resort to these capacities when urgently needed in other situations.

A bird feigning a broken wing need not read the mind of the fox endangering the nest. If the bird had such higher order intentionality, we should see more properly conventional behaviour (requiring nested mental attributions) in bird colonies, with local cultures etc. And we do not observe the presence of the useful supposed capacity here.

A more parsimonious explanation being available does not prove the absence of some capacity, it could just be there as a matter of fact, but it shifts the burden of proof to the ascription.

An isolated capacity ascription that does not fit into the overall profile of the animal in question seems always to be a theoretically very weak proposal.

The similarity between the behaviour of many animals (especially with respect to pain) and human behaviour seems intuitively striking. This *intuitive* certainty does not establish animal sentience in any *scientific* way, but it shifts the burden of proof to any theory doubting or even denying sentience in these animals. Arguments against animal sentience in these cases have to be as strong as our natural inclination to *perceive* some animals as sentient, it seems.

§6 *The Appeal to a Certainty Intuition as Paradigm*

The assumption of sentience within at least the strata of mammals and birds (cf. Humphrey 2022) seems unavoidable given our *intuitive conviction* that these animals are sentient. We 'just perceive' that they are sentient. This is one of our almost unshakeable beliefs about pets, domesticated animals, and at least larger animals in the wild. This is a 'certainty' in the sense of Wittgenstein's *On Certainty*: something we believe *without a need for further argument*; its certainty lays somewhat between the even more certain "I am not made of wood" (given

the literal meaning of the words) and the minimally less certain “The Earth is not flat”.

The paradigmatic argument *against* animal sentience in general proceeds by coupling sentience to consciousness and this to natural language (e.g. Macphail 1998). As we have a – more or the less – clear conception of consciousness and get rid of the methodological conundrums in the search for animal sentience this position can claim coherence and some parsimony in putting mentality in the world. As human culture evidently testifies that consciousness has survival value its evolutionary function (for the survival of humans) cannot be questioned, even if the side effects of our technological way of life might endanger human species survival in that current form.

In contrast many animal capacities can be understood in their role as adaptations without conceiving of these capacities as executed with an inner perspective (i.e. sentience).

Sentience has to start somewhere in the tree of life, maybe it is with humans! To locate its onset earlier in the tree of life needs either or both an argument about its evolutionary function with these animals or an argument why it occurred and than stabilized as a side effect of some other adaptive capacity.

Nonetheless, this denial of animal sentience seems to be outrageous – already in epistemology, even without considering the moral implications of this view for treatment of animals in farming and experiments, implications endorsed by some authors (e.g. Fox 1986, Leahy 1991).

The assessment of this theory of human *complete* exceptionality may in parts rest on weighing our trust in science against our trust in our manifest world view, given that we may very well err in science, as the history of science shows. So, even if we do not have no convincing – knock-down? – argument against complete human exceptionalism, we may defer the *justification* of ascribing sentience to some animals to a future state of science. And we may morally rather err on the side of caution and consideration, at least with respect to those strata where the argument of similarity has some hold. We should concede at the same time, then, that ascription of sentience to other strata hangs on very weak arguments, typically a problematic extension of the paradigmatic arguments from mammals

and birds even to insects. A weak argumentation for animal sentience for some animals constitutes *a test case* for theory construction in cognitive science.

One paradigm method in the philosophy of mind is the phenomenological description of structures of consciousness. Descriptivism of a wider kind is a paradigm in the philosophy of language and epistemology. Descriptivism aims to set out our linguistic life world with its constitutive practices. Descriptivism claims that these basic structures and their accompanying beliefs are not in need of explanation, but only of proper description, which also blocks inappropriate quests for further justification and (pseudo-)philosophical scepticism (cf. Strawson 1985). According to Wittgenstein philosophy is merely descriptive. There are lots of passages in the *Philosophical Investigations* stressing this point, e.g. "It leaves everything as it is." (§124) or "All explanation has to go, and description has to take its place." (§109). Descriptivism claims that philosophy describes regularities. Philosophy lays open the facts that speakers naturally behave in this or that fashion. The basic structures of our intelligent behaviour are just read off from an exact description of our linguistic behaviour. And these structures are justified by the fact that the practise which exhibits them is successful. Alternatives (including alternative philosophical claims on intellectual standards) stand on a far less firm ground by not being entrenched in our successful way of life ('life form' as Wittgenstein might say). Descriptivism does not exclude viewing linguistic behaviour as normative, as a described norm does not cease to be a norm by being described.

Descriptivism can be applied to our attitudes on animal sentience, especially the certainty contained in our manifest image of the world that mammals and some birds are sentient.

Outside some philosophical seminars – hopefully also inside most philosophical seminars – scepticism of a radical kind should not be taken seriously. The strength of our justifications need only be so strong as the type of knowledge or situation requires, as already Locke (1690, §§IV.I-IV) insisted upon. In a context of everyday life ("There is marmalade in the fridge") we do not need mathematical proofs for our claims, but trust our senses and testimony. Principle of sufficiency make us trust by default in reliable sources (like perception). Principles of credulity demand that to question a belief already held the critic has to provide a reason

for doubt. Radical sceptics hardly can provide such reasons. A sceptic who spells out a reason for doubting (e.g. the possibility of warranted belief) has to presuppose what she wants to be sceptic about and catches herself in contradictions.

Some of our certainties are beyond doubt. Talking is cheap, but not believing: it is easily said "I am made of wood", but given the literal meaning of the words, it cannot be believed. That I am not made of wood is a certainty I need no reason for to believe in, and I cannot meet a justified doubt that it is not true. Wittgenstein argues in *On Certainty* that our life consists of many such certainties. We can only attest and describe them. It is to the demerit of philosophy that some philosophers – in their proverbial arm chairs – ask for reasons for such certainties. These certainties constitute our lifeform. Scientific questions arise elsewhere.

And, again, one such certainty – for most, if not almost all, people – is that grown mammals and some birds (like ravens) are sentient. We believe this already as children without a reason. We typically do not ask for a justification. Only extraordinarily good philosophical and scientific arguments could make us doubt it. The arguments against animal sentience even in grown mammals may be strong, but I do not think that they are strong enough to doubt this certainty. And given the account of certainty alluded to above, I also do not think that this is just a naïve belief. Rather it is a certainty that needs elucidation, and this, not the certainty itself, poses many problems, as discussed in this book.

The certainty on sentience should not be confused with the claim [rejected in §1] that animals are *conscious just like we are*, or the rhetorical question how we know that they are not. This is not certain, but plain wrong [for all the reasons set out in §3]. It may be a childish belief, but we learn quickly that animals are quite different from us. We are certain that 'there is something going on' in a mammal's mind, *not that it thinks like we do*.

§7 *The Evolutionary Paradigm (I)*

Evolutionary reflections provide another (sub-)paradigmatic argument for sentience from analogy to the human case: "[G]iven evolutionary theory and given the demonstration of the survival value of consciousness the human case

provides, we have every reason to suppose that the members of other species also are conscious." (Regan 1983: 29).

This takes up the idea of identifying a functional role for sentience. The crucial question is: If sentience could plausibly be seen as an advantage in behaviour, are we *entitled* to conclude to its presence? Are we entitled to the Panglossian assumption that evolution provides everything that could be helpful? That goes against evolutionary theory. Survival under competition uses the physiological cheapest and easiest available means for advantageous behaviour. For the leap from non-sentience to sentience we either need a very good evolutionary explanation, or we believe sentience to come about as a side-effect and exaptation.

Another variant of evolutionary reflections points to the 'conservative character' of evolution: if something works good enough, it will be sustained without a need for further improvement. We can find, therefore, organs and brain structures of human physiology also in other mammals, at least. This will have to be weighed against the differences [to be seen], but provides a *prima facie* argument against too sudden changes in sentience.

§8 *Dualism and Sentience*

By the way: Suppose dualism was a tenable position (cf. Hasker 1999, Swinburne 1986). Dualism of emergent or independently existing substantial souls, although a minority position today, may be taken to present a more coherent (metaphysical) package than some other alternatives to mind/brain-identity.[8]

[8] Worse than these dualisms are, for example, 'property dualism' or panpsychism (e.g. Nagel 2012) and related approaches (like Reber 2010). Property dualism depends on a commitment to an elaborated metaphysics of properties and physical/psychological-laws, both of which may be questionable. It contrasts unfavourably to Anomalous Monism (as a form of a token identity theory) with which it shares the insistence on the irreducibility of psychological vocabulary. Panpsychism and related approaches can hardly be taken seriously: they posit sentience where we lack any scientific reason to assume it. They fail to recognize or to explain missing sentience in simulations run on PRAMs or by robots, which possess as complex architecture as some of the supposed sentient beings. They would become bizarre if, biting the bullet, they believed such devices to be sentient. Anti-functionalist biochauvinism is another such postulate-based mind metaphysics. All of them will not be considered here, because the problems they raise

The debate about animal minds had to be rewritten for dualism, in intricate ways.

For example, if Descartes' distinction in the *Meditationes* between the rational soul and the experiential mental life of the body was correct, animals could share with humans this – properly so called – 'animalistic soul', which meant that animal sentience consisted of states comparable in phenomenality to human experience. On Descartes' conception animals need not be automata.

On the other hand, if dualism worked by the attachment of 'soul stuff' to a sufficiently complex brain as receiver, small amounts of soul stuff could attach to less complex brains – and the debate whether this amount of soul stuff suffices for sentience might start all over again, this time under the description of 'complexity of the receiver', but in empirical matters equivalent to materialistic research.

If a dualistic world view – or a materialistic one for all that – entails a plan God pursues for this world, then one lands in theological debates about the role of animals in the plan of salvation. If God pursues his plan for humanity by evolution – as the Catholic Church admits – then everything in the debate about animal sentience can keep in place.

CS mostly works on the presupposition of some form of materialism. The intricacies of starting from some version of dualism can wait for another occasion, because dualism either leaves the empirical debate almost in place or it make the question of sentience only more difficult.

One supposedly easy answer to the question might be seen in religious beliefs of re-incarnation. Re-incarnation into an animal only makes sense, if at all, if one experiences something, one may presume, otherwise one will not recognize the impact of one's Karma. Thus, animals are not only sentient, but *conscious*! Re-incarnation beliefs of this sort not only fly in the face of cognitive ethology, because animals just do not behave as if they were re-incarnated humans with a consciousness reflecting Karma, whatever all this means, if anything. They are morally inacceptable as well.[9] They will not be taken seriously here.

---

with respect to sentience are of their own making, asking for metaphysical solutions CS cannot offer.

[9] Not to be argued here, cf. Edwards 2002.

§9 *The Non-Conscious Capacity Paradigm*

A paradigmatic argument type *against* animal sentience proceeds from human non-conscious capacities. Humans can learn and show flexible behaviour *without* these being conscious. Why then – asks (Macphail 1998: 138-74) – should similar capacities in animals be accompanied by sentience?

Natural language use is supposedly the most complex human cognitive capacity, and although our use of language is conscious, and semantic and pragmatic knowledge are explicit, the derivational syntactic system behind our language capabilities (the rules and representations involved in the mental derivations and computations) are, according to GG (cf. Chomsky 1980) not conscious. Why should less complex capacities be mentally accessible? In each case an argument is needed.[10]

Another paradigmatic argument type *against* animal sentience proceeds by pointing to similar types of learning and behaviour in animals which are prototypical candidates for sentience and in animals which by their neurological structure are otherwise no candidates for sentience.

Whereas the animal advocates are prone to extend sentience then down to, e.g., insects, the critic stresses that absence of warm bloodedness and absence of a CNS at all or of more than simple structure speaks against sentience in these 'lower' animals, and that *therefore* the seemingly intelligent behaviour indeed solves problems *as if* intelligent in the environment of the animal, but does not involve sentience. So why should this behaviour then be a criterion for sentience in the 'higher' animals?

[10] A capacity or skill not involving attentional focus or higher order thought need not be non-conscious. The criticism of RTM that going back to the phenomenological tradition appealed to the non-representational motor intentionality of 'skillful coping' (Dreyfus 2014), even if correct, considered skillful coping as "a kind of awareness". Some popularized versions of this are at least verbally confused when claiming, for instance, that driving home from work was done "unconsciously" or "without consciousness"! What they mean cannot be that they drove in a state of deep sleep – quite a feat – but that the driving was not in attentional focus or reflected upon. Skillful coping by itself, thus, provides no basis to deny sentience in skillful coping animals.

The argument does not exclude the possibility that despite the similarities the behaviour in the 'higher' animals is accompanied by sentience. The argument appeals to simplicity: an explanation for the behaviour of the 'lower' animals not involving sentience can be extended to other animals, and thus, *prima facie*, it should be extended to them, thus denying the attribution of sentience.

The overall basis of the argument lays in an account of insects or other 'lower' animals that explains their behaviour mechanically (e.g. by a mere physiological reflex model) or by building artificial insects, which clearly lack sentience, showing the same behaviour. Given a mechanical model of the behaviour there remains no evolutionary *rationale* for the presence of sentience, even if one has a general idea of the adaptive function of sentience. Given the cost in energy and maintenance of a brain an attribution of epiphenomenal sentience is no option.

A theory of insects and their seemingly intelligent behaviour thus can be a major factor in a theory doubting widespread sentience.

§10 *The Pain Paradigm*[11]

Pain and suffering are *prima facie* bad things. They are widespread. Both the *phenomenon* of pain or pain behaviour as well as the *concept* of pain play a fundamental role in investigating animal minds, especially with respect to attributing sentience to animals. In distinction to other mental states the occurrence of pain seems obvious. The concept of pain seems to be a paradigm case of functionalism. One may argue that the occurrence of pain or pain behaviour are inseparable from the occurrence of sentience in animals and from the supposed structure of animal cognition, and argue that pain is the paradigmatic case by which we – using a broadly conceived functionalist account of the mental – understand the workings of an animal mind from without and from within.

Pain as we know it from the human case is a phenomenal state of mind. A state for which it is *something like* to be in that state. Pain occurs *only* as a phenomenal state. You do not have to tell yourself "I am in pain" to be in pain, supposedly having pain does not require the possession of the concept of pain. 'the

[11] Some material on pain in §§10, 11, 13 is taken from (Bremer 2005a).

phenomenon of pain' means the occurrence of pain states and/or the occurrence of states, say in the brain, that correspond to these phenomenal states, and/or behaviour that is linked to these states. In some cases, it may be useful to use the tautological expression "phenomenal pain" to emphasize the pain as felt in contrast to the pain behaviour.

Pain plays a crucial role in attributing sentience and withholding attribution of sentience (and pain feelings) to insects. Starting from the human case of pain we derive two sets of criteria for attributing sentience (as being evidenced by being attributed pain):

(i) Criteria based on similarities of neurophysiology, pain being the paradigm case for establishing mind-brain-relationships;

(ii) Criteria based on pain behaviour, pain being a phenomenal state with a distinguished set of related behavioural symptoms.

Pain might not provide us with a knock-down argument for animal sentience, but it provides us with clues.[12]

In human neurophysiological architecture pain is channelled to the brain by way of nociceptors. Noxious stimuli applied to innervated body parts yield a message to the brain. The brain should be caused by this to initiate some motor activity protecting the inflicted tissue. In humans there are nerve endings in most bodily tissue that respond to stimuli like pinching that usually cause pain. Nociceptors respond to chemicals produced by damaged tissue, as well. Second-order neurons in the brain contacted by afferent nociceptors transform their impulses and might signal to further neurons, leading ultimately to pain recognition and avoidance behaviour. These nociceptors we also find in vertebrate brains (as we find there endogen morphine to alleviate pain by inhibiting the neurons linked to the nociceptors, it seems). The brain areas governing pain are at least in parts similar across the vertebrates. We find structurally similar brains to human brains in

[12] For the following paragraph cf. Bekoff 1988: 263-69; Churchland 1998: 40-41, 77, 144, 420-33; DeGrazia 1996: 97-128; Dubner 1984; Fields/Price 1993; Key 2016; Short/Poznak 1992. (Wall 1992) sees an "obfuscation of such terms as ‚pain'" here, since the attribution of pain given some criteria is even less regular in animal cases than in the case of humans (and their physiology or behaviour); he does not deny the phenomenon.

vertebrates (including a CNS). There is also an autonomous nervous system in vertebrates.

From the behavioural perspectives pain behaviour is more easily identified than, say, starting to look for a mate because of arousal. Pain is more easily identified, since often it is caused or accompanied by bodily injury. Pain is unpleasant and, therefore, linked to a behaviour (or attitudinal state) directed at its termination. Vertebrates in general show pain behaviour similar to humans. Rats, dogs, monkey etc. show changes in posture, vocalizing, temperament, locomotion/immobility and respiratory and musculoskeletal systems. They also show anxiety behaviour – supposing anxiety being closely related to pain or expected pain sensations – consisting in increased arousal, tension and inhibition of usual behaviour. Pain supports learning by avoiding the adverse situation or stimulus. As long as an individual reacts strongly to a stimulus and learns to anticipate situation of that kind to avoid them, we have clear *prima facie* evidence to attribute phenomenal pain.

Using animals in pain research obviously presupposes the similarity of human and cat or primate pain; experimental designs in pain research on animals are developed by using stimuli similar to ones painful for humans, and looking for responses similar to those of humans.

In distinction to vertebrates the evidence speaks against phenomenal pain in case of insects. In insects we find no CNS. Compared to humans and vertebrates in general there are quite different brain structures in insects (and some fish). Insects lack the central processing mechanisms of vertebrates. They do not have a nerve fibre system equivalent to the nociceptors we find in vertebrates. Insects notably lack pain behaviour which protects injured body parts in vertebrates (e.g. a spider losing a leg). If they show behaviour given noxious stimuli, it can be classified as a startle or protective reflex (involving no central mediation). Pain might be absent in insects because there is no selective pressure to protect the individual in contrast to the kind given the short life span of insects. Simple neural reflexes fired by a noxious stimulus might be sufficient for a species of short-lived individuals.

To sum up: If you look for sentience in an organism, look for pathways of pain and complex pain (avoiding) behaviour.

§11 *The Evolutionary Paradigm (II)*

There seems to be a simple evolutionary syllogism of attributing pain sensations to animals.

1. Pain has developed and has been maintained in evolution since it has a function.
2. When we look at our studies of evolutionary development, we see that each feature which developed had some precursors in evolution, different in some degrees.
3. Therefore, we may assume there to be, at least, pre-forms of pain in animals.

Assuming (1) to be right and notwithstanding the problem that pain might be the exception to the inductively established rule mentioned in (2) the main problem with this kind of reasoning is that it presupposes that phenomenal states (as we know them from the human case) admit of degrees. But seen from human phenomenology this seems to be straight wrong: consciousness is an all or nothing affair; you might be dizzy or drunk, but you either are conscious or you are not. There is no fading or flickering of consciousness (cf. Chalmers 1995). So, if the possession of phenomenal states does not come in degrees, it might jump from an evolutionary mutation without precursors. If there are different *kinds* of pains that might be another affair, but they do not differ in degrees of sentience.

From an evolutionary perspective we have to find a function for each cognitive trait of an animal. Only because it is adaptive to some problem did the trait survive. If benefits outweigh the costs an attribute can evolve and be incorporated into the living system. Taking the *Darwinistic stance* is to attempt some reverse engineering (cf. Dennett 1995: 48-60, 187-228): a trait occurs as a solution to an engineering problem relative to the organism's environment. We understand a trait by seeing how the creatures having it are better adapted with respect to a challenge posed by their environment. Looking back (therefore it is *reverse* engineering) we recognise why a trait is built into a system. An evolutionary process must also subserve the maintenance of the behaviours and structures that evolved. Within this evolutionary approach we try to explain what the evolutionary function of being sentient is. The phenomenon of pain might be crucial to make this

clear. There has to be something like pain in an organism with multi-modal sensory input and plasticity of behaviour: There are (following Plotkin 1994) *primary heuristics* like the structure of our body which embody knowledge about our environment (the knowledge of gravity is *built in* in our shape). This hard-wired or built in knowledge can usually not be accessed. It is given once the organism is there. *Secondary heuristics* are a little more adapted to changing situations. An example are plants which follow isolated features of their environment (e.g. a sunflower 'moving' the head with the sun). The secondary heuristic discriminates one feature of the environment, but it cannot be used to look for other things. In contradistinction to this animal behaviour shows *plasticity*. Animals possess *tertiary heuristics*, i.e. they have the ability to extract information out of a changing environment. A cat can adapt its foraging behaviour to new types of situations (places where it has never been or chasing animals, say newly imported guinea pigs, it has never seen before). This has to involve *representations* of some kind. It was evolutionary required to stay alive in an unpredictable environment (showing 'predictable unpredictability'). Pure instincts cannot deal with that. The animal with these representations has multi-modal input (to achieve a more coherent representation of its surroundings) and tends to some stimuli according to its learning history and its current goals (like feeding, mating, looking for shelter). Once behaviour exhibits plasticity there might very well be an evolutionary function for pain as setting priorities between inputs: "Mind the pain first!"

Being in a pain state has the obvious advantage of tending to immediate (bodily) problems within a highly complex environment. Multi-modal input and all the corresponding enticements given the animals goals could very well be dysfunctional if there was no alert system setting priorities. Pain is part of that system.

Evolution of a kind itself might involve pain *as building block*. A new kind evolves by mutation and selection of a more fit phenotype given the environment the animal lives in. Selection here is short hand for the elimination of the unfit. In case of sentient creatures the evolution of higher (more complex, more versatile) types of creatures presupposes that some forms that do not optimally fit the species' environment die out. Those which die are, of course, individual animals capable of feeling pain: Dying out involves pain. Without massive ancestral pain there would be no higher animals or humans around. If the theory of evolution is

our best theory to explain why we and other complex animals are around, the concept of pain is an indispensable part of that theory, pain – respectively – is an indispensable part of the world described by that theory.

§12 *Animal Attitudes (I)*

Talking about animal cognition usually does not start with talking directly about an animal's states of sentience, but rather starts with ascribing attitudes or states like attitudes as we know them from the human case (beliefs, desires, fears…) to animals. We say "the dog *feels* pain" ascribing sensation. We notice "My cat *wants* to get in the kitchen because *she thinks* there is some cheese left" ascribing beliefs and desires. Looking for intentional/propositional attitudes in animals seems as obvious as looking for sensation or sentience in general, but is confronted with a situation like the one with respect to consciousness. We have a highly complex model of propositional attitudes in the human case (cf. Davidson 1982, 1984) which involves capacities that make it highly unlikely that animals have beliefs and desires *in that sense*.

Humans have beliefs, higher order beliefs, and employ the concept of belief in assertion and in updating their web of beliefs. The concept 'belief' is tied to rationality assumptions concerning how beliefs have to be justifiable and interact with desires in rational action, interpretable and criticisable by rationality standards by an interpreter. Who has beliefs in this proper sense has to be rational (by capacity) and possess the concept of norms of rationality to be followed. Therefore, "belief" cannot easily be displaced by allegedly better (neurophysiological) terminology of CS, as has been debated under the title "the future of folk psychology" (cf. Greenwood 1991). Animals do not command the concept of norms of rationality to be followed nor the concept of belief, thus they *do not have* beliefs in the proper sense.

The *intentional stance* (Dennett 1971) can be adopted towards systems that do not have intentionality, but which can be described for some purpose as having it. In these cases the intentional idiom is employed only as a place holder for an explanation to come at the design or physical level of the system. You can talk about an ant in intentional terms: "The ant *wants* to get to the food and confronted

with the *choice* between two paths *it believes* the right path to be the better." There is, however, a sufficient explanation at the *design level* of the ant, since ants are controlled by olfactory input: An ant looks for food that gives more energy than needed to get it, and confronted with two paths the shorter one will have, after a while of use by co-working ants, more ant scent, so the ant takes it. There might be animals in case of which the intentional description is the simplest or even the only one we have so far. Reduction to the design level might be possible in the future only. And furthermore, there is a crucial distinction between build in intentionality (i.e. control of behaviour by some computational level that the system need not be aware of) and intentionality coupled with awareness of the intentional state. So, we may to be able to interpret the mouse in intentional terms, and maybe the mouse is a computationally controlled systems, but that does not settle the question whether the mouse experiences states with different intentional content. Humans do, since they can represent their intentional states in language. Complete reduction is wrong headed in case of such systems that describe themselves *as* intentional, even if we could revolutionise the intentional idiom. So – is it *like something* for the dog, ape… to be in the state *we* describe as "belief" or "desire"? The *instrumentalist* attitude akin to the intentional stance is not – apart from being a heuristic – an option for a *realist* cognitive science not only including ethology but also neurophysiology and phenomenology. Now, sensational states like pain might account for the presence of real belief like states in animals. Sentience, which we commonly ascribe at least to mammals, must be connected to states of 'recognising' and 'doing' since otherwise there would be no point in having it (cf. DeGrazia 1996: 129-36). Pain would be entirely otiose if you could not do anything about it. Animal research – seemingly presupposing this connection – employs "behavioral animal models which utilize an operant escape response to electric shock as a measure of pain" (Dubner/Beitel/Brown 1976: 156; cf. Dubner 1984).

If we know or can justifiably believe that an organism is able to feel pain, this is at the same time reason enough to assume that this organism has some *like* attitudinal states, states which are directed at the objects which have given rise to the sensations. These states need not be beliefs and desires in the full human sense, but we often can explain animals using belief/desire-psychology, so the states they possess have a *similar role like* beliefs and desires. Otherwise, explanatory

power within ethology would be lost. Desire like attitudes regulate behaviour within an *experienced* situation, so it would be queer if it was nothing like to have them. There might be a tacit level of information representation that supports ascribing something like beliefs to animals. After all, applying belief/desire-psychology to animals seems to be successful. In that case the logic rests on the side of the ascriber, who is human, of course, and is merely built in on the side of the animal (assuming a kind of computational level in the animal). Belief like states are not part of the accessible mind of such an animal, say a dog. The animals might have a mental content that as a sensation is tied to some belief like state, feeling "Wow!" in the belief like state with a content like 'That smells really good! I wanna take a look there'. We have no access to this representation. It cannot be like an articulated sentence, but the state a dog is in when expecting food is a state different from the one chasing a rabbit. Maybe these states are not just experiential states, their content might be more structured. So, we should say that those animals which require an intentional description or the behaviour of which requires some kind of belief/desire-psychology have *belief like states* or *desire like states* (BL/DL-states). In any case the phenomenon of pain points us to the presence of these types of states.

That *we* might take the intentional stance towards an animal is insufficient for ascribing real beliefs and real desires to the animal. Having *goals* and *registering* information are animal capacities, but admitting them only suffices for a quasi-intentional explanation, similar to an explanation of the course of a Cruise Missile given its registering information and having goals. Cruise Missiles are not sentient, thus, having and pursuing goals and registering as well as updating information *in themselves* (without further supporting evidence) are insufficient to justify attribution of sentience.

§13 *Pain as Functionalist Paradigm*

A theory of animal cognition involves a model how cognition works, what cognitive architecture it is based on. The concept of pain contains how we conceive of pain, what are the conditions to employ the expressions "pain" or "is in pain" with respect to an organism. One major tradition in cognitive science (the tradition that really started cognitive science [cf. Fodor 1974]) is based on the

Computational Theory of Mind (CTM) or functionalism. Pain is the paradigmatic case to introduce the functionalist's view. This concerns our use of the concept of pain as scientists. Besides that, we can consider whether an animal possesses something like a concept of pain.

Functionalism *defines* a mental state type by its functional relations to the individual's sensory input, other mental states and the individual's behaviour. The whole of these relations defines a causal role. Being a state of type α is being a state with that causal role that defines type α. Depending on the variants of functionalism pain states are either just those states that play the pain role (1st order functionalism) or pain itself is identified with that very functional role (2nd order functionalism). The whole cognitive system is described as a teleological unit at a computational level to explain its behaviour. Cognitive ethology, especially when using belief/desire-psychology, is nothing but a variant of functionalism, employed in this case to derive at a systematic theory explaining animal behaviour. The concept of pain is the concept of a functional state.

Pain is the paradigmatic mental state obviously accompanied by more or less distinct behaviour. Remembering what we have said so far that pain has a definite causal role (i.e. protecting tissue or avoidance behaviour) seems obvious, as well. It is, therefore, no accident that pain is the typical example in introducing or discussing functionalism. Taking pain to be defined by its causal role and being the typical functional states unites critics of functionalism with the functionalists (cf. J. Bennett 1976: §§1-3, 10; Lycan 1996). In the sense of 'paradigm' used here pain (respectively *the concept of* pain) is *a paradigm* for CTM, and thereby in cognitive science. Drawing analogies to pain is a paradigmatic argument for a functionalist account of some mental state.

§14 *Primates as Special Case*

The highest developed cognitive faculties in animals we find in the primates, the great apes. With respect to them we can ask whether they have something very near to a conception of themselves (cf. Parker et al. 1994). Having a conception of oneself, however, presupposes to distinguish between oneself and others as different cognitive or animate agents. So: Do primates see their flock (and maybe

other animals) *as* 'animate'? This may require discriminating them as having states of awareness, which involves on the side of the discriminating animal something like the concept of a phenomenal state.

Non-human primates are quasi-intentional beings (having belief and desire like states), but they do not understand the world in intentional and causal terms (Tomasello 1999). They do not point, show, offer or teach. They learn not by understanding a con-specifics strategy, but by focusing on clues in the environment. They see others as animate (not being stones, being unpredictable etc.), but not as intentional. Humans can take the other's point of view and by internalising the respective communicative encounters use a medium of internal description and redescription of themselves and others. Higher order intentionality enables humans to have linguistic beliefs or beliefs at all. Non-humans do not see the world in terms of intermediate and hidden forces (i.e. causality and intentionality). So, they cannot plan given an understanding of these forces. Nevertheless, seeing another animal as animate attributes phenomenal states to this animal (even if this attribution does not proceed as employment of a belief/desire-psychology). This seems to involve the ability to discriminate phenomenal states in others or even involves the possession of the *concept* of pain (recognising that some other ape feels bad *given* its corresponding behaviour).

Secondly: Primates have been the object of linguistic studies. Over the years there have been several experiments to teach apes sign use or even sign languages as those used by humans (cf. Premack 1976, 1983). Now, understanding a language results from a period of teaching and engagement in language acquiring situations. Language teaching situations involve a teacher, a learner and circumstances which allow the use of the expression to be acquired. This constellation has been called 'triangulation' (the angles being the teacher, the learner, the fact referred to). What does the acquisition of some term require on the side of the learner? In case of observational terms like "banana" the learner has to perceive the object (be in state of perception) and relate the salient feature of the percept (the qualia complex the learner is aware of) to the expression used by the teacher. The learner has to be in some (phenomenal) state, but she need not have any concept of being in such a state. The situation in case an expression for a feeling is acquired is quite different. First of all, what functions now as the object? It might be some other

animal or human pointed at, it might be the teacher to begin with, but at some stage in the learning process the object referred to (as being the bearer of the state talked about) is the learner herself. Acquiring the expression "is in pain" involves relating some observed behaviour (the third person perspective) to some phenomenal state (known from one's own first-person perspective[13]). The teacher is able to recognize the phenomenal state in the learner by observing some corresponding behaviour, and on that occasion conveys that that very state the learner is in *subjectively* shall be called "pain". Mental predicates have this dual aspects semantics (cf. Strawson 1959: 87-114), on which also the functionalist theory of mind is based. Pain is the paradigmatic case of being introduced to mental predicates. By showing appropriate usage of the expression "is in pain" the learner shows that she has acquired (or linguistically expressed) the concept pain. An organism able to use mental vocabulary, therefore, has not only phenomenal states but also (precursors of) concepts of these states. An ape engaging in signing behaviour involving mental vocabulary at least gives *prima facie* evidence of its possession of these second level discriminatory abilities or even of concepts. And that is what some of the language learning apes did. They employed not haphazardly expressions like "anger", "surprise" and even "sadness" (sadness being not only pain, but involving an evaluative component) showing some kind of understanding what it is like for themselves or others to be in these states (cf. Miles 1993, Patterson/Gordon 1993). So basic feelings such as pain might be things that at least the great apes have (something like) concepts for.

§15 *The Simplicity Paradigm*

Simplicity as a criterion of theory choice has a controversial career in the philosophy of science. Simplicity in ontology (e.g. number of basic categories) has to be distinguished from simplicity in basic predicates and model building, both of which are related to explanatory power as more simple models cover more ground (cf. Olliver 1996; Sober 1975, 1990, 1996). In case of cognitive ethology ontological simplicity may hinder introducing sentience (extending the metaphysical

[13] I loosely use the expression "person" here to refer to apes, although these are not persons in the full sense of the word – substitute "first ape perspective" if necessary.

realm of sentient beings) by proposing non-sentient models of species in question. Simplicity in basic predicates may hinder applying psychologically more complex (e.g. intentional) predicates where simpler descriptions (e.g. by associative learning) are also available.

Consciousness is a complex phenomenon. We can phenomenologically describe it and we work towards a theory of consciousness in CS. If one believes there is a 'hard problem' about explaining consciousness it brings a new type of entity into an otherwise mostly explainable material world. Even if one sees the 'hard problem' as pseudo-problem [cf. §34] one needs good reasons to posit more of it in the world.

Sentience, if there is sentience apart from consciousness, in a similar way brings something of a new quality in the world. *Therefore,* one should posit it only if no other explanation (be it physiological or merely computational, without phenomenality) can be given. *Therefore,* one paradigm case argument is the argument of parsimony (a variant of burden of proof principles). Whenever a cognitive achievement or faculty of an animal species can be explained without invoking sentience, we should assume that sentience is not involved.

Such explanations will often be computational. Highly complex information processing and seemingly intelligent achievements are implemented in PRAM computing machinery, which has no inner life at all. If an animal's achievement can be completely implemented in a computing machine, we have a reason to see it as executed insensitively. This presupposes that the implementation differs from a mock simulation as we may implement for models of human social interaction. A sufficiently rich implementation can only be expected in (small) robots which engage with their environment. Some such robots behave like insects, from the point on where they have demand for energy supplies they gather and have some equivalent of reproduction, their artificial life shows practically that insects of a similar ecological niche and similar behaviour need no inner life. Only followers of esoteric mysticism will then ascribe sentience to such robots.

Parsimony as quality of models should not curtail research aiming at identification of more complex properties. Diminishing the model of animal cognition can be as wrong as inflating it. The crucial test for the appropriate degree of complexity consists not in an exceptionless preference for simplicity (cf. Fitzpatrick

2019), but in judging whether the overall behaviour of a species justifies ascribing the more complex capacity, which should be avoided if no other behaviour of the species than the observed demands it, and should be rejected if other situations demand it, but it is not displayed.

The argument by parsimony cannot exclude that some animal despite simpler models nevertheless has sentience. Especially epiphenomenalists will admit as much. Epiphenomenalism, however, flies in the face of evolutionary theory and defeats itself in the human case (proposing a thesis the attitude of proposing which has no causal role itself by epiphenomenalism, but should have to have a purpose in a debate about epiphenomenalism).

Leaving epiphenomenalism to the side the argument of parsimony, even if it cannot guarantee the absence of sentience in a particular application of the argument, provides a very strong case against the attribution of sentience and emphasises that the burden of proof has not been carried by the proponent of sentience. Consciousness and sentience in their *qualitative difference* from merely computational processes will never be more parsimonious to assume than a large toolbox of computational routines.

§16 *The Fitness Paradigm*

An evolutionary process must subserve the maintenance of the behaviours and structures that evolved. This has several consequences:

(i) The intelligence we see in the behaviour of an animal really is the goodness of fit between some environment *and* the animal`s behaviour. Given a co-operating environment some behaviour may *look intelligent* to us which is just dumb but fits the scene (cf. Budiansky 1998: 27-33, 125-26). Remember: We are looking at the survivors only.

(ii) Since evolution has no foresight there are all kinds of side effects some of which might turn out to be beneficial. A side effect can originate with a single species distinguishing it from the rest of its branch in the tree of evolution.

(iii) Adaptations occur relative to an environment they fit it. Since the environment sets the problems we arrive at constrained cognitive faculties:

> You are only as smart as you need to be to have enough offspring. Everything else would be wasting energy. So, with respect to cognitive faculties assumed in animals we should ask: "Does it really need this?"

For instance, if all members of a bird species show some behaviour to lure away predators from their nest this is no sign of clever cunning but a hard-wired adaptation. Some pet behaviour looks intelligent to us since we *are* a most co-operating environment. Even complex behaviours might be analysable as a sequence of mindless steps each of which is fired by its predecessor (e.g. nest building in weaver birds).

'Fitness' can work as a paradigm critical argument. Fitting into the environment (the 'umwelt' of an animal) may be taken as basis of criticism of too hasty attributions of advanced mental capacities and sentience, as they are misattributed to the animal species instead to the *fit* between the species and its umwelt.

Otherwise not very intelligent animals can show highly complex suitable behaviour in their umwelt. The system of both umwelt and animal constitutes the complex suitable behaviour. The animals need little, if any, phenomenally experienced skills to exploit their umwelt successfully. Robots mimicking such invertebrates, lacking any sentience themselves, are further proof to the problem of misplaced capabilities.[14]

§17 *Animal Concepts (I)*

Philosophy asks what concepts are. Are they abstract entities, symbols in the mind, patterns of behaviour or what? Does being in possession of a concept mean to have some mental representation, to manipulate some symbols or to show a systematic pattern of behaviour? These are important questions, but more important are the properties philosophers think concepts have. We start here with an elaborated theory of concepts – a clear concept of 'concept' so to say – and then

[14] Cf. (Barrett 2011) on several examples, covering both animals and robots, which are illuminating the fit in the combined system of animal and umwelt, where used information can reside in the umwelt without any mental representation in the animal. (One has to ignore Barrett's misinterpretation of Turing, which directly contradicts claims in Turing's four papers on intelligent computing machines.)

consider whether it makes sense to describe some animal`s discriminatory abilities as possession of concepts.[15] Concepts are:

(i) *fine grained* ('the older brother' is distinct from 'the first son' even if coextensive)

(ii) come in a *system* (given an even moderate holistic account of natural language meanings, even if many LOT-types are atomistic)

(iii) Are often *socially acquired and employed* in conditions of fit (at least a lot of concepts, some concepts are supposedly innate)

(iv) some might be causally rooted (given a moderate account of observation language)

If there is an analytic/synthetic distinction concepts can be expressed by analytic statements. If there is not, a concept you have of *x* is given by the totality of what you (firmly) believe about *x*. In both cases – although the theory of meaning is different – concepts are meanings of linguistic symbols. In both cases the question of concepts is tied to the question of having beliefs and having language. What concepts you possess is revealed by your verbal behaviour. Since beliefs are conceptually tied to the other propositional attitudes to have a single concept means to show a very complex pattern of behaviour (cf. Davidson 1999).

Animal behaviour can neither be interpreted in a fine-grained fashion (the dog doesn`t care about his owner being the older brother or the first son) nor is their employment of a supposed concept rooted in social behaviour patterns. Language would allow for fine grained belief ascription and the use of language shows command of the subjective/objective contrast essential to the concept of a belief or judgement. Unfortunately, dogs do not have language. Having a *system* of beliefs means to keep this system coherent, so it requires the possession of meta-representations and *rationality* – things way beyond the dog`s mind (cf. J. Bennett 1964, Davidson 1982).

[15] The main parts of §17 and §19 are taken from (Bremer 2008: 75-78). On concepts one has to combine insights from Davidson (1984, 1999) on natural language concepts and Fodor (1975, 1987, 1998) on concepts and innateness in the *Language of Thought*.

Therefore, animals possess *no* concepts in the sense we speak of human concept possession, especially no theoretical concepts. They possess *systematic discriminatory abilities* which are the precursors of concepts. Discriminating a cat is not having the concept 'cat'. If some animals have something like parts of our LOT, then with respect to some observational concepts they might possess the core of what makes an observational concept in humans. We humans identify such a supposed LOT symbol by the intersubjectively shared distal stimulus. In animals it can be supposed given their discriminatory abilities and the way their brain resembles ours. Elaborated discriminatory abilities (cf. Allen 1999) consist in: (a) self-monitoring with error-detection (e.g. in pigs), which involves (b) an internal representation of what is discriminated and memory of how it has been done, so that this results (c) in an improvement of discriminatory abilities [cf. §27 on the limited metacognition involved]. Even if we should prefer not to call the involved mental representations and nodes in the animal`s categorisation scheme 'concepts' – avoiding to blur important differences – they are something like *schemata*, precursors of concepts, but unfortunately less systematic than the syntactic categories which compose into sentential structures.

Animal schemata are schematic as non-experiential representations and percepts occur in *similarity spaces* relative to the discriminatory abilities of an animal (including their grain and range) and to the chance of exploiting a situation for the needs of the animal. The individual representations can generate a schema by overlap and morphing into each other. So, a cat need not distinguish individual mice but finds itself in situations where the schema get activated. The underlying processes are not experiential, of course. The capacity for schematic representations in animals resembles pattern recognition we find in ML systems (cf. on which Kelleher 2019). One may even speculate that animal LOT might combine schemata into larger units [cf. §29].

§18 *Animal Attitudes (II)*

Beliefs in humans are fine grained (i.e. involve fine grained concepts). So, beliefs require language, which is able to supply words with fine grained meanings, to their expression. Beliefs form a system that has to be coherent. Individual acquired new beliefs are not just put in a belief box, but have to be integrated coher-

ently into your web of beliefs. Therefore, individual beliefs have to be represented as being believed ("I believe A", "I believe B", "So I believe A&B"). So having beliefs requires higher order beliefs and so requires having the *concept of belief*. Beliefs (as opposed to knowledge) live from a distinction between mere belief and true belief (i.e. they involve the concept of *truth* and the concept of mere belief). Higher order beliefs represent lower order beliefs using the concepts of belief and truth. Animals do not represent in language. So, animals do not represent beliefs, at least as far as their awareness goes.

A theory of belief like states would have to work itself bottom-up towards belief. It would explain features of 'belief like' states that serve their purpose without making them full-blooded beliefs. Building blocks of such a theory can be found in Bennett`s theory of *registration* (J. Bennett 1976: §§14-26): Registrations are simpler than beliefs, goals are simpler than desires, despite there being a structure similar to belief/desire-psychology. A system *a* registers ψ, if *a* is in a sensory state which is similar to a ψ-operative state, a state being ψ-operative if a behaviour because of ψ was not accidental. Registering need not be transparent to *a* (even a Cruise Missile can register ψ), but given assumptions about what *a* registers we can suppose what goals *a* has. Registering does not require language nor does pure registering require awareness. We approach belief like states when registering is supplemented with further faculties, for example being able to *learn* given conditional registrations or being a system that *strives* for new information to extend its behavioural repertoire. Developing such an account might give us BL-states which are not beliefs but serve their explanatory power in the animal case and allow for BL-phenomenology. Similar with desire like states.

Some BL and DL states involve phenomenality by embedded percepts of some type [as will be argued for later], BL and DL states in themselves need not involve phenomenality. In the human case beliefs and desires need not involve phenomenality, as they can be non-conscious. So, ascribing representational BL and DL states to an animal species is *not* equivalent to ascribing sentience to that species. A case in point may be bees. Bees process systematic representations *describable* in a BL/DL-framework, but insects in general are not good candidates for sentience (by the other dimensions of the reflective equilibrium, like neural physiology and general plasticity of behaviour).

§19 *The Language Paradigm (I)*

A language is a finite transformational system which given an alphabet and some axioms/start symbols generates a word set. Natural languages are:

(i) *compositional* (exhibiting structural rules that allow to understand the [new] whole given the meaning of the parts, which contains their logical function within the whole),

(ii) *productive* (allowing to make infinite use of finite resources by rules of generation)

(iii) *discrete* (the signs [phonemes, letters] can be kept apart by their physical properties)

(iv) *semantic* (signs are symbols in that speaker and audience share the meaning by conventions).

A language is a discrete combinatorial semantic system. Semantics governs sentence construction, there being mental modules accordingly explainable within a computational perspective. Animals use sounds and interact using sounds, but it is a misnomer to call these interactions "language use".

Let's consider the characteristics of language in reverse order:

(ad iv)

Animal signals are almost ever used to make some other animal do something. For this purpose it isn`t necessary that speaker and audience share a meaning (one vervet monkey says: "I see a eagle" and the audience hears: "Go to the trees!"). If meaning requires mutual knowledge (i.e. higher order beliefs and expectations) it is way beyond the animal mind, lacking ordinary belief. Even signing apes use 95% of their signs in imperatives. Nevertheless, in these apes 5% of their utterance are used to refer to something. Chimpanzees can translate real-world objects into mental representations, but are surprisingly weak in doing the opposite (taking photos, maps etc. as a guide to the real world). They can transfer abstract ideas (e.g. 'colour'), pigeons cannot. Apes even seem to understand the abstract relation between an arbitrary sign and an object, but they cannot represent second order relations (cf. Premack 1976, 1983). The bee`s dance exhibits a causal chain of reference to food and the direction towards it, but apart from being inflexible and

innate the use of the dance is imperative and used only by foragers (i.e. not all members of the species 'speak'). Most animal language is stimulus bound and shows no displacement (of reference); dolphins understand references to objects not within their visual field. A lot of animal sounds (e.g. in squirrels) can be explained by the effects of their acoustic features (e.g. on attacking hawks).

Taking individual signals as representations, one may give an externalist semantics for them, analogous to externalist semantics in conceptual atomism with respect to human LOT (cf. Fodor 1987, 1998). The signal carries informational content by tracking a property in the typical situations of the animal (cf. Dretske 1988: 83-105). For a signal to have objective content in this way it need not be conventionalized as a symbol. This implies, however, that two animals although both signal with the signal do *not share* a common meaning.

(ad iii)

Apart from apes and dolphins trained in sign use and sign language animals almost never use discrete signs. A dog has one type of bark, which may become louder and faster, but cannot be split into discrete units. Baboons grade their calls (intense barking signing warning to very intense barking expressing fear or commanding escape). Animals use *graded tokens* (cf. Dobrovolsky 1989). This applies also to whales. That severely restricts the possibilities of recombination.

(ad ii & i)

Despite some claims to novel signs produced by trained apes there is no evidence for productivity and compositionality in apes. There seems to be a elementary sensitivity to word order in trained dolphins (cf. Schustermann et al. 1986) and some apes, but distinguishing word order is not syntax. Apes cannot produce structural innovation, but only lexical substitution. Dolphins 'only' form a learning set and transfer it to novel examples. Some birds (e.g. a chickadee) have combinatorial songs, but the units have no function.

A common 'clever quip' on the inability of animals to use language asks rhetorically: "Haven't we failed in learning their languages, as well?" [This is the third most common rhetorical question in the debate, besides those two in §1.] Apart from presupposing that there is something systematic enough to be learned, this overlooks that concerning learning animals we endorse them and look for minute

signs of correct behaviour, ready to classify it as good enough to be further endorsed and developed. Animals do not do this. As with tool use – give or take some anecdotical evidence – animals do not teach. There might be occasions of animal teaching, but mostly animals learn the behaviour of others by watching and then engaging in trial-and-error efforts. For language this does not suffice.

Language is the essential feature of human uniqueness. Thought and consciousness depend on it. Descartes, therefore, anticipated Turing in the idea that the crucial test for the presence of thought is the fluent and creative use of natural language in a conversation with us.[16]

Although there is no animal language there are complex patterns of animal sign use and animal communication. Although they make no language, they obviously involve several cognitive faculties and recognition of sign production in the animal itself or some of its flock. Some animals have the ability to *classify* objects and *to link* the classification system used to arbitrary signs for the purpose of communication/interaction with another animal. These faculties – especially in apes – go beyond sentience in general. The beginning of reference and use of pronouns can be situated in a level above BL-states and beneath language and beliefs. What is crucially missing in apes is a theory of mind which allows them to see their comrades as intentional agents.

§20 *The 'Theory of Mind' Paradigm*

In reflection one act is the object of another act. This reflexive act can be reflected upon. The object of reflection then is a more complex act represented by a more complex representation involving embedded cognitive state (i.e. embedded cognitive operators). Iterated higher order reflection thus *represents* ever more complex states of embedding. From a *procedural* perspective there are always *only two* levels: the ongoing mental process (even be it one of thinking about thinking) and its represented object (be it acts of iterated thinking about thinking).

[16] The intimate relation between thought and language defines the tradition of 'Cartesian Linguistics' (Chomsky 1966) up to current Generative Grammar.

Reflexive or higher order states are pervasive in our mental life (e.g. in the dynamics of belief acquisition, update and revision, or in the following of conventions). Consciousness, however, cannot (for the reasons present in the philosophical tradition starting with Fichte) be *explained* or modelled by a Higher Order Theory – and it *should not* be so explained in the light of good theories of conventions and belief dynamics, as the higher order states involved in our background (implicit) reasoning are *not* conscious and would overload conscious life if any higher order state was conscious. Not any higher order state, thus, must be conscious. If the mere presence of higher order states yielded consciousness, even artifacts which we do not consider to be conscious (as some IT-systems with self-monitoring representations of some of their states) had to be conscious.

If a Higher Order Theory now liked to make a distinction between those higher order states and those leading to consciousness, the theory shifts to those distinguishing features as defining consciousness, instead of placing the crucial feature in the higher order structure.

Anyway, the higher order state supposedly presupposes the existence of the first order mental state, and thus follows it *in time*. This delay may be one problem to ascribe causal powers to conscious states. A further problem in this vein rests in the higher order state not changing the given first order state, so that the causal powers/functional roles associated with consciousness cannot rest in that state, but only in the new state, i.e. the higher order state, which *itself* is not conscious. Causal efficacy of consciousness gets lost or cannot be explained.

It is – at best – misleading to speak of a mental state/act 'becoming conscious'. Non conscious mental states and conscious mental states are structurally distinct. What 'becomes conscious' is the content (state of affairs) that is represented in the non-conscious mental state. The content is conscious content in a conscious mental state, a representation token of that content is present in the conscious state. The non conscious state is not present in the conscious state, and it does not undergo some transformation to become conscious. Why some *content* enters consciousness is an epistemological/psychological important question. It points to some procedures of tending to epistemic issues or pressing decisions to be made by the cognitive agent. The mental operations and the (non-conscious) self-monitoring of the cognitive system may give rise to conscious states, and if there

was an explanation to be had for why some content becomes conscious, it might be found here (cf. Baars 1997, 2021). The structure of conscious states does not *explain* – as some Higher Order Theories pretend – why some content is conscious, it *describes* what is involved in consciousness.

A non-higher order representational account of consciousness resembles a Higher Order Theory by emphasizing the role of embedded (self-)representation for the occurrence of consciousness. Embedding (e.g. in the scope of operators), however, need not be higher order in the sense of a state being the object of another state (i.e. in the sense of reflexion of Higher Order Theories). A representationalist account that relies on the – traditional – idea of immediate self-access or self-presenting representation can be taken as a 'First Order Theory' of consciousness.

Embedding of (quasi-)intentional states occurs not only in reflection, but also in attributing intentional states to others. To do this one has to perceive the others as intentional agents with an intentional mental life. One has to have TOM in perception of conspecifics and other animals.

Physical 'awareness' (registration which parts belong to one`s body) is present in many animals. Some have supposed that the use of pronouns (e.g. "me") in trained apes might be a sign of self-awareness, and most famous is the mirror recognition test (cf. Parker et al. 1994). This evidence has been contested (cf. Budiansky 1998: 161-88), especially since some animals we otherwise would not rate as highly developed as the great apes (e.g. some monkeys, maybe even some ravens) pass the mirror test. Which behaviour could be the basis of an attribution of self-awareness? Having the resource to understand oneself, one would have the resource to understand others as (intentional) agents. Reasoning about others might even evolutionarily precede reasoning about oneself. Some researchers training apes in sign language claim that *trained* chimpanzees, for example, are able to recognise intentions in other chimpanzees. A dolphin is aware of the effects of its behaviour on others; anecdotes are claimed to show that a dolphin can behave as if he knew that the behaviour was illicit and did it only when no people were around. Imitation also ascribed to some animals like apes would require understanding the other as intentional. One can argue that animals lack the required TOM, and this being the decisive difference between a chimpanzee and a one-year-old child (cf. Tomasello 1999). Only humans understand conspecifics

as intentional which enables a new form of cultural learning and enables conventions (cf. Searle 1995). Non-humans do not see the world in terms of intermediate and hidden forces (i.e. causality and intentionality). Non-human primates are quasi-intentional beings (having BL and DL states), but they do not understand the world in intentional and causal terms. They see others as animate, but not as intentional. They can see others as seeing a situation, because they themselves see the situation, no representations of the others' representations are needed, because the situation itself *as seen* by the 1st animal itself fulfils this role. Ascribing a false belief, in contrast, requires representation of the representational vehicle employed by the other, which requires (inner) speech to access this representation phenomenally, which requires language use. A truth evaluable representation of *a situation* has to have constituents to *locate* reference points (i.e. singular terms) and to *characterize* them (i.e. general terms), i.e. has to be sentential. Ascribing a false belief uses a sentential meta-representation of a sentential representation. This requires meta-linguistic competence. One can switch the approach and start to investigate how animals can reliably *track* what others are up to (cf. Butterfill 2019), without representing their mental content, but, for instance, by being attuned to behavioural cues.

Animals in contrast to humans lack TOM, because as lacking language they lack propositional attitudes, and thus lack ascribing propositional attitudes to others, without which TOM does not get off the ground (cf. also Bermudez 2009). Humans as language users are additionally able to construct and consciously access explicit linguistic meta-representations (cf. Bremer 2012). Humans not only possess the concept 'belief', they can even *quote* their own beliefs and those of others.

Can an animal which has no beliefs be sentient? Once again, we are at a loss to specify *how* the animal is aware of something, what animal representations are like. Encounters with conspecifics and ritualised behaviour in the flock surely have a distinctive feel to them. Animals like apes and wolfs that show ritualisation and co-operation seem to be on some level beyond the turtle and below the two-year-old. The difficulty lies in outlining their representational resources. Theories of mind, on the other hand, that entail that animals have no phenomenal mental

content and sentience *at all*, fly in the face of the evidence of most approaches in cognitive science and are thus highly questionable.

§21 *The Simulation Paradigm*

One function paradigmatically associated with sentience is direction of attention to gather more relevant information and (to start) to learn. If the repertoire of action is a set of fixed conditional behaviours (connecting information and motor output) there is no need to sustain the costs of sentience.

Some invertebrates have a fixed neural connectome (i.e. their brains are hard wired). Thus, they cannot learn new conditional reaction types. Thus, there is no need for sentience in these invertebrates.

Machine Learning (ML) has occupied the title "AI" in the last years, although its methods neither cover all AI nor all forms of learning in humans. The talk of artificial 'neural nets' in technical systems has captured public imagination. ML's achievements in category and associative learning, and in some predictions are impressive (but cf. Narayanan/Kapoor 2024). The algorithms use a *simplified model* of NNs and run on PRAMs – there are no more neural nets in such a run than rain in a computer running a weather simulation.[17]

The success of Recurrent and Bayesian NNs in prediction made these conceptions re-enter CS and contribute to a theory of cognition as crucially involving 'predictive coding' both in perception and in motor control (cf. Clark 2016, 2023).

Some insect like robots which successfully cope with a situational task (like collecting garbage) possess a 'subsumption architecture' (cf. Brooks 2002) where individual modules work in parallel without coherent integration by a central unit of process control (a GWS). So, *prima facie*, insects may be successfully cope with their umwelt without some inner realm of integrating their information flow and motor output, thus without sentience.

[17] With some programming skills you can easily implement the basic algorithms on your standard computing machine (cf. e.g. Raschid 2016).

The impact of ML on cognitive ethology has to be seen as providing new paradigmatic counter-arguments to those arguments for sentience which are based on learning. Extensive attributions of sentience to invertebrates typically argue by the learning behaviour of, say, bees (cf. Chittka 2022). These learning achievements, although one may consider them to be impressive, are often and maybe every time not beyond the reach of ML. ML – to repeat – runs on standard vNA PRAMs, which have no sentience at all.

Whereas sentience advocates like to infer by *Modus Ponens* from similar learning behaviour between some mammals which are sentient and insects that insects have to be conscious or at least sentient, critics may refer to ML to apply *Modus Tollens* and infer that not just the argument for insect sentience based on learning fails, but that the support from learning behaviour to mammal sentience is questionable as well.

§22 *Reflective Equilibrium (II)*

An overview of some of the mentioned scientific evidence for sentience is summarized in the following table.[18]

Apart from our intuitive certainty that mammals and some birds are sentient the positive evidence collected across different parts of CS supports ascribing sentience to them, at least, given this evidence, the burden of proof shifts to the deniers of sentience. The paradigmatic arguments employed have their mentioned weaknesses, but together may be taken to justify a *default* assumption of sentience for mammals and some, if not all, birds. The further argument shifts, therefore, to questions how to conceive of their sentience, i.e. how to conceive or model the content of their experiential mental life. [Which will occupy the discussion in the second part of this book.]

On the other hand, the controversies how far to extend the assumption of sentience are fierce with respect to fishes and invertebrates. The negative evidence displayed in the table *prima facie* justifies the denial of sentience with respect to insects. The burden of proof still lays on the proponents of insect sentience.

[18] The table and some material in the preceding paragraphs are translated from (Bremer 2005).

Source of evidence	Positive evidence	Negative evidence
Neurophysiology	- presence of nociceptors, pathways for pleasure, and morphine - areas governing pain similar across vertebrates - structurally similar brain to humans in vertebrates (and CNS) - sympathetic autonomous nervous systems in vertebrates	- no CNS in insects, quite different brain structure in insects and (some) fish. - cold blooded nervous systems may lack speed for information to travel - completely fixed neural system in some invertebrates
[Cognitive] Ethology	- adaptivity to novelty - pain behaviour patterns - anxiety behaviour (increased arousal, tension, inhibition) - pleasure-seeking behaviour	- insects notably lack behaviour which protects injured body parts - accomplishments of insects may be better attributed to the hive/colony than the individual
Evolutionary Theory	- pain seems to focus attention to a harmful situation, event - plasticity is required and might involve a central 'theatre' of control and co-ordinating input (GWS)	- pain might be absent in insects because there is no selective pressure to protect the individual body because of the short life span. - similar for small fish in large swarms
Cognitive Psychology	- *pleasure* supports learning in humans - perception can be *focused*	- learning abilities similar to learning abilities not involving consciousness in humans
RTM / CTM	- representations may have phenomenal aspects - partially shared LOT	- mental information processing need not be and generally is not conscious
Robotics / AI	- a world model in control structures profits from including a *self-model*	- steering behaviour of some insects can be rebuilt in mindless little robots

§23 *The Anthropomorphic Fallacy*[19]

What's wrong with anthropomorphism? One could say that we project human characteristics onto animals. It's not just that we describe animals using terms that are usually applied only to humans. First, for any minimal realism, describing something as *F* does not necessarily entail that it is *F* (except in the special cases of institutional speech). But this cannot be the fault of anthropomorphism: the possibility of error, of misapplication of an expression like "thinks he sees three birds" is also present in the description of humans. Second, however, it is the case that we can reasonably assume that the categories of our best theory of *x* capture the real properties of *x*. Thus, we assume that the categories of psychology, as they concern humans, capture the real psychological properties of humans. If these terms could be appropriately applied to animals, then there would be no doubt that we have identified the corresponding properties *just as in the human case*. The point is that these categories/expressions cannot be entirely appropriately applied to animals. The fact that they can be applied at all is due to the fact that complex concepts and corresponding expressions refer not only to one criterion of their applicability, but to multiple or even context-specific applications. We use such expressions only in loose language (or 'analogous language') when we apply them based on the mere approximate fulfilment of some of their application criteria, often even knowing either nothing about the applicability of the other criteria or, moreover, knowing that some of these criteria are not fulfilled.

The *anthropomorphic fallacy* regarding an expression δ in human psychology consists in assuming, in a case such as the one just described, that the *complete* concept of δ can be applied to *x* or that *x* actually *has the properties* corresponding to δ.

A simple example: Pressing a button on the TV remote control, I say "the device isn't responding" (perhaps the batteries of the remote control need replacing). Here, I use "respond" in reference to a television. My statement makes sense. This is possible because the object described received a command but showed no response, whereas in other situations it responded correctly. "Respond" is used anthropomorphically — and in the strict sense incorrectly — because a response

[19] §§23 – 25 are mostly a partial translation of three paragraphs from (Bremer 2006).

is an intentional (speech) act, and the conditions of an intentional system that this presupposes are simply not met by televisions. We can, out of laziness, expand the concept of 'responding' to include reactions from technical devices. In this sense, the expression "responding" would have been used correctly in the example – but only at the cost of exchanging a more precise, narrower concept for a broader one. In fact, this would have been done purely out of laziness, since we already have the non-intentional concept of 'reaction'. In order to be able to talk about the intentional responses of people as opposed to technical devices, after having expanded the term "responds," we would have to introduce a new, narrower term, or we would sink into vague and undifferentiated language, which is acceptable in everyday life but the end of it for cognitive science.

The critique of anthropomorphism in descriptions of animal behaviour therefore means that opinions and desires in animals are like answers in television sets.

Beyond identifying the fallacy, a comprehensive theory will seek to answer why this fallacy is so widespread. Uncovering the cause of an obvious fallacy refutes it once again — this time due to its deviant genealogy. In this sense, it has been suggested (cf. Carruthers 1992: 125-26.) that the root of such hasty transfers lies in the fact that we simply transfer our innate explanatory patterns of everyday psychology to non-human objects. Within cognitive science (particularly learning theories and theories of language acquisition), there is some evidence to suggest that the basic patterns of everyday psychology (i.e., rationalizations of actions by attributing opinions, desires, etc.) are innate (cf. Fodor 1987; Premack/Premack 2003). This provides us with – in addition to the equally basic causal explanation pattern – an orientation pattern for predicting the behaviour of moving objects, which we simply apply to them from an early age and only withdraw when special implausibility arises.

The difficulty therefore lies in distinguishing the overgeneralizations of anthropomorphic fallacies from appropriate psychological descriptions (of animals).[20]

[20] An extreme position is that of Kennedy (1992), who concludes from the fact that we have a tendency to apply intentional explanations not only to other people but also to animals that the application in general is *wrong*, and therefore proposes the development of a neo-behaviourist terminology to overcome this anthropomorphism.

§24 *Heterophenomenology*

One could object to the possibility of animal psychology — and it is occasionally objected in one way or another — that such assumptions are in the air anyway, since we have no access to the minds of animals. We simply do not know what goes on inside them and therefore couldn't have a theory of animal cognition. This argument is often used when animal lovers claim that animals have comprehensive abilities and want to dismiss all empirical evidence against this. If this objection were relevant, then at least a reluctance to judge any animal emotion would be appropriate — if not, for the sake of simplicity, a denial of animal emotion — because speculation and fantasizing in the face of a lack of data are not part of the cognitive science toolkit.

This objection can, however, be rejected. A psychology from the third-person perspective is possible. Daniel Dennett has developed the corresponding approach under the title *heterophenomenology* (see Dennett 1991: 72-81, 444-53). The heterophenomenologist collects all data that presumably point to cognitive abilities, sentience or consciousness. This includes descriptions and recordings of behaviour, but also all utterances made by the being in question, especially those concerning its inner life. Such utterances are taken ad face value. Heterophenomenology does not begin by excluding data or devaluing it in relation to others. Its primary concern is a comprehensive collection of data. From this data, the heterophenomenologist first constructs 'the heterophenomenological world' of the subject (Dennett 1991: 81), i.e., in the case of humans, the subjective world with all the mental occurrences that the subject believes and expresses. The object of study acts as the final authority on which data are to be included. Theories are then developed to explain the occurrence of this data or the events underlying them. Mental events and cognitive abilities are the theoretical entities of such theories. Only in the course of such theories (in the context of establishing a coherent overall picture and a reflective balance between the theoretical concerns and the collected data) does the reliability of individual data or individual data sources come into question. Generally, however, the mere collection of data assumes the – initially purely instrumentalist – stance that the living beings under study are intentional systems whose behaviour and utterances have meaning for them and which, in the face of (something like) opinions, pursue their goals.

Within the framework of theory development, this "if" discourse can then possibly be transformed into a non-relativized description. In this case, the theory aims at adequately capturing the psyche of the theoretical object. Contrary to the objection that a third-person psychology is not possible, it is not at all clear what is missing here. After all, the subject can report all mental occurrences that it considers relevant. A residue would only remain if the objection were to be based on the (linguistically and epistemologically) dubious thesis that one can either have conscious occurrences without believing that one has them, which – in Kant's words (1786: 131-32) – would mean that they are simply not for me, or that one can mean something without being able to express it (in any way). But even if this were possible, the impossibility of expression would at the same time mean the end of a psychology that seeks to base its science on the first-person perspective.

§25 *A New Anthropomorphism*

A psychology from the third-person perspective is therefore possible. A variant of heterophenomenology can be a form of cognitive ethology, since a theory of animal cognition is based exclusively on an external perspective — here, the question of which subjective data we lose with such an approach is beyond an answer. The consciousness we possess is not the sentience that is examined there. What matters is classifying the cognitive performance of animals from our perspective and assessing their scope. In the case of animals — with very few exceptions: signalling primates and dolphins — there are no direct reports of their inner lives expressed in a common language. What matters first is a comprehensive collection of presumably relevant behavioural data. The selection of these data and the interpretation of animal vocalizations is already more theory-guided than the uncritical listening to the subject within the framework of a heterophenomenology of humans. A set piece of theory is required to begin the heterophenomenology of sentience. This leads to a *new anthropomorphism*. The heterophenomenological approach provides the starting point. Furthermore, the *heuristic* use of the vocabulary of human psychology plays a major role here. The new anthropomorphism differs from naive anthropomorphism in that it recognizes the non-transferability of the terminology of human psychology to animals. The goal

of a methodically reflected cognitive ethology can only be a science with its own taxonomy, with its own vocabulary, which is clearly distinguishable precisely with regard to the differences between human and animal cognition. On the path to this goal, however, properly understood anthropomorphism represents a necessary *transitional* stage. The anthropomorphic description of animals in a sense that will be clarified shortly may thus be the beginning of animal psychology, but it must not be its end.

The vocabulary of human psychology is used in two ways in the new anthropomorphism. First, in a *heuristic* way. Starting with an initial transfer of categories from human psychology, the question is asked, on the one hand, what justifies the application of these terms to animals, and, on the other hand, what is lacking in animals in order to speak of a completely appropriate use of these terms. Second, the question arises as to which cognitive abilities and states can be identified in animals at all. The new anthropomorphism takes the identification of cognitive abilities and states from heterophenomenology. The aim here is to specify as precisely as possible the behavioural data that, when present, allow a justifiable inference of corresponding cognition from the third-person perspective.

However, not every application of a psychological term to animals constitutes an anthropomorphic application. Humans, and thus human cognition, evolved from animals. The special abilities of humans (especially linguistic coding and the emergence of consciousness) are not isolated phenomena, but rather shape a person's other mental or conscious states. Presumably, there must nevertheless be areas of cognition that are found either completely or, at least if their specific transformation in human cognition could be disregarded — which is by no means certain — in a similar form in closely related animals (especially higher mammals). The more evolutionary psychology develops, the more an interspecies-applicable cognitive vocabulary may emerge. The task of cognitive ethology in this context is:

(i) to specify the conditions under which corresponding states or abilities occur and what role they then play in the animal's behaviour;

(ii) to identify which animals possess the corresponding conditions or abilities;

(iii) to establish a dividing line between an unproblematic cross-species way of speaking and an anthropomorphizing way of speaking regarding more complex cognitive performances.

Specifying expressions that fall into this category is difficult, as many expressions for typically human abilities (such as "answering someone") are often merely applied figuratively to technical devices. Approximate candidates for species-independent expressions could be expressions such as "orienting oneself to the right in the visual field in response to the information ..."

The paradigmatic anthropomorphic pattern in this case is: If plastic behaviour systematically occurs that is accompanied by psychological states of type *F* in humans, we are *prima facie* justified in inferring *F*-like states in the corresponding animals.[21] The restriction to plastic behaviour is intended to exclude behaviour that reacts fixedly to specific trigger stimuli due to genetic anchoring and, due to its fixed mode of action, does not require any accompanying consciousness. The difficulty arising from this rule, however, is what the phrase "*F*-like" is supposed to conceal. The similarity is supposed to refer not only to the causal role, but also to the corresponding internal episodic or dispositional event — assuming that we have already observed the behaviour in question.

Specifying the corresponding causal role in such a way that it can be generalized to animal behaviour will generally not be so easy. In the paradigmatic case of a mental state with a causal role — the case of pain, which is often used as an example — this may be quite obvious, since avoidance behaviour or expressions of pain occur immediately in the situation of injury. Analogously, states such as hunger, fatigue, and fear are quite closely related. However, when dealing with more complex states, it becomes more difficult, since in this case one also has to determine the mental content of the state in question approximately.

Evolutionary explanations of animal behaviour and their cognitive abilities are based on identifying the functions of these abilities. In the case of cognitive

[21] A similar principle can already be found in Romanes, who directly followed Darwin; see (Romanes 1898: 6); for the rise, decline, and rediscovery of this methodology, see (Rollin 1989). The overstretch of analogies to human persons also motivated Morgan at this time to disagree, which came to be known as 'Morgan's Canon', that a simpler explanation of the behaviour of an animal should always be preferred. [cf. §15]

abilities, this essentially has to do with the representational content of corresponding states, which places the animal in a behaviour-guiding relationship with its environment. The attribution of content (the attribution of quasi-intentional states) is what makes this type of explanation possible. Determining the content of representational states is already problematic with regard to the reference objects of such states. Although we are fairly certain that mammals, due to their shared neurophysiology, discriminate between objects in a similar way, the object we perceive through sensory impressions already exists in contexts typical of our way of life and falls under corresponding concepts. A dog knows neither the meaning of wardrobes nor tool cabinets, nor even that of tables, and certainly not that of ballpoint pens. In this sense of the classification systems used, we and dogs do not live in the same world, even though, like dogs, we relate to corresponding objects of reality. Regarding the objects we identify as the same, it is not at all clear whether an animal that displays different reactions to an object in different situations considers the cause of these reactions to be one and the same object. The attribution of content to the dog's states either proceeds by assuming concepts beyond the dog's horizon or requires a simulation/empathy into the dog's world – possibly guided by imagination. Ignorance of the contextual relationships of this world (such as the role of lip curling and frontal gaze as signals of aggression) leads to corresponding miscoordination in human-animal behaviour and is a source of mistreatment of animals. The difficult task of adequately describing the behaviour of animals, to which we attribute mental representations, consists in finding the way of describing the situation in which they would describe it if they had a minimal language at least. Simulating/empathizing with the dog's world carries the great risk of not being able to disregard our conceptual framework and thus becoming entangled in overly anthropomorphizing language. A difficulty that was probably not unrelated to the decline of historical anthropomorphism. It is questionable that the animal itself propositionally represents the content of its state, which we express propositionally, because it lacks language [see below].

However, to the extent that it is possible to predict animal behaviour on the basis of corresponding attributions, the corresponding heterophenomenological attributions prove themselves – quite analogous to the validity of (psychological)

theories in general, provided that criteria of explanatory strength and simplicity are also used when weighing up competing explanations.[22]

An anthropomorphic description proves its worth to the extent that, on the one hand, general intentional descriptions can be justified even for instrumentalists (cf., for example, Dennett 1983), but especially by successfully answering the following questions (cf. Bekoff/Allen 1997: 69-71):

(i) What exactly is cognitive ability *x*?

(ii) Due to which properties of the animal is ability *x* an intentional, reliable, productive and plastic ability?[23]

(iii) *How* is this ability normally exercised?

(iv) How does this ability relate to the animal's other (cognitive) abilities?

(v) *Why* has this ability become established through evolution?

In summary, this list of questions integrates the justification of the attribution of the corresponding ability in general with reference to the quasi-intentional contents to be attributed.

§26 *Sentience and Animal Welfare*

Identifying animal mental states is not merely of theoretical interest. If animals have such things as sensations or desires, then — given a corresponding moral conception — they must also be considered morally.

---

[22] The appeal to the fact that, given the abundance of alternative explanations that would otherwise have to be provided and the mass of behaviour explained in this way, the intentional description of animals is *simplest* and is therefore preferable, is found among several cognitive ethologists (see, for example, Dawkins 1993; Griffin 1992). The problem with such appeals to simplicity is that many cumbersome replacement explanations may still be simpler overall than introducing a new explanatory model regarding some kind of entity that postulates complex capacities and mental entities. [cf. §15 on arguments by appeal to parsimony]

[23] In other words, a skill that can be transferred to new situations and must therefore be adaptable to situational parameters. Reliability refers to the chance of being successful in important cases where this skill is used – even if these are not the majority of cases.

The possession of certain cognitive abilities often coincides with the possession of morally relevant properties. This is one respect in which a theory of animal cognition is relevant for animal ethics. A number of corresponding justification programs now exist. Given such a justification that animals are to be considered morally, the more difficult question arises as to how they are to be considered. The real problems of animal ethics today seem to be less problems of the general justification of the moral consideration of animals than the theories of 'animal welfare' required for this purpose, and interspecies utility comparisons. The (quasi-)interests of animals can only be considered morally if, on the one hand, we know what (quasi-)interests animals have in the given situation and, on the other hand, we can compare the extent and importance of these animal interests with competing (quasi-)interests of other animals and human interests. A theory of animal welfare must be based on cognitive ethology. Fundamentally, there are several ways to understand animal interests/valuations (cf. Dawkins 1980, Webster 1994):

(i) On the one hand, we can infer the presumed extent of animal suffering or pleasure from *observing animal behaviour*, especially the strength of its reactions (e.g., the speed of withdrawal or eating). This requires a behavioural theory that allows for the correct assessment of reactions (e.g., not mistaking lip curling in non-domesticated dogs as an expression of pleasure, but viewing it in domesticated dogs as an adaptive effort to coordinate with humans, whom the dog has observed signalling friendliness in this way). [For example, one can determine that in laboratory rats, blood pressure and heart rate are higher when kept on wire mesh floors than when kept in cages with a continuous floor (Krohn 2002).]

(ii) Secondly, we can attempt to assess the presumed extent of animal suffering or pleasure by measuring *the effort* animals make to escape from the corresponding unpleasant situation or to enter the corresponding pleasant situation. In such studies, the animals themselves are 'asked' through their reactions (for example, when a chicken or a rat chooses either one cage flooring or the other by being able to move freely between the options). However, it remains difficult to say at what level of effort the desired state becomes essential for the animal's well-being.

The reliability of such statements can be increased by methods such as factor analysis and species comparison, supplemented by physiological knowledge (for example, to be able to distinguish between the short-term well-being resulting from high sugar consumption in a dog and the long-term damage this habit causes to the dog).

(iii) Thirdly, a sound heterophenomenological *use of analogies* to the human case may be permissible to a limited extent.

Investigations of this type are the beginning of welfare relevant mental content attribution.

What is permissible to do with animals? Do animals have rights, and do we have moral obligations toward animals?

These questions are complex in several respects. On the one hand, the answers will depend on which ethical approach one considers correct in general and specifically with regard to animals (e.g., utilitarianism, compassion ethics, or reciprocal moral rights, etc.). On the other hand, the vast diversity within the animal kingdom, ranging from bacteria and insects to vertebrates, mammals, and especially primates, must be taken into account.

This is where the relevance of ethology lies: ethologically, hypotheses regarding the existence of morally relevant characteristics in different animal species can be justified. This will lead to very different moral obligations towards animals of different species. At one end of the spectrum, primates could be animals whose communicative abilities and rudimentary self-image place them close to the human moral community. At the other end, insects (and all simpler animals such as bacteria) could be animals towards whom there are hardly any moral obligations, insofar as they are not sentient. The complexity and extent of presumed preferences of animals determine the extent to which they should be included in our moral considerations. Mammals such as cows and pigs will be ascribed not only sentience, but also a multitude of lifestyle preferences (ranging from exercise to social contact) that are incompatible with conventional (mass) animal husbandry.[24]

[24] Cf. on animal welfare in general and ways to decide between forms of husbandry: Broom 2014: 91-107, Krohn 2002, Webster 1994.

The attribution of abilities thus has an ethical dimension, which is why morally relevant characteristics should not be attributed to or denied too quickly. In particular, the determination of whether a species of animal can feel something like pain (for example, in the case of fish) points to an ongoing ethical controversy, in addition to the ethological problem (whether ethological evidence can be found in a species' complex avoidance behaviours or neurological evidence in analogies to mammalian brains).

Talk of 'animal rights' (Regan 1983; Sunstein/Nussbaum 2004) cannot be taken literally. Taken literally there would have to be a plurality of animal rights. This is doubtful. First, it is doubtful that animals have any *rights* at all, because they are not persons, and the foundations of ethics employing the 'rights'-concept rely either on the concept of 'personhood' or reciprocity between persons (e.g. Kantian Ethics, Discourse Ethics, Contractualism). Second, it is doubtful that animals have *multiple* rights, meaning that these rights can be differentiated as they can in the human case. In the human case specific rights correspond to specific morally relevant properties of a person (e.g. a right to politically assemble). In case of animals one may understand their welfare in different dimensions, but these are not rights in the former proper sense.

One might nevertheless use the talk of 'animal rights' and lay down the principle: *The basic animal right could be to live in a species-appropriate manner* (i.e., without unnecessary human interference). This right of animals would be a right not to be harassed. This right can be justified by recognizing sentient animals as moral patients: their welfare can be affected by human action. Morally considered the permissibility of an action depends on its rational acceptability in principle by those affected. Sentient animals can be affected by our actions even though they cannot rationally consent or dissent from being affected. Widescale ignorance about the quality and extent of them being affected warrants assuming it best to leave them to themselves. There correspond no immediate duties to this right.

Non-interference by humans also means that all naturally caused animal suffering will continue. This suffering is part of the natural life of animals. Animals cannot (without the absurd consequences of a complete human transformation of nature) have a right to freedom from suffering. Non-interference means that we do not cause *additional* animal suffering. Within the framework of a moral consideration

of animals, one could view this as the moral right of animals and thus speak – in an extension of usage – of 'animal rights'. (In addition, animals generally have positive *legal* rights.)

Such a minimalist conception of animal rights assumes that animals, insofar as they are affected by human actions in their sentient vulnerability, must also be morally considered – albeit to a lesser extent than persons. This can be expressed as the 'right to *consideration*'.

This title immediately raises the concrete question: How should they be considered? Answer: In the interests of their respective species-specific and individual welfare.

In this respect, questions of animal welfare must be answered. The general title for an animal's welfare, in turn, can be understood as 'a species-appropriate life', which is why the right to moral consideration leads to the right to a species-appropriate life. What constitutes a species-appropriate life can only be determined within the framework of an empirical ethology of the relevant species. Here, the discussion of animal rights again depends on ethology. Given a conception of the appropriate life of a species (e.g., whether it is a wild animal or a domesticated animal), specifications of the right to a proper life can be derived, for example, regarding the necessary food, the required habitat, etc. One could now call these specifications 'derived rights' (in the plural). These (species-relative) 'rights' will often sound strange (a 'right to grass', 'right to 1 square km', etc.) and depend on ethological theories and facts about the context, which can change. Whether one nevertheless calls them 'rights' rather than 'specifications of the right to a proper life' is perhaps a dispute over words.

It should also be noted that a right to a species-appropriate life need not entail protection from being hunted (inasmuch as this resembles natural prey fate), exemption from usage by humans (inasmuch as this could be considered natural, as in ploughing with an ox or sheep keeping the bushes down), and need not entail protection from being slaughtered (inasmuch as this takes place at an ordinary life span and happens quick and painlessly, if that is possible). In the state of current animal use one may doubt though that this permissions in principle can be put into a practice which does not relevantly increase animal suffering. As a moral rule of thumb an advocate of animal rights better works with “Let animals

keep to themselves, as far as possible!" The possibilities are limited by humans and animals sharing living space and meeting each other, be it with domestic animals or in the wild.

The moral consideration of animals arises very differently from different ethical approaches. Human exceptionalism applies to morality as well. Morality can be best justified by person centred (anthropocentric) ethics like Discourse Ethics (Habermas 1995).

Within the framework of the *anthropocentric* justification of animal ethics, at least three justifications are available:

(i) For reasons of prudence and based on the rights of children, we have reason and an obligation to preserve the natural foundations of life. Animals are components of such ecological systems. Factory farming and meat consumption contribute significantly to environmental pollution. A rejection of the current amount of consumption of animals can be justified in this way alone, regardless of whether animals themselves possess moral rights. Such a justification for the moral consideration of animals also includes animals to which neither complex preferences nor sentience are ethologically ascribable. In the preservation of ecosystems, our lack of knowledge about how changes will affect the continued existence and well-being of humanity provides the fundamental aspect of caution. – No ecosystem must remain as it is according to a natural law. From nature's 'perspective,' every equilibrium is as good as any other. In this respect, extinction, immigration, and mixing of animal species are natural processes. As such, they are not relevant to us – except for aesthetic reasons. These processes become relevant at most out of anthropocentric caution. Measures that seek to follow such caution, however, risk themselves becoming factors in changing equilibria. Nature conservation, understood as the preservation of a system once found, brings human evaluations (including those of existing fauna) closer to a non-judgmental nature. We may not know whether the changes in the equilibrium would have been good for us. Here, our inadequate knowledge encounters a technological imponderable.

(ii) Violence against animals, *ceteris paribus*, has a brutalizing effect. For reasons of moral education, careful treatment of animals is therefore recommended as a protection of fellow human beings. (A lack of compassion for animals often indicates a psychopath at an early stage.) This can be taken even further: Insofar as humans are not inherently evil, evil usually arises from a failure to consider the many (side) consequences of our actions on others; others often suffer because we lack foresight and empathy. The development of empathy and compassion is therefore a virtue. Empathy is virtuous because it indirectly leads to moral action toward others. The development of empathy thus contributes to the positive development of moral character. Greater empathy does not stop at species boundaries. A virtuous person feels compassion for animals and thus includes them in his or her moral concern, even if this virtue initially aims at avoiding lack of consideration between people. In addition to the environmental protection justification, there is thus an anthropocentric virtue-oriented justification for the moral consideration of animals.

(iii) An anthropocentric aesthetic argumentation is also possible. Morality allows for a range of additional duties that are suggested to others without being as compelling as the core of moral duties. These duties are similar to the appeal for an aesthetic engagement with art. Unharmed animals can be viewed as more beautiful, and the way of life that also encompasses such aesthetic aspects can be viewed as more successful. Here, aesthetic conceptions of life and self (for example, in the context of designs for a meaningful life) overlap with moral duties (for example, justifications of type (ii)).

Justification type (i) appeals to human rational self-interest. The focus here is more on nature conservation and an obligation to maximal possible non-harm, with consideration of the rights of individual animals being subordinate to this. Justification type (ii) applies generally to any animal – regardless of their sentience. Empathy could even be developed with regard to toy animals. The justification itself depends on the strength of the correlation between harm to animals

and harm to humans. Justification type (iii) may persist in hedonistic pleasure; other people may also feel pleasure in hunting trophies.

Overall, an anthropocentric justification of obligations towards animals tends towards a recognition of non-interference (i), and (ii) an expanded doctrine of virtue (cf. Franklin 2005).

A justification of animal ethics that emphasizes morally relevant characteristics such as the ability suffer partially succeeds in forms of utilitarianism in its ethical form (i.e., in a form of calculation that goes beyond enlightened self-interest). This fails partially because the intuitive appeal "suffering is *per se* bad" captures our basic understanding of 'suffering', but overlooks the fact that much suffering is unavoidable and does not contain a call for its elimination, especially the suffering experienced by animals in the context of their natural, species-appropriate lives. What some utilitarians correctly see, however – and in this respect this justification model succeeds – is that some animals are among those affected by our actions (in the sense of being able to feel the effects of our actions). We understand ourselves as agents in coexistence with others. This includes the concept of reciprocal rights and the consideration of others in our actions beyond a calculation of enlightened self-interest. If we have understood this, we can extend our consideration to animals as well (cf. DeGrazia 1996). This is similar to the virtue-oriented anthropocentric approach. What utilitarianism also correctly recognizes in its happiness accounts is that the additional causing of suffering (beyond that which is natural) requires a special justification. While suffering is not bad *per se*, it does seem to require justification if we bring additional suffering into the world. Therefore, all those affected by it must be considered in a utilitarian balance sheet. In such balance sheets, the mass of hypothetical animal suffering will often outweigh the gain in – often relatively insignificant – human satisfactions. The emphasis on sentience as a criterion for moral consideration in some forms of utilitarianism makes them particularly suitable for including animals in moral considerations. Even if one considers utilitarianism in any form to be inappropriate for the foundation of a human-centred morality (for example, because it does not contain fundamental rights in the narrow sense), then a form of utilitarianism could be the way in which animals should be included in our moral deliberations.

One may see a division of ethical frameworks: Discourse Ethics for persons and Utilitarianism for animals, which expresses the special moral status of animals. Those who understand 'rights' in a correlative sense (wherever there are duties to certain actions, there are corresponding rights for those in need of protection) can accordingly speak of 'animal rights' in such utilitarian considerations.

While these general considerations apply to animals in general and their abilities in general, specific ethological and ethical questions arise regarding animals that occupy a special role with respect to humans: companion animals and 'farm animals'. Towards these groups, on the one hand, we have special ethical obligations because they have been removed from their natural environment and weaned from it (but cf. Budiansky 1992). In the case of companion animals that have been adopted into our immediate social circles, their harm appears particularly reprehensible. In the case of companion animals and 'farm animals' their owners also have a greater moral responsibility due to their control over the living conditions of these animals. Ethologically, these animals raise questions about their altered abilities (for example, that of dogs in relation to humans) and the extent to which this results in special interactions, even across species boundaries.

Animals are massively used in experimentation. A lot of the research using animals in experimentation (apart from toxicity tests) presupposes that animals are sentient in a relevant way. From an ethical perspective animal experimentation faces a dilemma with, for instance, pain research: either animals have sentience and pain similar enough to human pains, then the experimentation may give the pain and preference like states of the animal not their moral due, or the animal does not have pain similar enough to human pain, then the experimentation seems useless.

Especially problematic are experiments in the field of psychiatry. They presuppose not only animal sentience, but complex emotional and attitudinal states the human counterparts of which require consciousness and language. If animals could have states sufficiently similar to human psychiatric disorders and are forced into them by research, the mistreatment of the animal mind seems inacceptable.

Both a denial of sentience, which allows experiments *tout court*, and an endorsement of complex animal minds can thus lead to an increase in animal

experimentation. A proper and differentiated theory of animal sentience might delineate which burdens can acceptably be placed on which animals in experimentation.

Some experiments should even be endorsed: those non-invasive behavioural tests used to ascertain an animal's preferences (or like states) in animal welfare research. Even for animal rights activists some such research has to be endorsed as long as there are domestic animals and human/animal-contacts and competition in the wild.

§27 *Subjectless Awareness?*

'Ego' shall refer to the agent of thinking. 'thinking' always means conscious thought in distinction to mentation (mental events which are not conscious). The Ego should be distinguished from the occurrence of cognitive agency on a subconscious level (a 'functional Ego') and a narrative self-representation of a thinking agent's biography and self-understanding: the 'Self'.

A tokening of a certain, specific representation results in a conscious state. This representation is a representation of the thinking agent in *this very act* of thinking (and tokening conscious representations). This is an occurrence (a process) of representations which cover the occurrence itself. The thinking agent as agent is always not just represented but *active* in thinking and representing. Being active does not exclude being represented as active, but this representation *of* the agent is not the agent: one of the traditional problems of having a theory of the 'transcendental ego'.

Talking of a thinking 'agent' and conscious 'acts' does, of course, not mean that conscious acts are actions. Actions result from conscious acts like beliefs and volitions, so these (on pains of a vicious regress) cannot be actions. Talking of 'acts' stresses that consciousness develops in (inner) time as flow and that we are present to ourselves as the agents of our conscious ongoings, even if some content is pressed upon us (e.g. in perceptions) by our situation. Even if some painful event happens to us, we experience ourselves as the ones who 'do feel the pain'. Part of the Cartesian evidence is that 'I am thinking' (i.e. both being and being actively conscious), not that thoughts merely happen to me.

In a conscious act the Ego is thinking agent and part of the total content. We know the Ego in its activity, not its ontological essence or metaphysical nature. Thus – as Kant stressed – awareness of ourselves as cognitive agents is compatible with our ignorance about the ontological nature of ourselves – say as brains or souls.

The Ego, although a functional aspect of a person who is not in doubt about being the agent of consciousness, has *no worldly content* in itself. Worldly content is transcendent (i.e. posited as distinct from the experiencing Ego), even content pertaining to imaginary objects. Worldly content is intentional content, including biographical beliefs. Although the Ego is the Ego of a person, who knows about herself in her self-narrative, the Ego can – with mild imprecision – be said to be 'contentless'. The Ego exports self-related content into the self-narrative, which is transcendent. Precisely speaking the Ego is not completely contentless as its core is the experiencing subject given to itself: it has the crucial – verbally hard to grasp – Ego-content defining consciousness, but has no worldly content.[25]

Consciousness or awareness in humans always involves some form of awareness of the thinking agent (the Ego). In that sense every human consciousness is self-consciousness/self-awareness. A thesis going back at least to Aristotle's *De Anima,* and present in the tradition from Fichte to Sartre as an explicit critique of reflexion or higher order theories of consciousness. This, however, should not – as unfortunately in parts of the 'philosophy of mind' – be confused with consciousness of the self with 'self' understood as the narrative, biographical self-representation of a thinker/person. The 'self' in this sense (and so is 'self' employed here) is not the Ego. The Ego is what often has been called 'the I'.[26]

---

[25] In this vein one may see the rationale in Sartre's thesis that in distinction to transcendent being consciousness is a nothing (of content).

[26] In what one may see as a reversal of terms, but in stressing the same point, Sartre speaks in the same titled book of the 'Transcendence of the Ego' [i.e. the self] with respect to consciousness, which always is immediately [i.e. without reflection] aware of itself. Whether Sartre talks of the self or the Ego may be a philological issue not to be considered here. That Sartre rather reverses terms fits to making a threefold distinction between the self (as narrative or intentional object), immediate self-consciousness, and pure reflections. All these terms have been used and abused in different theories of self-consciousness. Hopefully my use of "Ego" will become clearer in the context developed here. I do not think a completely new term can be helpful because of the intimate relation between our personal use of "I" and self-consciousness. The

Human consciousness involves self-awareness, not just in general as capacity, but in every conscious act. Some animals possess sentience, but not self-consciousness. Some talk here of 'animal consciousness' with the adjective stressing the essential difference to human consciousness, but [as said in the Preface] that concept is misleading, 'sentience' being a more appropriate concept. Most animals aren't even sentient. A few animals (supposedly primates, dolphins, and ravens, and maybe cephalopods) seem to have – at least occasionally – something like a form of self-awareness. How to model these distinctions in animals poses a challenge to cognitive ethology and a philosophy of mind broadly taken. It need not concern a descriptive model of human consciousness, since this does not come with the explanatory goal of a stepwise genesis of consciousness.

In human consciousness we experience ourselves as the agents of consciousness. I may say that some experience happened to me or that I underwent an experience. I was experientially confronted by something. I do not say that some thought (especially in inner speech) happened to me. Sometimes a thought 'springs to mind' or one 'has an idea', but these are not the standard cases. In thinking (in inner speech) I graft thoughts, ponder ideas and questions, and reflect on these thoughts or thought in general (including experiential encounters) In all this I experience myself as a thinking agent. My conscious life not just happens to me. I am not an observer of some conscious life, I am engaged and 'participating' in my conscious life.

The agent of my thought as conscious is not distinguished from myself, i.e. there is no anonymous agent and a personal Ego, but the person experiences herself as the agent of her thoughts. Several functions of agency have been traditionally distinguished (e.g. as 'transcendental' vs. 'empirical' Ego, *inter alia*), but these must not be hypostatized as distinct entities, which would immediately raise the question of their relationships and ontological status. These philosophical terms highlight functional aspects of the agency present in consciousness. This agent is

---

Latin expression "Ego" indicates its role as a theoretical term in a philosophical elucidation of self-awareness. It also helps to avoid confusions with talk of the embodied person or the biographical self-narrative. The Ego will have a correlate in non-conscious mental acts, which may be called the 'functional Ego'. That correlate plays no great role in our considerations here.

not the narrative (biographical unity) in which a person understands herself (the 'self' as object), because a thinking agent is not a narrative (a narrative is a static object). Nonetheless we are conscious of ourselves also in the sense of conscious of our *selves*, as we always understand the thinking agent as a phase of a larger whole we understand as our self.

The unity of conscious content became unified as contents of one Ego. We understand ourselves in correlation to the whole of our conscious life. We generate a self(-description) out of our lived experience. In this way unity of self and unity of content are correlational. Nonetheless the activity of the thinking agent is the activity which establishes the correlation. The relation is not symmetric, only in our understanding do both sides depend in each other. In the broader metaphysical picture the thinking agent depends on the environment it is embedded in (i.e. the referents of some of the representations in consciousness). The metaphysical conditions the conscious agent depends on need not be unified into the same whole the content of its thoughts are unified into.

So, whereas in the human case consciousness is equivalent to self-awareness there may be a level of sentience in organisms which have phenomenal states and distinguish themselves (in different degrees) from their environment and their flock, but lack self-awareness. One may speculate to think of this sentience in mammals to be somewhat like the right hemisphere thinking in average humans; there has to be something *functional* similar to an 'I think' lacking although the step to explicit self-awareness in consciousness. Sentient animals of this kind experience the world from a perspective, but not as a self given a self-narrative and personal point of view, experiencing a subject as the agent/observer of mental life. They live as the (perceptual) centre of their world and distinguish themselves from other animals and con-specifics, but lacking reflexive concepts and concepts covering deliberation and intentional action they cannot experience themselves as 'subject of a life'. Although animals exhibit individual behavioural differences – different 'characters' as some say – animals lack subjectivity *for themselves*.

Phenomenal states present to them not only their situation but also their inner states. Some of these experiential feelings may be inner percepts: representations of their inner overall state (e.g. in coping with a situation), which are in their function meta-cognitive when accompanying their intelligent coping, without

being meta-representational. They need not be sentential nor *about* any representation to give feedback *about* their cognitive engagement and its success or failure. They may show in behaviour indicating surprise, frustration or exhilaration.

An exception closer to self-awareness – if this admits of degrees at all – might be the great apes. They are exceptional in several respects (like learning simple signing and problem solving), and their cognitive ethology writes a chapter quite different even from other mammals. Classifying great apes as "close to humans" will hardly be disputed, although trying to identify the content of their mental states still remains challenging (cf. Camp 2009).

§28 *The Language Paradigm (II)*

Human reasoning and language are recursive. TOM and linguistic conventions require higher order (recursively embedded) intentional states.

Speculation: Recursion may be the single mutation that set humans apart from animals (cf. Chomsky/Berwick 2016, Hornstein 2009). Animals lack TOM and language, but they – or at least some, primates and ravens – have a limited planning capacity which might may require some embedding of intentional states, embedding of limited degree, say 2. Recursion then will be the step to *indefinite* embedding. At least degree 4 is required for mutual role expectations, indefinitely further embedding may be the structure of mutual knowledge, e.g. in conventions. Positing some precursor capacity in some animals facilitates understanding the step to recursion and language. Language itself then is not an adaptation but an effect of a mutation with respect to the adaptive capacity of some planning. Once in place language and TOM equip humans with hugely beneficial capacities like better coordinated actions and cultural transmission of knowledge. Once in place the mutation of indefinite recursive embedding will be inherited.

There is no question that language and culture are evolutionary advantageous and thus have a beneficial function, although supposedly the result of a mutation in a somewhat related area. The difference between humans (human culture) and animals shows that they are not epiphenomena or non-functional. As consciousness is tied to these capacities or – better said – part of the type of mind evolved by

recursion, consciousness is part of this functionality and no epiphenomenon either.

§29 *Animal LOT (I)*

Presupposed in the preceding speculation on the onset of humanity (the human mind) was some planning capacity in animals. This planning involves some form of LOT (to simulate the possible actions and their outcome decoupled from execution), and – one may assume – sentience, since animals of this type exhibit plasticity of behaviours and can attend to aspects of their environment selectively, all of which – one may assume – require sentient awareness.

The ascription of sentience to some animals thus allows for a coherent 'just so' story of human evolution. On the one hand the evolution of recursion leaps from animal cognition to cognition of a new qualitative form. On the other hand, there is some continuity that diminishes the miraculous gap between animals and humans. The speculation can be seen as a paradigm argument for the assumption of animal sentience as facilitating the explanation of the evolution of the human mind.

Part of the building blocks of the human mind are conceptual representational structures (RTM). Some animal LOT phylogenetically contributed a basis for the human mind. A complete restructuring of the then present representational building blocks by the newly evolved recursive mind resulted in human language/LOT, which qualitatively differs from animal LOT, in those animals which have a LOT at all. The language of human thought (human LOT) may (mostly) just be natural language, which is obvious in inner speech, but may extend even to non-conscious mental processes. [A topic to be returned to.]

A productive combinatorial system need not output sentential structures. For instance, a production system implementing a grammar for legal license plates outputs possible license plate combinations. Animal LOT need not output sentential structures, maybe rather combinations of schemata abstracted from percepts. Even if the underlying LOT structure computed with sentential representations, in the mental experiential life of the animal such sentential structures would not be present, as the computational LOT is not present to a sentient mind. What is

present to the experiential mental life of an animal can be phenomenal aspects of underlying LOT combined as combination of percepts. Those animals which have to be seen as quasi-intentional systems with corresponding explanations of their behaviour have to have a rich enough LOT, because there cannot be computations without representations.

Humans can imagine language in inner speech, animals cannot. They cannot imagine sentences, and thus cannot have propositional attitudes as part of their mental *experiential life*. Their unconscious mental states could have sentential structures. These might be simple sentential structures (i.e. lacking complexities like sentential embedding, quantification or counterfactuals). These structures would be operational in their goals and registrations. They can be the target of low-level intentional stance explanations of their behaviour. One might speculate that some more developed animals have a LOT with such elementary sentential structures (say, at least mammals), while less developed animals have a LOT, if any, with a non-sentential combinatory output.

For *types* of a combinatorial system some abstraction has to have taken place. A single situational percept cannot function in general derivations. Animal LOT as derivational system – even with no sentential output – has to use schematic more abstract representations. "Until digitalization has occurred, nothing of cognitive significance has appeared" (Dretske 1999: 201). Heterophenomenology has to try to identify these types.

§30 *Shared LOT Types?*

Conscious experiences and events are not always higher order states. Higher order in the narrow sense is reflexion. Consciousness can be modelled in its representational structures as *first order* (FOT), making use of self-representational concepts.

This does not exclude that these representational resources *depend* on a cognitive architecture of recursion and developed self-concepts. An architecture of GWS and recursion results in a consciousness that in all its first order states involves representations of the cognitive agent/observer and the state quality/type itself. GWS includes self-representation. Thus, consciousness may depend on this

architecture, creatures lacking it also *lack consciousness* of the representational form explicated in FOT. Crucially they miss the self-concepts. Their type of awareness – if any – *cannot* be understood as like a first order consciousness, just lacking a higher order capacity, since full blown (human) consciousness is first order consciousness, most of the time (i.e. when not engaged in reflective thinking).

TOM and language yield consciousness. They may have resulted in a restructuring/re-wiring of cognitive architecture (like in Dennett's image of the 'Joycean Machine' of always available inner speech). Sentience in humans *is not a remnant* of previous sentience in this form.

Conscious self-access is of one type: 'I am conscious of me ___' does not come in different qualities. Qualities and types occur in the states one is conscious of: seeing, hearing, seeing something red, something blue. The content of consciousness harbours a manifold of qualities and represented objects apprehended by states of different types (including believing, desiring etc.), whereas the aspect of being conscious in these acts is always the same.

In

(1) I am conscious of me believing the pen is black.

the part "the pen is black" expresses the content of the belief, and thus the content of that act of consciousness. This content is present in a belief-type state. The whole state 'believing the pen is black' is content of that act of consciousness. Also, content ('agent content') of that consciousness is 'me believing the pen is black'. We have thus three types of content:

(i) objective content (a sentence or a pictural representation)
(ii) state-type content (the mode in which the content is present)
(iii) agent content (representing the thinking agent).[27]

Representational structures are also important because of their functional role. Two sentential representations may share their referential/informational content, but differ in the role they play in the mental life of a thinking agent. This

[27] This distinction has been marked early in the philosophical tradition. Making this threefold distinction Reinhold (1791) considers the 'foundation' of all philosophical knowledge.

difference has been the cornerstone both of theories of oblique contexts as well as theories of essential indexicals or 'indicators'. A temporal expression may refer to the same time as the indicator "now", but the functional role of beliefs involving the different expressions may vary widely. Similar remarks apply to the subject's use of "I" and the use of a representation that objectively identifies the thinking agent for an audience. A theory of the mind interacting with others and reality thus must involve a theory of indicators and the representational format of self-access. The functional role of a representation in consciousness does not reduce to its referential content. Therefore, indicators like "I" and "now" are ineliminable for such a theory.[28]

Within some version of RTM the attitudes are taken to be sentential/propositional attitudes.[29] The content of (conscious) mental states can be expressed by sentences of some internal or external language. These sentences as employed in a situation of usage have referential content (computed by anchoring indexical expressions to appropriate entities). As representations these sentences also have a mode of representing their referential content. The same referential and even the same semantic content can be represented by different sentences. Understanding attitudes and content as involving such (tokenings of) sentences accounts for the hyper-intensionality of attitudes. The way or mode of representation often results in a different functional role of the individual mental event. Informative identity statements have the same referential content as uninformative, but consist in a mode of representation with a unique functional role. E.g. knowing 'Monday is Monday' and knowing 'Now is Monday' differ not in their referential content, but in its mode of representation, which accounts for the different functional roles corresponding beliefs may have. RTM, thus, allows for indicators/indexicals in representations of mental content, without denying that there is an objective referential content. And objective referential content does not exclude a perspectival mode of representation from a subjective point of view. Accounting for the

[28] As has been stressed by several authors in the philosophy of mind, for instance in Castaneda's *The Phenomeno-Logic of the I.*

[29] Some version of RTM may be appropriate for (parts of) cognition and representational structures, without a commitment to a general computational theory of the mind, which faces many controversial issues, not relevant here.

representational structure and status of self-awareness does not require the introduction of special objects of self-ascription in self-awareness (e.g. special properties or predicates vs. propositions or sentences) nor special individualized kinds of the attitudes (like 'believing-of-oneself that ___' vs. 'believing that ___') beyond the distinction between referential content and representational mode. Representational content is referential content represented in some way/mode/character. The elucidation of self-awareness proceeds at the aspect of the cognitive agent of consciousness, not its conscious content.

The pen, in (1), is an enduring object. A perception of the pen is a fleeting experience. A constant perception unites a series of successive or overlapping perceptions into a longer period of a constant perception of the pen as the objective ground of the series. In the series or movie an enduring perception of the pen corresponds to the enduring pen. In a similar way the Ego presents the thinking agent. Conscious acts are united (instantly) diachronically by retention and protention, and synchronically by a form of immediate conjunctive equivalence of the form "$I(\varphi \,\&\, \psi) \equiv I\varphi \,\&\, I\psi$" into the stream of consciousness. We experience a *single* (not several co-occurring) and *lasting* Ego. This experienced lasting Ego does not contradict the fact that each conscious act has an Ego, no less than the enduring movie representation of the pen contradicts the presence of a pen picture in each movie frame. The subjective experience of a lasting subject of consciousness corresponds to the lasting thinking agent as the underlying object. The Ego is the self-presentation of this objective ground, be it the brain, the soul, or whatever.

For perceptual mental states awareness of the state is constitutive. Some structure of self-representation has to be involved – or better: the belief has to be involved in this structure – to switch from an unconscious belief to a conscious belief. Whereas we can have non conscious beliefs, mental states involving perceptual/phenomenal concepts (like: 'seeing', 'hearing' …) are *always* conscious. The concept 'black' involved in the belief

(2) I believe the pen is black.

is – if not different – though only partially (referentially) identical to the phenomenal concept in

(3') I see a black pen.

Phenomenal qualities occur only in conscious states. (2) can be represented by words (i.e. merely sentential). Phenomenal qualities are represented in a different *way*, which from a subjective point of view can be characterized as analogue holistic rather than symbolic digital. Their informational content cannot be exhausted by a few sentences, i.e. they possess high resolution even of only co-present features. Within that type of representation resolution and informational density can vary widely (from high, in perception, to low, in imagination).

An *equivalent* form of (3') is

(3) I am conscious of me seeing a black pen.

because phenomenal states are always conscious, while (2) and (3) are not equivalent, because beliefs can be unconscious or conscious. In this perspective, an act of seeing does not become conscious by being the object of a higher order state or by *becoming* embedded into a more complex structure involving self-representation. It exists *only* in those structures. There are, then, different constraints on the mental use of concepts which are not phenomenal and those which are. The presence of an 'I'-symbol may be part of an elucidation of the presence of self-awareness (as ever-present aspect of consciousness in humans), still there are other *constraints with respect to phenomenal qualities* in consciousness, constraints of a type and function which – although we have no idea how – may have precursors in those animals which have awareness (without self-awareness).

Theoretically there are two options to model the difference between an unconscious belief involving 'black' and a sensory state involving 'black':

(i) There are two referentially related but not identical representations, representations with phenomenal aspects being only present in conscious states, the entering of consciousness by a so far unconscious belief requiring a substitution of representations. Some representations yield by their tokening phenomenal content, and result *inter alia* in perceptual beliefs by further processing.[30] Those animals which show sentience

[30] Chisholm in *The First Person* tries to capture this by the model of self-ascription of phenomenal properties. As said above, these are not required to account for consciousness itself, but they may point to a typology of representations. Rey's 'Computational/Representational

have to have some form of internal representational medium (a precursor of human LOT) which contains such symbols.

(ii) There are not two distinct representations. The difference between a representation of 'black' being present in an unconscious belief and being present in a conscious state works by the phenomenal aspect of a representation of this type being suppressed in unconscious state, resembling the suppression of the phonological component of a natural language sentence in unconscious mental acts.

Option (i) could be classified as a representationalist version of 'qualia realism', in which the special status of qualia is modelled by usage of a special type of representations. The disadvantages of this option seem obvious: there would be a huge amount of computational effort in the needed substitutions, and one needed a further explanation how this doubling of representational resources came about. On option (ii) cognitive architecture and its mechanisms of working on some part of a complex representation and not others avoids these difficulties. Option (ii) should be preferred on reasons of parsimony. Given option (ii), however, one can no longer argue that animals which have a precursor of human LOT enjoy phenomenal states for *that* reason.

Creatures without consciousness may still have phenomenal concepts in representational structures of a simpler kind. This is the question of original sentience again. They must have a form of LOT that suffices for a sufficiently complex representation of their environment and their own body. This LOT precedes language and some of it might still serve in unconscious modular processing in humans. Humans might represent mostly in natural language as LOT_1, but also in some module specific LOT_2 different from natural language. This module LOT_2 might be still in place, because the restructuring of the human mind by natural language did not reach all modules, as there might be no evolutionary incentive of re-wiring all the modules. Some more complex recursive human LOT_2 occurs in the derivations of the language modules, underlies the derivational system of the language faculty. Generative Grammar claims: The language faculty derives sentences (as *structural descriptions)*. These possess syntactic, semantic and

Theory of Thought and Qualitative States' (Rey 1997) incorporates sensations and phenomenal concepts into a sentential RTM; cf. also (Maloney 1989: 228-51).

phonological features. In procedures using sentences not always all features need be used. There are purely syntactic procedures, syntactic/semantic procedures operate at the *interface* to the *conceptual system*. Syntactic derivations and descriptions fit to the constraints of the conceptual system (cf. Chomsky 1995). The conceptual system may also contain some phylogenetic old LOT and human specific concepts. In case of outer or inner speech the phonological features are passed on to *Spell Out*. These phonological features can also be muted.

Therefore, natural language may very well be the representational medium in unconscious mental states using *natural language sentences* as LOT_1, with the phonological features muted. So, even non-verbalized thought, as intuitively given in introspection, may be processed and represented by sentential structures with the phonological component (the imagined sounds of inner speech) suppressed. Thoughts may be expressed in sentences which are neither covertly nor inwardly uttered.

That we cannot remember the time of our life before the period of acquiring language testifies the restructuring of our mental life by language. Most people will believe toddlers to be sentient, but this sentience – if given – is cut off from our present phenomenal life and phenomenal self-concept. We cannot say 'what it is like to be a toddler' although having been one ourselves. Animal sentience should be as least as far off in conceivability as toddler sentience.

§31 *Animal LOT (II)*

Language developed stepwise with the occurrence of indefinite recursion, extending a previous LOT. A structured LOT with indefinite embedding may serve as the basis of nested intentions and beliefs, which allow mutual knowledge and conventions.

Once a proto-language has developed it can gain complexity if children can acquire it quickly. Thus, a Language Acquisition Device (LAD) is adaptive, and a specific language module for complex natural languages evolves. The deployment of a LAD presupposes some language to be in place, thus language supposedly developed stepwise by ever more elaborate proto-languages.

These, especially their semantic conventions, presuppose higher-order intentions, which originally should have been non-verbal, even if acquisition of a full capacity for natural language might result in a cognitive architecture rebuild, such that after acquiring the full natural language capacity intentions (conscious and unconscious) are represented in natural language.

The LAD could be a stepping stone with no further function. It might even be discarded like a setup program for a computer program one installs.

One might extend the speculation [of §29], thus, by the idea of a *stage* of protohumans with some more elaborated (recursive) LOT different from natural language being in place. Supposedly this LOT had to share some structure with animal LOT out of which it developed. The question related to the issue of sentience is whether this animal LOT – presumably preserved today – has enough structure to base sentience, and how we could come to know this by empirical research and model building.

§32 *The Gricean Paradox*

Given language and conventional meaning a speaker can use a sentence ψ to express her belief that ψ. She can also use ψ in a situation to bring across something different φ by exploiting some features of the situation (e.g. by grinning while uttering ψ). Conventional meaning and speaker meaning thus can be kept apart and this plays a role in a theory of conversation in language pragmatics. That a Gricean analysis of speaker meaning can be part of language pragmatics should not be controversial.

Controversial is a Gricean account if it attempts to ground linguistic meaning in speakers' intentions. The intentions employed in the Gricean account involve propositions and 3rd order intentions. Even for introducing *further* conventions (in Lewis' analysis) once language is already in place this should not be problematic. A Gricean account is questionable as an account to explain original linguistic meaning. The intentions and intentional contents involved are so complex that even if one grants some complex LOT operational before the advent of language this complexity is too much.

A better explanation – or at least 'just so' story – about the arrival or natural language as we know it may proceed by several stages of co-evolution between proto-languages, human (proto-)LOT and (proto-)LADs.[31]

A Gricean taking up the challenge had to stepwise build up the intentional complexity (as J. Bennett 1976 attempts). Even if such a stepwise Gricean theory does not convince us as a theory of meaning it can contain ingenious ideas for precursors of complex intentions like humans have them.

If a stepwise Gricean account was feasible, it would land itself in the paradoxical situation that animals could ascent to linguistic meaning – but they did not. So, even if one acknowledges animal LOT of some form, and maybe human (remnants of) LOT distinct from humans using natural language for thought, this LOT should be conceived as of limited complexity to avoid the paradox of the successful Gricean. The features of propositional content and higher order intentionality at least are doubtful.

§33 *The Language Paradigm (III)*

Once natural language is in place the constituted thinking format of inner speech replaces former structures with natural language thought and consciousness becomes a 'Joycean Machine'. The requirements of language have shaped then both human physiology (e.g. in the vocal tract) as they must have shaped architectural modifications in cognition and thus the brain (cf. Deacon 1997).

Given the complexity and systematicity of natural language the language-based thinking format will boost expressive power, versatility and cultural transmission of human cognition. In comparison to animals a new quality and a new level of cognition will be reached, which itself shares little with its animal precursors.

At the former stage (the stage of proto-humans), at least, one has to work with non-verbal complex thoughts. Given these a *partial* Gricean account of the origin of the first conventions and the beginnings of communication seems feasible. This does not mean, however, that the primary function of language is

[31] We do not have a complete picture yet (cf. Tallermann 2005; Jackendoff 2002: 231-65), and we need not wait for it for a distinction between human minds and animal minds today.

communication: (i) because language once developed serves at least at the same time as an internal medium of representation, and (ii) because the more complex (recursive) LOT used in the development of proto-languages has to be in place as an internal medium of representation, which then might have been almost completely substituted by natural language.

Language changed human cognition dramatically. These changes hinder us in *conceiving* the cognition of animals, which lack language. Whatever – if anything – goes on in the mind of an animal (from its 'point of view') is not a sentence. Animal cognition is *not* human cognition minus language. Animal brains are not just human-like brain with the speech areas and modules missing.

§34 *Human Cognitive Architecture*

In a conscious thought one is conscious both of the content of the thought as well as *having* a thought with that content. For the latter aspect one has to be conscious of the representations present, and these can only be representations in a language (or perceptions embedded in a language frame), 'deeper' levels of mental code do not enter consciousness.

This is one main argument for conscious thought to be linguistic. One other paradigmatic argument for the conclusion that *all* thinking is linguistic is that thoughts can 'enter' consciousness. Mental processes compete for access to the GWS: pressing issues, debugging problems and unknown/unresolved encounters (external or internal) may suddenly become conscious. If they were not linguistic the access would require an extra procedure of translation (lexical lookup etc.), which may compromise the (temporal) efficiency of on-the-spot problem solving and attention direction. Cognitive processing presumably runs smoother when using one representational format instead of two, and the linguistic format is needed anyways.[32]

There might be some translation and some LOT-representing present, but not in central cognition and the GWS, but in the encapsulated modules of the mind.

[32] Ignoring his then HOT there is much to be agreed to and learned from (Carruthers 1996), but cf. (Fodor 1998a: 63-74).

Their inner workings may use whatever representational format as long as their output and messaging to overall cognition are translated into the natural language format.

Human cognitive architecture may contain two types of representational interfaces:

(i) Transducers are operative in converting sensory signals into a format fit for processing in the modules,
(ii) Other transducers (or 'translators') are operative in module output entering general cognitive processing.

And – one may add for completeness – cognitive architecture supposedly also involves 'compilers' by which an acquired skill descends from 'knowledge that' into 'know how' executed by compiled procedures in a module.

In a computing machine with von-Neumann architecture (vNA) the signal bus, the CPU, cache and Assembler (machine instructions) are hard-wired functionality. In cycles data and programs are fetched and executed given the basic instructions and architecture. Programs, that have been compiled to the basic instructions, are loaded and executed. Memory external to the central processing may be vast and stores program codes and data. This storage is inert. All activity takes place on the motherboard with the interacting elements of the basic architecture. No re-structuring of the memory needs to take place in each situation of program execution. New situations are handled by new programs which have to be stored in external memory, the machine being a Universal Turing Machine.

In contrast the brain has *dedicated circuits* which compute specific functions. To compute a new function a new network has to be set up by rewiring. Standard and habitually used functions are encapsulated in corresponding dedicated *modules*.

Rewiring is slow and thus no way to deal with situational or occasional tasks. Central processing (in the GWS) has to have a structure more similar to vNA:

(i) a central fetch (bind and broadcast) functionality and computational processing of data fetched either from modules or from memory,
(ii) a read/write memory structure of not-dedicated brain networks; large parts of the brain may be physically similar because they serve as memory to load (symbolic) data and programs.

Higher cognition corresponds supposedly to global processes in the brain. These resemble vNA. Inasmuch the computer model of cognition may not just be a functional model but a model with anatomical analogies. Neural networks – somehow – implement the GWS architecture, neural structures of a not task dedicated sort. Neural nets which compute functions – often studied in cognitive neuroscience (say, in perception) – constitute the automatic, unconscious processes in the brain. These nets may be Bayesian networks, which compute representations from sensory input. These networks in part resemble the models and computational paradigms of Machine Learning. Coordinative ongoing (situational and plastic) control cannot reside there. Control and conscious deliberation in inner speech is implemented in the GWS architecture, which resembles serial processing in vNA.

The common complaint that the brain is no computer as we know it from our desktops although, of course, partially right misses this. In its modularity of dedicated parallel modules and automatisms the brain does not work like such a computer, but in higher cognition and human brains' crucial capacity (namely, consciousness) the brain's structure may very well resemble vNA and PRAMs.

The 'plasticity of the cortex' (in case of partial damage) that some region can step in for another region without loss of functions *could* indicate that the cortex in large parts fulfils a role similar to storage in vNA computers, where knowledge can be copied here and there. Turing in his 1948 report "Intelligent Machinery" observes the unorganized cortex, and argues that the brain could implement in parts a universal BTM.

Modules containing stores of specialized knowledge (e.g. of psychological concepts or grammar) may be unaffected by the place where these representations are stored. Computational subsystems (actively computing modules) might be transferable to other locations the more their computations are encapsulated and accessed only by few – if any – other computing modules and central processing. If they depend on local neural connections cortical damage may show in corresponding misfunction. Thus, 'plasticity of the cortex' could be an argument that there are more representation storing than computational modules. An animal missing a cortex with plasticity could indicate, for a sceptic, the absence of sentience, because it may indicate the absence of a global workspace.

The distinction between processes in the mind's modules and general processes in GWS reminds one of Descartes' distinction between those mental events belonging to the body (namely perception and feeling) and those belonging to the soul (rational thought). General processes are neither informationally encapsulated (by definition) nor can all be deterministically algorithmic (cf. Fodor 2000), because they entail evaluations (say, of coherence), judgements, practical deliberation, and decision making.

Not surprisingly cognitive neuroscience (cf. Gazzinga/Ivry/Mangun 2019) delivered on perception and motor control, but not in comparative detail on the correlates of rational thought. Apart from the idea of the pinnal gland Descartes' distinction still cuts at a vital difference.[33]

Consciousness is *somehow* implemented in brain states (cf. Shoemaker 2007). Theories of neurocomputation, thus, are highly relevant for an account of consciousness. Given the complexity of brain structures and our limited knowledge of the details of neural computation, neurocomputational models have to make use both of abstraction (from implementation details) and idealization in the sense that they propose *a model how* the brain might compute some function which the brain – say, in vision – *actually* computes, but supposedly not literally as the model says (cf. Williams 2025). Thus, the current stage of neurocomputational theory centres on 'how possibly' explanations and not on reductive explanations based on the details of an implementation (like in vNA hardware). For a PRAM with vNA computer technology explains the functional design in implementation details (cf. Petzold 2000). For a CTM and a theory of consciousness the focus stays on abstract cognitive architecture and individual algorithms, not on a verified implementation.

[33] Even dualism of some form might be a remote but unrefuted option – one not to be considered here. Descartes (in the *Meditationes*, his replies to objections to them, and in his *Philosophical Letters*) admits that animals (mammals, at least), although they lack a soul, because they lack thought, because they lack language, have some sentience; because of the similarity of their sense organs to ours they can enjoy sensations (i.e. it is not the case that 'nothing is going on' in their mind). Descartes is often erroneously lumped together with some of his followers who have claimed that all animals lack sentience.

The brain has only finite storage, and thus has at best the computational scope of a Bounded Turing Machine (BTM). This in itself poses no devastating limitation: BTMs recognize Context Sensitive Language, to which set natural languages belong. How the brain stores (intermediate) results of computations, however, poses a problem, as synaptic connections gain weight slowly and synaptic activity cannot be constantly maintained as a form of short-term memory (like the cache of a PRAM). A simple half-adder might be more easily modelled by wood and pinballs than by a neural network (cf. Gallistel/King 2010, Marcus 2001). Understanding and forming sentences in natural language also poses unanswered challenges to neurocomputational models, so far (cf. Jackendoff 2002: 58-67). NNs of the kinds we know from ML and connectionist modelling fail at processing natural language.

A conscious state is phenomenally linked and refers back by retention to previous conscious states, as phenomenology of inner time revealed (Husserl 1928). In information terms this should lead to exponential growth of informational content. Under the assumption that the brain stores memories in NNs using synaptic weights one may calculate the brain's storage capacity and from the comparison to required storage demand of a person's biography – even neglecting the individuality of each person's biographical information content – argue that the brain as current theories conceive of its workings has insufficient storage capacity for a mental life (cf. Knight 2025).

Algorithms behind basic epistemic capacities like belief update and revision are sophisticated and exponential in time complexity in the number of beliefs in the belief system to be maintained (cf. Bremer 2025: 87-112). The complexity bottleneck can be partially circumvented by switching to parameterized versions of the otherwise infeasible algorithms (cf. van Rooij et al. 2019). Our knowledge of neurocomputation of any algorithms of general cognition, however, falls short of a level towards any serious idea of implementing the 'Joycean Machine' or BTM computations in the brain. Theories of neurocomputation so far identify computational NNs in the brain, but current knowledge of the brain's structure and memory do not live up to the demands of computability and a CTM.

Setting up the 'hard problem' (of explaining phenomenality in neurophysiological terms) some philosophers (like Chalmers 1996) have surmised that the hard

problem is hard in contrast to the 'easy' or 'soft problem' of finding neurophysiological correlated to the functional and computational states of the mind, which allow to explain them (as implementation basis). One may disagree: the 'soft problem' has turned out quite hard once algorithms of general cognition (i.e. beyond perceptual processing) are considered. The 'soft problem' is the seriously 'hard' problem.

In the supposed 'hard problem' of consciousness the picture presented consists of outlining some neural processes and *then* asking how they yield consciousness. This picture suggests two events: the set-on of the neural process and then the emanation of consciousness. The second event then seems mysterious, its aetiology questionable. Like 'qualia' cannot be separated from conscious states with their functional role – there are not two ingredients which can be separated – the same applies to the neural correlates of consciousness. The on-set of some neural process *is* the on-set of consciousness. These neural processes are the correlates of our conscious experience, better to say: they *are* our conscious life. Outlining when such processes occur and where means outlining the conditions – necessary and sufficient in a complete theory – for consciousness. One cannot witness the stepwise set-on of consciousness from a 1st person perspective. The 'hard problem', though, trades on imagining this epistemic impossibility. The 'hard problem' is an artificial pseudo-problem.

A theory of the conditions of consciousness (brain structures in the cortex and thalamus, temporal coding of specifiable frequencies etc.) is a *complete* synchronic theory of consciousness, supplemented by a diachronic theory of its evolution. This theory can be established by the experiments of neuroscience with test persons reporting what they experience during brain scanning, and by an evolutionary theory of the co-evolution of language and consciousness. The resulting cognitive architecture enabled consciousness and human exceptionality.

§35 *The Versatility Paradigm*

The degree of complexity or number of resources in animal LOT determines whether the animal has sentience of a type. The neural system of an animal and its interconnections have to have a degree of complexity that allows for

representation of sensory modal units (like sounds or images) and for internal affective units (like evaluations and bodily feelings). It is necessary for sentience that an animal can 'do something' about its situation. Mere conditional behaviour (like in bacteria or sponges) needs no sentience. To 'do something' about its situation an animal must represent parts of the situation (salient or relevant for its metabolism, say), update these representations and (re-)evaluate them on the spot: a fixed situation type schema and value can be covered by a mere conditional behaviour. Animals with a neural substrate enabling them to represent in this way could be further up on the ladder that leads to sentience. They need not be sentient, as even some flexible behaviour can be automated, as witnessed by AI systems. Surely animals below this level of neural differentiation cannot be sentient.

Another anatomical complex of features relevant for the origin of sentience is the sensory apparatus of an animal. Once the animal possesses *multi modal* sensory representations there follows the need for intermodal coherence and thus for at least some degree of *central* processing (cf. Humphrey 1992, 2022; and §34). Multiple sensory representations do not *entail* sentience, since they could be stored in a vector like data structure without overall integration, where algorithms with motor output only use their respectively relevant subpart of the vector.

§36 *Versatility in Behaviour*

Versatility of behaviour implies that behaviour is not fixed to a conditional motor output. This raises the question how one behaviour is occurring rather than another one also in a situation of this type (the type being relative to an animal's BL/DL state space and relevant discriminations available in a situation, its 'umwelt').

Lacking inner speech as medium of deliberation and having inbuilt rationality rather than weighing decisions rationally, even a sentient animal with plasticity of behaviour is not an 'agent' (in the sense of the theory of action). How then plasticity, and what takes the role of choice?

One option to model this plasticity may employ opportunistic default principles. These principles express *a pattern* of animal behaviour. They model plasticity by an explanatory schema which takes an animal's registrations and goals into

account, but does not ascribe an intentional action to an animal. They take a realist (not instrumentalist) stance about the presence of animal representations and experiences in the situation. An animal also has representations of dangers and different strengths of goals (in some ordering). An opportunistic approach could then say: "Given the registration of an opportunity to achieve a goal (i.e. registering a likely antecedent condition to attain the goal state) go for the opportunity or any other available with a comparatively evaluated goal unless (this is the defeasibility) a more preferred goal becomes compromised (e.g. by putting yourself in danger)."[34] The anthropomorphic picture of such an animal 'going with the flow' of its situation, if not esoterically loaded, may be not far off the mark.

Assuming a complex enough environment and goals of equal interest a situation of similar type need not show the same behaviour pattern. The behaviour on an occasion may depend on a contingent order of observation, distractedness in observation or goal discernment, or be random to a degree.

The representational repertoire of such an animal (its LOT) must contain types for behaviour *onset* and *concurrent adjustment*, and, of course, representations issuing in experiential joyful states. These states and their interdependence correspond in the mental life of the animal to the opportunistic principle above. How or whether the moment of going for opportunity No.1 against opportunity No.2 is phenomenally experienced by such an animal remains difficult to conceive by us, as we in some form or others always *comment* on our choices in similar situations. Nonetheless, this model shows that without ascribing human like intentionality (the principle *only* refers to registrations and goals) plasticity of behaviour can be understood as between deliberated choice and mechanism.

An animal with a type of memory registering the frequencies and maybe temporal distance of a behaviour (in the past), none of which need be accessible mental content, may *experience* joy to a greater extent when a goal that was not fulfilled for long can now be realized (e.g. cows entering pasture after winter time).

---

[34] It should be clear by now that *this* imperative is phrased from the human perspective *as if* an animal would follow it. Expressions like "yourself" address persons. The animal needs no phenomenal mental event resembling this imperative, because it is constitutionally embedded in its umwelt, always ready to go with the opportunities.

§37 *Animal LOT (III)*

In the flow of consciousness we find not just inner speech, but dominant is the flow of experience in the different sensory modalities. Sensual experiences are not verbal. The respective perceptual representations (percepts), however, may well be embedded as icons or by pointers in a sentential frame (pointers like in low-level computer programming). The sentential (digital/symbolic) nature of consciousness in its computations and processes does not contradict the presence of analogue percepts, by a further code that is manipulated by indexes and pointers resembling algorithmic procedures in AI.

Percepts and experiences are most likely what we share with mammals and birds. They have to represented in some carrier structure of LOT, which may be similar to such representational carrier structure in mammals and birds. Advocates of sentience may thus think that the phenomenality of those experiences founded in the LOT concepts may be present in these animals as well.

The LOT types of such representations are LOT concepts. Some of them are referentially tied to properties in reality. Some of those constitute the core of the meaning of our observation vocabulary.[35] Some meanings are tied to a purely human LOT concept (i.e. a concept expressed in natural language). In semantics, therefore, we meet both direct referential anchoring of some meanings as well as meaning holism. Even directly referentially anchored meanings are part of a holistic semantic system, so that human concepts differ from animal schematic representations, which are concepts in animal LOT, but [as said before] only concept like with respect to human LOT. Sharing some LOT up to similarity, therefore, does not entail sharing concepts in the narrow sense, but could entail *sharing similar percept content*. How this content is inwardly *apprehended* by an animal is the question of the experiential unity of sentience again.

[35] Of first degree, as even theoretical vocabulary of some theory may be observational vocabulary for a theory level founded on this; on observation, percepts, perceptual beliefs and coherence cf. (Bremer 2008: 63-71).

§38 *The Priority of Behavioural complexity*

Given a (vague) commitment to an ontological materialism the advent of sentience depends on complex neural structure, as the advent of consciousness depends on them. These can also be identified by neurophysiology from the outside, with respect to consciousness and with respect to sentience. As we have access to consciousness we can ask, which neural structures in the human brain play a decisive role in its advent. As we have no access to the first level of sentience in animals, we do not know which brain to look at. Methodologically, therefore, behavioural criteria for the ascription of sentience have to *precede* the identification of brain structures. We have to know first which brain to look at. Starting from complex neural structures one might be led to postulate sentience in strata of animals on very general analogies to human or grown mammal brains. As said before, this can be highly misleading. Cognitive ethology must precede neurophysiology.

What use would be a sentience that does not show in versatile behaviour? In the sense of 'conceivable' employed nowadays widely in philosophical thought experiments one may find it 'conceivable' for there being sentient animals without complex plastic behaviour. Form an evolutionary perspective this seems so unlikely that complex plastic behaviour constitutes a necessary condition for sentience.

Once some hypotheses on neural correlates of sentience have been put forth the placement of the advent of sentience can be predated in the evolutionary tree, maybe. This depends on finding such structures in the brain of some species. Their absence is evidence of absence for corresponding phenomenal states. The claim that some species developed a functional but not physiologically equivalent neural system needs empirical confirmation and justification in a hypothesis on adaptational functions. All this will be a question of reflective equilibrium between these approaches, and between these and AI (including 'Artificial Life', cf. Boden 1996, Levy 1992). Mechanical and algorithmic implementation constitutes *prima facie* a counterargument to some process being accompanied by awareness of some kind. The exhibition of some behaviour in itself does not suffice to ascribe sentience. In case of fishes and pain: given their lack of cortical structures which correlate in humans to pain and an explanation that their

nocifensive *behaviour* (behaviour that shields body parts from further damage) stems from spinal cord circuitry, it can be denied that fishes are able to feel pain (cf. Key 2016).

§39 *Varieties of Sentience*

A human person enjoys consciousness and sentience as a member of the species, like any other human person. As said, animal sentience need neither be identical to human sentience nor comparatively conceivable from our perspective. Actually, the situation might be worse. Animals belong to different species in different strata. Nothing guarantees that mammal sentience is identical to bird sentience. What should be identical to speak of 'sentience' as a *genus* at all is the presence of phenomenality. This presence of phenomenality, however, does not explain sentience, but rather re-describes what we understand by 'sentience'. The different strata and species of animals surely have phenomenal states of different qualities, if they have any. The common denominator of their respective feelings and experiences reduces to the claim that 'something is going on inside of them'. We might have parts of an evolutionary theory of the occurrence of sentience in general (cf. Humphrey 2022), looking at sentience from the outside objective side. A developed account of this sort will look at the clear cases of sentience first (say, grown mammals). The theory will focus on the *that* not on the phenomenal *what* of sentience. How far back in the tree of life it could trace sentience will be a more difficult second step. If we only had a good theory of the subjective side of sentience in general, the next step could be to look at different kinds of sentience. Possession of sentience is an all or nothing matter. Still the scope and intensity of a species' phenomenal states may vary widely. A smaller scope of phenomenal states will not disqualify an animal as sentient, but this smaller scope, if identified by us, could have consequences for animal welfare and differential treatment of animals.

§40 *Human Dual LOT Coding*

Rational thought and general cognition use natural language as LOT (LOT_1). Besides this there is a LOT_2 used in (some) mental modules, perception, and playing

a decisive role in the LAD and the lexicon of natural language (cf. Bremer 2008). These are computational systems, and following Fodor's slogan "No computation without representation", these systems have to use a LOT_2. LOT_2 used in natural language acquisition can have less complex structure than natural language, its role in the LAD is to allow generalizations over syntactic variables and categories.

Human cognition thus employs two representational systems, even if unconscious thought more likely will be expressed in natural language (LOT_1). With natural language as LOT_1 being the constitutive vehicle of thought, thought *production* employs the lexicon and the syntactic system of LOT_1. A content (concept of LOT_1 or LOT_2) is chosen at the start (typically a predicate type serving as a sentence frame), the lexical look-up delivers the semantic/syntactic features for formation of a larger unit (by Merge, according to GG) following the constraints laid down in the lexical entries and the general rules of composition, further content inclusion forming a larger unit and derivational transformations (by Merge and Move) result in a structural description (of a sentence) containing syntactic, semantic, and phonological features (like in a three part vector). Phonological features can be sent to Spell Out (in case of inner or outer speech). Semantic features may interact with the conceptual system at large. In case of declarative sentences truth conditions are computed in correspondence with the syntactical composition.[36] Inference depends on syntactic features and derivational rules (like in a logical formal system).[37] The syntactic component of the language faculty

[36] To be understood in the tradition of syntax-semantics correspondence of *Montague Grammar* (cf. Montague 1974).

[37] It is a mistake to doubt the inferential role of inner speech because it occurs as phonological imagination. The phonological imagination is just one part of the current structural description. Inference can employ the other parts. In overt speech recognition phonology does not block inference, which would reduce communication efforts *ad absurdum*, but drives transduction (by parsing) in a complete structural description of the heard utterance, with the relevant syntactic and semantic features.

It is also at least misleading to assert (in a tradition that goes back at least to Vygotsky's studies in the 1920s) that inner speech is "interiorized" outer speech (cf. Vygotsky 1986: 86), because this invites the idea of temporal precedence of developed outer speech that then (supposedly on hearing oneself speaking) becomes internalized. Rather, in acquisition of a specific natural

accesses the lexicon and composition of sentential structures accesses semantic information (like semantic features) from the chosen units, which play a role in syntactic principles (like 'Full Interpretation' and 'visibility' of subcategorized or bound elements). Composition, thus, following general syntactic/semantic constraints proceeds as interaction between the components of the language faculty.[38] We do not have fully developed thoughts in the semantic component which then get expressed in some language. This *performance* model sees natural language (and thus its complex syntax) at the centre of thought production and processing. This implies and explains, again, that languageless creatures have no complex thoughts.

We have no access to LOT_2 and to compiled algorithms in *know how* or module computation (cf. Fodor 1975, 1983). We experience the phenomenal features coming with some representations of LOT_2 as carrier structure of percepts. In inner speech or thought we can reflect on these contents. In a similar fashion some animals *experience* the phenomenal features coming with their LOT, being sentient. They experience perceptions with percepts as units. They are missing reflexion and knowledge *that* they experience, because this representational medium is not directly accessible itself. As they cannot express or describe the type of state (type of sensory content like smelling, or getting information about) they experience, the specific type occurrent has to have a specific phenomenal quality itself. There has to be a distinct phenomenality to the types of sensory states available and to the types of BL and DL states, without which the animal could not recognize them. Animals capable of planning must be able to run mental simulations which the animal can recognize as simulations (cf. M. Bennett 2023), in contrast to hallucinations.

Animal sentience by reflecting on the necessary unity of its mental content thus supports the thesis that mental event types (in humans: beliefs, seeing, desiring, intending etc.) have a phenomenal quality in themselves.

---

language both its phonological form in outer and inner speech (as imagined outer phonological form) are learned simultaneously. With this thoughts appear in the child's conscious mind.

[38] The details are of no importance here. The rough story, so far, is compatible with different version of GG, maybe (dropping Move) even *Lexical Functional Grammar*.

§41 *Animal Attitudes (III)*

Language less creatures do not have sentential mental content, no *propositional* attitudes. It has been claimed: "It does not seem difficult to imagine languageless creatures who have propositional attitudes." (Jamieson 2009: 21). It does seem *quite* difficult as there is no directly accessible propositional content (like sentences of inner speech) to have such attitudes to. We can take the intentional stance, but this stance is taken instrumentally, as it might be on occasion towards all kinds of technical devices, in which case there might be at least the attitudes of the designers operational in the device, and towards inanimate things (say, "The ball wanted to roll leftwards, but the wall did not let it"). Even if the perceptual content of a sentient animal's mind presents objects with properties, this does not entail that the animal's sentient experience at any time processes propositional content with singular terms referring to objects and general terms referring to properties.

To ascribe phenomenal content to a grown mammal like a cat one, therefore, should not use sentential constructions like "The cat sees that the mouse is too far from its hiding place" but *transparent objectual constructions* like "The cat sees a mouse too far from its hiding place". Such attributions target BL and DL states, which allow to explain the cat's behaviour below the level of attribution of beliefs and desires. Being clear about the exclusively instrumental use of an intentional stance which interprets cat behaviour with externally constructed beliefs and desires, the intentional stance can be adopted. The danger of taking it succeeding at face value lurks here. The *caveat* that within the cat there are no sentential (and thus intensional) beliefs and desires must be kept in mind.

The justification for transparent objectual attribution of content stems from the hypothesis that the similarity between cat perceptual LOT and human perceptual LOT_2 may suffice for cats and humans (given their similar visual apparatus) to have similar *percepts*. Perceptual cognition in humans contains a connection between perceptions and perceptual beliefs, and the perceptions are conscious. The percepts can, however, be isolated as a dependent part of human perceptions. A part that may have a similar correlate in animal perception. Humans and animals can in their interactions focus and make use of the perceptual features of an object (say, a dog fetching fitting toys when shown a colour card or hearing a colour

command). The associative learning in the animal need not be associated with an experience, but the direction of attention and situational appropriate behaviour can be assumed to have phenomenality. It cannot be habitual in a novel situation, the attention *locus* not being part of an inbuild program. Of course, humans are a most suitable environment for an animal's state to work out the way the animal quasi-desired. We quickly infer to the best explanation of that stare, employing the intentional stance. By this animals can develop the habit to stare by associative learning. No TOM needs to be involved here, but nonetheless the animal has to have a schematic representation of humans as live agents that are useful given their presence in a situation with a salient focal object related to DL states.

The concepts employed with respect to the percept by a mammal and a person will differ widely, as is obvious with respect to artefacts: language less creatures cannot have concepts like 'book' or 'pencil', not to mention all theoretical concepts like 'electricity' or 'group cohesion'.

Domestic animals and pets show behaviour where their quasi-desire contains the quasi-desire that the human present does something about the object they have in focus. This need not be interpreted as involving 2nd order intentionality, but it shows that animals can accustom to the experience of joint focus. As a mechanic explanation of this behaviour is neither available nor on the horizon (of animal neuroscience) nor even likely given its versatility, the best explanation refers to intelligent situational behaviour given percepts, i.e. sentience.

§42 *Goal/Registration-Explanations*

In simple technical systems like a thermostat or trip wire alarm one finds conditional connections between different states of the system. State x being the case is followed (by some likelihood p) by state y. Nowhere in these simple systems one finds a conditional representing the conditional connection. The conditional connection is purely mechanic. Somewhat more sophisticated technical systems like remote controls or calculators are in parts programmed using a programming language with conditional branching. After compiling the code nowhere in the control unit of the system, if there is any, is a conditional statement accessible which expresses the operational conditional connection.

The conditional branching may work by using measures on some registered property, which again can be coupled with motor output related to goal states. In this way preference orderings and situationally preferred behaviour can be implemented in the system, again with no accessible numerical ordering as statement or number value directly accessed by a control unit.

Negation can be operationally present as absence of particular information. On a slightly more complex level the system represents conjunction, negation, and *simple* conditionals (indicative, not counterfactual, not universally quantified lawlike conditionals). Such a system corresponds to some animals (e.g. mammals, or even the intelligent bees as conceptualized by [J. Bennett 1964]).

An explanation based on goals and registrations of information appeals to:

(i) ascription/identification of goals and their preference ordering
(ii) ascription/identification of registrations of situational information
(iii) available motor output (where the system likely registers which output *o* is feasible in a registered situation *s*)
(iv) platitudes of the form: if *x* has motor output option *z*, and *x* registers {*y*, if *z* then *w*, if *y* then *z* feasible}, and has goal *w*, and there is no preferred goal *w** such that *x* registers {if *z* and *y* then not *w**} then *x* outputs *z*.

This type of explanation scheme needs more fine-tuning, but the idea of such Goal/Registration-Explanations (GRE) should be clear enough. They do not involve ascription of propositional attitudes. They can serve as lower level of an intentional stance towards technical systems – and animals.

Example: Why does the squirrel collect the nut from the yard? Explanation: the squirrel is able to run to the nut collecting it; the squirrel has the goal to have a nut (indeed many of them, as available); the squirrel registers nut collection is feasible in the given situation; the squirrel registers no preferred goal (like avoiding the farm cat) is compromised by collecting the nut.

This needs more fine-tuning as well. The amount of plasticity in an animal's behaviour correlates with the complexity of the GRE needed for its behaviour. The viability of such GREs, however, especially for animals lower in the tree of life, seems probable.

The next level are systems which operate, *inter alia*, on conditional representations of different types (strict conditionals, defaults etc.), having a broad spectrum of representations (including access to their preference orderings) and extended capabilities of behaving (like collecting relevant information). Such systems Newell (Newell 1990) conceptualized as '*General Problem Solvers*' (GPS). AI aspires to implement GPS, 'symbolic' AI allowing for internal symbolic representations as well. The GPS may have more resources for intelligent behaviour – 'borrowed', of course, from the intelligent designers – than many animals.

In computing machines like PRAMs/BTMs sentience is absent. This means for CS that the sub-theories on animal minds (i) concerned with GREs on animal behaviour and (ii) concerned with sentience, will *overlap at some crucial juncture* or point of elaboration, but are *not* identical and can be kept apart for some animals.

§43 *Animal Concepts (II)*

A theory of sentience cannot start top-down from our conception of consciousness. Consciousness has unique aspects, looking for them in the animal mind has a clear answer: they are missing. Animals speak no language, them talking about us behind our back is nice nonsense. They lack a self-narrative as they lack the language to narrate it in. They most likely lack *propositional* attitudes. They lack reflexion, because they lack an inner medium accessible to higher order thought (like inner speech), and they lack TOM. Had they TOM and recursive higher order thought they would have invented language, which they have not.

A theory of sentience has to invent conceptual tools to conceive of sentience, therefore proceed bottom-up, *conceptually bottom-up, not evolutionary*. The animals very low in the tree of life are too far off from cognition at all. No bacteria cognitive concepts pave the way to understand sentience. Evolutionary the theory of sentience has to start with animals close to us: domestic animals. Once a conception of their sentience has been developed, we then can look for *its ingredients* in other animals lower in the tree of life.

First conceptual tools are the concepts of BL and DL states and the hypothesis that mammal LOT has to have some similarity with human module LOT (LOT_2),

including representational structures which carry phenomenality in perception. Mammal LOT-types are quasi-concepts (in comparison to human concepts) as schematic representations (say of entity types like trees, foxes etc.).

The next step in conceiving a mammal's sentience is to identify such types. They will include:

(i) sensory types related to the animal's sensory apparatus, BL states and salient environmental features given the DL states of the animal,
(ii) desire types operative in the DL states, which express at least the basic quasi-desires of a mammal of this kind,
(iii) internal sensory types operative in control and self-monitoring of the animal,
(iv) acquired types in learning to deal with the contingent situation a group or even an individual animal is in.

In case of a domestic cow we may surmise:

(ad i) types like 'wet', 'loud', 'light', colours and many items similar to an early empiricist's list of 'impressions'.

(ad ii) types like 'fed' or 'safe', where types related to enjoying play or other activities are harder to identify.

(ad iii) types like 'stable' (in footing), 'hungry' and pain types.

(ad iv) types which relate to water bowls or a milking machine.

The individuation and collection of a list of such types could be achieved by well-meaning and open-minded cow experts like farmers or welfare scientists.

Given these types and behaviour explanations using objectual transparent descriptions of the situational mental content of the cow in BL and DL state quasi-psychology, we may try to understand – as far as possible for us – the life of a cow roughly from its perspective. Such a heterophenomenological description is – in principle – feasible. Of course, we are not cows. We do not have the experience of the cow as of its perspective. This, however, does not exclude an objective description of the cow's sentient states and sentient content, just like – in this case a real parallel to humans – the fact that each person has a 1st person perspective does not exclude the possibility and semantic comprehensiveness of a 3rd person

objective description of her situation, including her mental states and mental state content.[39]

Domestic animals appear as the best candidates for such systematic heterophenomenological accounts of their mental life. With them also issues of animal welfare are connected to knowledge of their mental life. We tend to slide into anthropomorphism, but scientific observers both of domestic animals as well as animals in the wild can methodologically suppress this tendency.

Assuming such conception building works for some mammals, CS may proceed:

(i) trying to do something resembling this to animals resembling the animals dealt with so far,

(ii) trying to work in a similar fashion like human neuroscience in establishing neural correlates of the sentient states, although, since interactive participation of a test 'subject' is limited here, this will be even more difficult than in the human case.

Given reasonable confidence in results delivered by (ii) CS can look for the presence of the neural correlates or a highly similar structure in other species or taxa. This has already been done by researchers in cognitive ethology, neuroscience and CS in general.

Critics can point out that the descriptions often used are semantically overshooting, they extend psychological vocabulary instead of regimenting a vocabulary for animal minds [like this book complains as well]. Critics can also point out that the extension of ascribing sentience to some other species or even taxa moves far too quickly [like this book complains as well]. The methodology hinted at does not secure but rather undermines that animals apart from mammals and some, if not all, birds are sentient (in a way that can be conceived by this methodology). Somewhere the line sentient/non-sentient has to be drawn, rejecting panpsychism and mysticism. Given the paradigmatic arguments considered here

[39] Philosophers who deny this (e.g. Nagel 2012) confuse the *epistemic* irreducibility of the 1st person perspective with the *semantic* incomprehensiveness of a 3rd person objective description of mental life. I agree with Dennett on heterophenomenology, and with Weisberg (2008: 143): "It's all there, just accessed by a different route. So the physicalist story is complete, even though we can't imagine what it is like to be [the cat] Poopy."

against widespread sentience, the line may be drawn below birds. There are difficult cases like cephalopods (cf. Godfrey-Smith 2016) where one may justify at least a thorough look at the animal's intelligent behaviour, although their short life span, many offspring and cold-bloodedness make sentience less likely. The possibility that sentience evolved independently with cephalopod brains should not be excluded *a priori*.

§44 *Conditions of Sentience*

BL states and DL states are like beliefs and desires in their functional role. Animals with BL and DL states fall under explanations that are like those in belief and desire psychology. BL and DL states differ from beliefs and desires in not using sentential content. In the proper sense even sentient animals do not have *propositional* attitudes, nor thoughts.

A theory of sentience contains criteria for scribing sentience. Form a realist pint of view (ideally) the conditions for ascribing a property correspond to conditions for having that property. Thus, the debate of the last few paragraphs leads to a (defeasible) definition.

Sentient animals are exactly those which:

(i) can be interpreted in their behaviour by employing the intentional stance using beliefs and desires the content of which is specified from the outside by the human interpreter;
(ii) themselves have BL and DL states, which cannot be directly specified from the outside, but be interpreted in GREs;
(iii) have to be interpreted from an intentional stance, since their behaviour has no predictively and explanatory equivalent non-intentional explanation.

Condition (ii) combines the platitude that we cannot access the mind of an animal with the viability of a heterophenomenology which goes beyond covering animal mental states by vocabulary of human psychology. Of course, any animal can be interpreted from the intentional stance, as can many technical devices. Sentient animals need to be interpreted thus because of their quasi-intentional states,

which effect their behaviour. Sentient animals exhibit behaviour of a complexity and versatility (situational plasticity) that defies reduction to conditional reflex like sensory-motor-programs.

The evolutionary rationale of sentience lays in allowing for such complexity by compensating with situational adaptation by updating BL and DL states. The investment into the required neural foundations pays off for species with few longer-lived offspring, which may depend on longer periods of rearing and learning.

Sentient animals have phenomenality (by definition of 'sentience'). Animals with phenomenal states and learning behaviour can also relive parts of their experiences or like situations in dreams. This lived experience involves a perspective and phenomenal states representing internal (bodily) states, but not a self or subject. Sentient life lives in the immediacy of its lived experience. Sentient animals distinguish not only themselves from the environment, but can also distinguish members of their species or flock from other animals. Their inner representations contain socially directed BL and DL states, although they lack TOM. They see their conspecifics as causal agents – interfering in their own situation – but not as intentional agents with a mental life.

Some sentient animals (like primates or ravens) can detach partially from the immediacy of the situation by planning, which involves phenomenal states of imagination. Their self-directed BL states involve a self-representation that can be understood from the outside as a precursor of self-knowledge.

These and all other more versatile sentient animals have no concept of truth and possess no natural language, but they communicate the content of their BL and DL states to human interactors in learned ways, and to their conspecifics in natural ways of signalling. Vocalizations of sentient animals can be seen as expression of externally inaccessible and difficult to conceive quasi-attitudinal states and feelings.

This conception of sentience follows a bulk of empirical evidence from cognitive ethology (and CS in general) in substance, but not in terminology. Condition (iii) is the centre of controversy about fishes and invertebrates, apart from Cephalopods, which seem to be too intelligent and versatile for their behaviour to be

explained any other way than by GREs. Concerning other invertebrates and fishes the debate in CS shifts burdens of proof back and forth. Advocates of sentience stress that some behaviour seems irreducible quasi-intentional and that no *complete* reductive alternative has been provided. Deniers of sentience stress that is only *seems* quasi-intentional, and that although complete reductions or simulations are difficult to achieve and corroborate, algorithmic or mechanistic models have often enough been indicated to be confident that advanced CS may deliver the details. Given the general burden of proof resting with the *proponents* of some existant (object or property) condition (iii) does not lead to inclusion of those invertebrates and fishes into the class of sentient animals.

This conception of sentient animals suffices to systematize and articulate the idea of sentience contained in the certainty of our manifest image of the world that at least grown mammals and some birds are sentient. As sentience also implies feelings and states similar enough to human pain or anxiety, the conception also suffices as a foundation of animal welfare considerations.

Epilogue

What this book has not provided is a new comprehensive theory of sentience as theory building in CS. That should be pursued mostly by empirical research in CS, most notably cognitive ethology as field research of animals in the wild, supplemented then by comparative neuroscience of animal brains. In principle behaviour centred approaches (like in field research) should dominate over neurophysiology, and be counterbalanced by considerations of parsimony in the light of simulation by non-sentient computational equivalents (in AI, robotics or ML).

What this book has tried to provide was a meta-reflection on arguments concerning sentience and descriptions of animal minds in CS. Individual papers and books on animal minds often rely only on a very limited range of paradigmatic arguments in favour of their theses. Further on, an enrichment of methodology and descriptive resources when dealing with mental content of supposedly sentient animals seems to be called for. With respect to this the book engaged in some limited theory building.

What this book pleaded for was a more balanced consideration of paradigmatic arguments on a way towards reflective equilibrium between them and their background covering animal minds.

What this book hopes for – presumably in vain – is the introduction of more regimented language when theorizing about animal minds. Unbridled anthropomorphism and use of human psychological vocabulary applied to animals (including even fishes or invertebrates) undermines the credibility of theories of animal cognition. Moral motives to liken animal minds to human minds may be understandable and laudable, but, because of their obvious shortcomings in the eyes of critics, endanger the better moral consideration of sentient animals. The slogan "putting animals in their place", although negatively connotated, could mean putting them in the proper place of distinctive descriptions and cognitive models in

contrast to persons (i.e. conscious beings), which by itself need not exclude them from moral considerations. Popular science infotainment books which claim that flies "have personalities" or ants "have a sense of their mortality", and much more (e.g. Morell 2013) are beyond repair. Criticism of them had from a social science perspective to look at the economics and social logic of infotainment, as well as from a psychological or philosophical perspective at the human desire to find companionship in animals, not to be alone in face of the rest of nature. These are interesting and anthropologically fascinating topics, which will not change any anthropomorphizing habits, of course. Outside of science there is – up to ethical issues – nothing wrong with anthropomorphizing in cartoons, memes, and our treatment of companion animals. Obviously, lots of fun can be found here, fun mostly aiming at humans using animals as props. Companion animals understood as 'personalities' have educational and therapeutic importance. Inside of CS, however, terminology and theory should not be unbridled anthropomorphic (i.e. only as so far unavoidable), and should – especially after the moral consideration of animals has been admitted – untied from ethical issues and human psychological fascination with animals and our companionship with them. Dennett's 'killjoy' lower-level explanations of supposedly advanced cognitive capacities have their proper place here (cf. Shettleworth 2010).

Bibliography

Alcock, John (2001). *The Triumph of Sociobiology*. Oxford.

Allen, Collin (1999). "Animal Concepts Revisited", *Erkenntnis*, 51.

Andrews, Kristin/Beck, Jacob (2019). (Eds.) *The Routledge Handbook of Philosophy of Animal Minds*. New York.

Baars, Bernard (1997). *In the Theater of Consciousness*. New York.

- (2021). *On Consciousness.* Science and Subjectivity. New York.

Balcombe, Jonathan (2016). *What a Fish Knows.* The Inner Lives of Our Underwater Cousins. New York.

Barret, Louise (2011). *Beyond the Brain*. How Body and Environment Shape Animal and Human and Minds. Princeton.

Barwise, Jon/Perry, John (1983). *Situations and Attitudes*. Cambridge/MA.

Bayne, Tim/Hohwy, Jakob (2013). "Consciousness: Theoretical Approaches", in: Cavanna, A. et al. (Eds.) *Neuroimagining of Consciousness*. Berlin.

Beauchamp, Tom/Frey R.G. (2011). (Eds.) *The Oxford Handbook of Animal Ethics*. Oxford.

Bechtel, William/Abrahamson, Adele (1991). *Connectionism and the Mind*. An Introduction to Parallel Processing in Networks. Cambridge/MA.

Bekoff, Marc/Allen, Collin. (1997). *Species of Mind*. Cambridge/MA.

- /Burghardt, Gordon (2002). (Eds.) *The Cognitive Animal*. Empirical and Theoretical Perspectives on Animal Cognition. Cambridge/MA.

Bekoff, Marc/Meaney, Carron (1998). *Encyclopaedia of Animal Rights and Animal Welfare*. Westport

Bekoff, Marc/Jamieson, Dale (1996). (Eds.) *Readings in Animal Cognition*. Cambridge/MA.

Bennett, Jonathan (1964). *Rationality*. Towards an Analysis. London.

- (1976). *Linguistic Behaviour.* Cambridge.

Bennett, Max (2023). *A Brief History of Intelligence*. London.

Bermudez, Jose Luis (2009). "Mindreading in the animal kingdom", in: (Lurz 2009).

Birch, Jonathan (2022). "The search for invertebrate consciousness", *Nous*, 56.

Boden, Margaret (1996). (Ed.) *The Philosophy of Artificial Life*. Oxford.

Bremer, Manuel (2005). "Tierisches Bewusstsein als Testfall für die Kognitionswissenschaften", in: Hermann, Christoph et al. (Eds.) *Bewusstsein*. München.

- (2005a). "Pro Pain. On The Role of the Phenomenon of Pain in Studying Animal Minds", in: Weingartner, Paul (Ed.) *Das Problem des Übels.* Frankfurt a.M.
- (2006). „Tierisches Bewusstsein, Anthropomorphismus und Heterophänomenologie", *Philosophisches Jahrbuch*.
- (2006a). "The Egological Structures of Consciousness. Lesson from Sartre for the Analytic Philosophy of Mind", in: Dreyfus, H./Wrathall, W. (Eds.) *Blackwell Companion to Phenomenology and Existentialism. Oxford.*
- (2007). „Methodologische Überlegungen zu tierischen Überzeugungen", *Journal for General Philosophy of Science*.
- (2008). *Conceptual Atomism and Justificationist Semantics*. Frankfurt a.M.
- (2008a). "Do Cats Have Beliefs?", in (Hales 2008).
- (2012). "How are Semantic Metarepresentations Built and Processed?", *Kriterion*.

- (2025). *Against Metaphysical Necessity*. Modality and Belief Revision in Updates Logical Empiricism. Berlin.
- /Cohnitz, Daniel (2004). *Information and Information Flow*. An Introduction. Frankfurt a.M.

Brooks, Rodney (2002). *Robot*. The Future of Flesh and Machine. London.

Broom, Donald (2014). *Sentience and Animal Welfare*. Wallingford.

Budiansky, Stephen. (1992), *The Covenant of the Wild*. Why Animals Chose Domestication. New York.
- (1998). *If a Lion Could Talk*. London.

Butterfield, Stephen (2019). "Tracking others' mental states", in: (Andrews/Beck 2019).

Camp, Elisabeth (2009). "A language of baboon thought?", in: (Lurz 2009).

Carruthers, Peter (1994). *The Animal Issue*. Cambridge/MA, 2nd Editio
- (1996). *Language, Thought and Consciousness*. Cambridge.
- (2000). "The evolution of consciousness", in: (Carruthers/Chamberlain 2000).
- (2005). *Consciousness*. Essays from a Higher-Order Perspective. Oxford.
- (2006). *The Architecture of the Mind*. Oxford.
- (2018). "The Problem of Animal Consciousness", *Proceedings and Addresses of the APA*, Vol. 92.
- /Boucher, Jill (1998). (Eds.) *Language and Thought.* Interdisciplinary Themes. Cambridge.
- /Chamberlain, Andrew (2000). (Eds.) *Evolution and the human mind.* Modularity, language and meta-cognition. Cambridge.

Castaneda, Hector-Neri (1999). *The Phenomeno-Logic of the I*. Essays on Self-Consciousness. Indiana.

Chalmers, David (1995). "Absent Qualia, Fading Qualia, Dancing Qualia", in: Metzinger, Thomas (Ed.) *Conscious Experience*. Paderborn.
- (1996). *The Conscious Mind*. In Search of a Fundamental Theory. Oxford.

Chisholm, Roderick (1981). *The First Person*. An Essay on Reference and Intentionality. Minneapolis.

Chittka, Lars (2022). *The Mind of a Bee*. Princeton.

Chomsky, Noam (1966). *Cartesian Linguistics*. A Chapter in the History of Rationalist Thought. New York.
- (1980). *Rules and Representations.* New York.
- (1995). *The Minimalist Program*. Cambridge/MA.
- (2000). *New Horizons in the Study of Language and Mind*. Cambridge.
- (2016). *What Kind of Creatures Are We?* New York.
- /Berwick, Robert (2016). *Why Only Us*. Language and Evolution. Cambridge/MA.

Chris, Eileen (1999). *Images of Animals*. Anthropomorphism and Animal Mind. Philadelphia.

Churchland, Patricia (1998). *Neurophilosophy*. Cambridge/MA, 10th Edition.
- (2013). *Touching a Nerve*. The Self as Brain. New York.
- /Sejnowski, Terrence (1992). *The Computational Brain*. Cambridge/MA.

Clark, Andy (2016). *Surfing Uncertainty.* Prediction, Action, and the Embodied Mind. Oxford.
- (2023). *The Experience Machine*. How Our Minds Predict and Shape Reality. London.

Corballis, Michael (2011). *The Recursive Mind.* The Origins of Human Language, Thought, and Civilization. Princeton.

Daston, Lorraine/Mitman, Gregg (2005). (Eds.) *Thinking with Animals.* New Perspectives on Anthropomorphism. New York.
Davidson, Donald (1982). "Rational Animals", *Dialectica*, 36.
- (1984). *Inquiries into Truth and Interpretation*. London.
- (1999). "The Emergence of Thought", *Erkenntnis*, 51.
- (2001). *Subjective, Intersubjective, Objective*. Oxford.

Darwin, Charles (1871). *The Descent of Man*. London.
Dawkins, Marian (1980). *Animal Suffering*. London.
- (1993). *Through Our Eyes Only?* The Search for Animal Consciousness. Oxford.

Deacon, Terrence (1997). *The Symbolic Species*. The Co-evolution of Language and the Brain. New York.
DeGrazia, David (1996). *Taking Animals Seriously*. Cambridge/MA.
Dennett, Daniel (1971). "Intentional Systems", *Journal of Philosophy*, 68.
- (1991). *Consciousness Explained*. London.
- (1995). *Darwin`s Dangerous Idea*. London.
- (2005). *Sweet Dreams*. Philosophical Obstacles to a Science of Consciousness. Cambridge/MA.
- (2017). *From Bacteria to Bach and Back.* The Evolution of Minds. London.

Descartes, Rene (1641). *Meditationes de Prima Philosophia, in qua Dei existentia et animæ immortalitas demonstratur*. Paris.
- (1644). *Principia Philosophiae*. Amsterdam.
- (1970). *Philosophical Letters*. Edited and translated by Antony Kenny. Oxford.

Dobrovolsky, Michael (1989). "Animal Communication", in: O`Grady, William et. al. *Contemporary Linguistics*. New York.
Dretske, Frank (1988). *Explaining Behaviour.* Reasons in a World of Causes. Cambridge/MA.
- (1999). *Knowledge and the Flow of Information*. Cambridge/MA.

Dreyfus, Hubert (2014). *Skillful Coping*. Essays on the Phenomenology of Everyday Perception and Action. Edited by Mark Wrathall. Oxford.
Dubner, Ronald (1984). "Methods of Assessing Pain in Animals", in: Wall, P./Melzack, R. (1994). (Eds.) *Textbook of Pain*. 3rd Edition, Edinburgh et. al
- Beitel, R./Brown, F. (1976). "A Behavioral Animal Model for the Study of Pain Mechanisms in Primates", in: Weisenberg, M./Tursky, B. (Eds.) *Pain*. New Perspectives in Theraphy and Research. New York/London.

Dunbar, Robin (2000). "On the origin of the human mind", in: (Carruthers/Chamberlain 2000).
Dunlosky, John/Metcalfe, Janet (2009). *Metacognition*. Thousand Oaks.
Edwards, Paul (2002). *Reincarnation*. A Criticial Examination. New York.
Feinberg, Todd (2024). *From Sensing to Sentience*. How Feeling Emerges from the Brain. Cambridge/MA.
Fichte, Johann Gottlieb (1794). *Grundlage der gesamten Wissenschaftslehre*. Leipzig.
Field, H.L./Price. D.D. (1994). "Pain", in: Guttenplan, Samuel (Ed.) *A Companion to the Philosophy of Mind*. Oxford.
Fitzpatrick, Simon (2019). "Against Morgan's Canon", in: (Andrews/Beck 2019).
Fodor, Jerry (1974). "Special Sciences (or the Disunity of Science as a Working Hypothesis)", *Synthese*, 28.
- (1975). *The Language of Thought*. Cambridge/MA.
- (1983). *The Modularity of Mind*. Cambridge/MA.
- (1987). *Psychosemantics.* Cambridge/MA., 2nd Edition.

- (1998). *Concepts*. Where Cognitive Science Went Wrong. Oxford.
- (1998a). *In Critical Condition*. Polemical Essays on Cognitive Science and the Philosophy of Mind, Cambridge/MA.
- (2000). *The Mind Doesn't Work That Way*. The Scope and Limits of Computational Psychology. Cambridge/MA.
- (2008). *LOT 2*. Oxford.

Fox, Michael (1986). *The Case for Animal Experimentation*. An Evolutionary and Ethical Perspective. Berkeley.

Franklin, Julian (2005). *Animal Rights and Moral Philosophy*. New York.

Gallistel, C. R./King, Adam (2010). *Memory and the Computational Brain*. Why Cognitive Science Will Transform Neuroscience. Malden.

Gazzaniga, Michael/Ivry, Richard/Mangun, George (2019). *Cognitive Neuroscience*. The Biology of the Mind. 5th Edition. New York.

Godfrey-Smith, Peter (2016). *Other Minds*. The Octopus and the Evolution of Intelligent Life. London.

Gould, James (1982). *Ethology*. The Mechanisms and Evolution of Behavior. New York.

Green, David et. al. (1996). *Cognitive Science*. An Introduction. London.

Greenwood, John (1991). (Ed.) *The Future of Folk Psychology*. Intentionality and Cognitive Science. Cambridge.

Griffin, Donald (1992). *Animal Minds*. Chicago/London.

Gutfreund, Yoram (2024). "Neuroscience of animal consciousness: still agnostic after all", *Frontiers in Psychology*.

Habermas, Jürgen (1995). *Justification and Application*. Remarks on Discourse Ethics. Cambridge/MA.

Hales, Steven (2008). (Ed.) *What Philosophy Can Tell You About Your Cat*. Cicago.

Hasker, William (1999). *The Emergent Self*. Ithaca.

Hauser, Marc (2000). *Wild Minds*. What Animals Really Think. Markham.

Hawking, Stephen/Mlodinow, Leonard (2011). *The Grand Design*. New Answers to the Ultimate Questions of Life. London.

Hornstein, Norbert (2009). *A Theory of Syntax*. Minimal Operations and Universal Grammar. Cambridge.

Humphrey, Nicholas (1992). *A History of the Mind*. Evolution and the Birth of Consciousness. New York.

- (2022). *Sentience*. The Invention of Consciousness. Oxford.

Husserl, Edmund (1928). *Vorlesungen zur Phänomenologie des inneren Zeitbewußtseins*. *Jahrbuch für Philosophie und phänomenologische Forschung*. Volume IX.

- (1950). *Cartesianische Meditationen*. The Hague.

Irwin, Louis et al. (2022). "Editorial: Comparative animal consciousness", *Frontiers in Systems Neuroscience*.

Jackendoff, Ray (2002). *Foundations of Language*. Brain, Meaning, Grammar, Evolution. New York.

James, William (1890). *The Principles of Psychology*. Harvard.

Jamieson, Dale (2009). "What do animals think?", in: (Lurz 2009).

Kant, Immanuel (1786). *Kritik der reinen Vernunft*. 2nd Edition. Königsberg.

Kelleher, John (2019). *Deep Learning*. Cambridge/MA.

Kennedy, John (1992). *The New Anthropomorphism*. Cambridge.

Key, Brian (2016). "Why fish do not feel pain", *Animal Sentience*.

Knight, Andrew (2025). “Why the Brain Cannot Be a Digital Computer: History-Dependence and the Computational Limits of Consciousness”, *History and Philosophy of Physics*.
Kowalski, Robert (2011). *Computational Logic and Human Thinking*. Cambridge.
Krigel, Uriah/Williford, Kenneth (2006). (Eds.) *Self-Representational Approaches to Consciousness*. Cambridge/MA.
Krohn, Thomas (2002). *Method Development and Assessments of Animal Welfare in IVC-Systems*. Dissertation Universität Frederiksberg (Denmark).
Leahy, Michael (1991). *Against Liberation*. Putting Animals in Perspective. London.
Levesque, Hector (1989). “Logic and the Complexity of Reasoning”, in: Thomason, Richmond (Ed.). *Philosophical Logic and Artificial Intelligence*. Dordrecht.
- (2012). *Thinking as Computation*. Cambridge/MA.
Levy, Steven (1992). *Artificial Life*. The Quest for a New Creation. New York.
Lewis, David (1969). *Convention*. Harvard.
- (1980). “Mad pain and Martian pain”, in: Block, Ned (Ed.) *Readings in Philosophy of Psychology*. Vol.1, Cambridge/MA.
Libet, Benjamin (2002). *Mind Time*. The Temporal Factor in Consciousness. Cambridge/MA.
Lipton, Peter (1991). *Inference to the Best Explanation*. London.
Locke, John (1690). *An Essay Concerning Human Understanding*. London.
Lurz, Robert (2009). (Ed.) *The Philosophy of Animal Minds*. Cambridge.
Lycan, William (1996). *Consciousness and Experience*. Cambridge/MA.
Macphail, Euan (1998). *The Evolution of Consciousness*. Oxford.
Maloney, Christopher (1989). *The Mundane Matter of the Mental Language*. Cambridge.
Marcus, Gary (2001). *The Algebraic Mind*. Integrating Connectionism and Cognitive Science. Cambridge/MA.
Martinho-Truswell, Antone (2022). *The Parrot in the Mirror.* How Evolving to be Like Birds Made Us Human. Oxford.
Mayr, Ernst (2001). *What Evolution Is*. New York.
McDermott, Drew (2001). *Mind and Mechanism*. Cambridge/MA.
Mikhalevich, Irina/Powell, Russell (2020). “Minds without spines: Evolutionary inclusive animal ethics”, *Animal Sentience.*
Miles, Lyn White (1993). “Language and the Orang-Utan”, in: Cavalieri, Paola/Singer, Peter (Eds.) *The Great Ape Project*. London.
Montague, Richard (1974). *Formal Philosophy*. Selected Papers of Richard Montague. Edited by Richmond Thomason. New Haven.
Morell, Virginia (2013). *Animal Wise*. The Thoughts and Emotions of our Fellow Creatures. Brecon.
Nagel, Thomas (2012). *Mind and Cosmos*. Why the Materialist Neo-Darwinian Conception of Nature is Almost Certainly False. Oxford.
Narayanan, Arvind/Kapoor, Sayash (2024). *AI Snake Oil*. What Artificial Intelligence Can Do, What It Can’t, and How to Tell the Difference. Princeton.
Newell, Allen (1990). *Unified Theories of Cognition*. Harvard.
Natorp, Paul (1888). *Einleitung in die Psychologie nach kritischer Methode*. Freiburg.
Olliver, Alex (1996). “The Metaphysics of Properties”, *Mind*, 105.
Parker, Sue et al. (1994). (Ed.). *Self-awareness in Animals and Humans*. Cambridge.
Pena-Guzman, David (2022). *When Animals Dream*. The Hidden World of Animal Consciousness. Princeton.

Petzold, Charles (2000). *Code*. The Hidden Language of Computer Hardware and Software. Redmond.
Plotkin, Henry (1994). *Darwin Machines and the Nature of Knowledge*. London.
Premack, D. (1976). *Intelligence in Ape and Man*. New York et al.
- /Premack, A. (1983). *The Mind of an Ape*. New York.
Pylyshyn, Zenon (1984). *Computation and Cognition*. Toward a Foundation for Cognitive Science. Cambridge/MA.
Radner, Daisie (1994). "Heterophenomenology. Learning About the Birds and the Bees", *Journal of Philosophy*, 91.
- /Radner, Michael (1996). *Animal Consciousness*. New York.
Raschid, Tarig (2016). *Make Your Own Neural Network*. Scotts Valley.
Reber, Arthur (2016). "Caterpillars, consciousness and the origins of mind", *Animal Sentience*.
Regan, Tom (1983). *The Case for Animal Rights*. Berkeley.
- (2004). *Empty Cages.* Facing the Challenge of Animal Rights. Lanham.
Reinhold, Karl Leonard (1791). *Über das Fundament des philosophischen Wissens*. Jena.
Rey, Georges (1997). *Contemporary Philosophy of Mind*. A Contentiously Classical Approach. Malden.
Rogers, Lesley (1997). *Minds of Their Own*. Thinking and Awareness in Animals. Boulder.
Rollin, Bernard (1989). *The Unheeded Cry*. Animal Consciousness, Animal Pain and Science. Oxford.
- (1992). *Animal Rights and Human Morality.* Revised Edition. Amherst.
Romanes, George (1898). *Animal Intelligence*. London.
van Rooij, Iris et al. (2019). *Cognition and Intractability*. A Guide to Classical and Parameterized Complexity Analysis. Cambridge.
Rowlands, Mark (2002). *Animals Like Us*. London.
Sartre, Jean-Paul (1936). "La Transcendance de L'Ego", *Recherches Philosophiques*. English: *The Transcendence of the Ego*. An Existentialist Theory of Consciousness. New York, 1960.
- (1943). *L'Etre et le neant*. Paris. English: *Being and Nothingness.* London, 2003.
Shettleworth, Sara (2010). "Clever animals and killjoy explanations in comparative psychology", *Trends in Cognitive Science*, 14.
Shoemaker, Sydney (2007). *Physical Realization*. Oxford.
Short, Charles/Poznak, Alan Van. (1992). (Eds). *Animal Pain*. New York.
Schustermann, R. et al. (1986). *Dolphin Cognition and Behaviour*. Hillsdale/London.
Searle, John (1995). *The Construction of Social Reality*. London.
Sneddon, Lynne et al. (2018). "Fish sentience denial: Muddying the waters", *Animal Sentience*.
Sober, Elliot. (1975). *Simplicity*. Oxford.
- (1990). "Let´s Razor Ockham´s Razor", in: Knowles, Dudley (Hg.). *Explanation and its Limits*. Cambridge.
- (1993). *Philosophy of Biology*. Oxford.
- (1996). "Parsimony and Predictive Equivalence", *Erkenntnis*, 44.
Sperber, Dan (2000). (Ed.) *Metarepresentations*. A Multidisciplinary Perspective. Oxford.
Strawson, P. F. (1959). *Individuals*. An Essay in Descriptive Metaphysics. London.
- (1985). *Scepticism and Naturalism*. Columbia.
Sunstein, Cass/Nussbaum, Martha (2004). *Animal Rights*. Current Debates and New Directions. Oxford.
Swinburne, Richard (1986). *The Evolution of the Soul*. Oxford.

Tallerman, Maggie (2005). (Ed.) *Language Origins*. Perspectives on Evolution. Oxford.
Tattersall, Ian (2002). *The Monkey in the Mirror*. Essays on the Science of What Makes Us Human. San Diego.
Terman, Folke (1993). *Reflective Equilibrium*. An Essay in Moral Epistemology. Karlshamn.
Thagard, Paul. (1978). "The Best Explanation: Criteria for Theory Choice", *The Journal of Philosophy*, 75.
Tomasello, Michael (1999). *The Cultural Origins of Human Cognition*. Cambridge/MA.
Trestman, Michael/Allen, Colin/Birch, Jonathan (2025). „Animal Consciousness", *Stanford Encyclopaedia of Philosophy*.
Turing, Alan (2004). *The Essential Turing*. Seminal Writings in Computing, Logic, Philosophy, Artificial Intelligence, and Artificial Life. Edited by Jack Copeland. Oxford.
Veit, Walter (2022). "Towards a Comparative Study of Animal Consciousness", *Biological Theory*.
Vygotsky, Lev (1986). *Thought and Language*. Cambridge/MA.
de Waal, Frans (2001) *The Ape and the Sushi Master*. Cultural Reflections by a Primatologist. London.
Wall, Patrick (1992). "Defining 'Pain in Animals'", in: (Short/Poznak 1992).
Webb, Christine (2025). *The Arrogant Ape*. The Myth of Human Exceptionalism and Why It Matters. New York.
Webster, J. (1994). *Animal Welfare*. A Cool Eye towards Eden. Oxford.
Weisberg, Josh (2008). "What's It Like to Be My Cat?", in (Hayes 2008).
Wheeler, Michael (2005). *Reconstructing the Cognitive World*. The Next Step. Cambridge/MA.
Williams, Danielle (2025). "How to be realist about computational neuroscience", *Synthese*.
Wittgenstein, Ludwig (1953). *Philosophical Investigations*. Oxford.
- (1969). *On Certainty*. New York.
Young, Rosamund (2005). *The Secret Life of Cows*. Animal Sentience at work. Preston.
Zahavi, Dan (2020). *Self-Awareness and Alterity*. A Phenomenological Investigation. New Edition. Evanston.