Information, Intelligence and Idealism

Martin Korth

Contents

Chapter 1

Introduction

Why are computers so smart these days? And why are humans apparently still a bit smarter? Does this have something to do with the difference between data and meaning? Does this in turn mean that at least some abstract entities, such as numbers, exist independently of human thought? Wouldn't that require an expansion of our scientific world view? And would that at all be compatible with what we know about our world from physics and chemistry, philosophy, psychology, neuroscience and the theory of evolution? Finally, what would this tell us about ethical and aesthetic value theories?

These and related questions will be discussed in this book. We will find that the difference between data and meaning, i.e. quantitative and qualitative information, does indeed appear to be of central importance for understanding both artificial and natural intelligence. And then the independent existence of abstract entities not only appears to be a particularly promising hypothesis, but also one that is entirely compatible with the sum of our scientific knowledge, especially with regard to value theories. The book thus arrives at the exploration of a scientifically tenable, panpsychistically inspired, objective idealism that can be derived from our most fundamental intuitions as subjects that perceive qualities, but that can also take into account the structuring of the world already at the micro-scale, found in the modern natural sciences.

The result is a Platonic, but in a second step also a scientific realism and a naturalism in the sense that it is informed by the natural sciences in terms of an inductive metaphysics. An objective idealism, not in a rationalistic maximum form, but in a pragmatic minimum form; without eternal truths, but dependent on the continued philosophical-scientific and also philosophical-social dialog.

The proposed model could offer interesting solutions to a number of problems at and near the mind/matter boundary: Proposals are being considered for the interpretation of quantum mechanics, the problem of molecular symmetry, the neuronal code and the binding problem in neuroscience, mental causation, a more holistic understanding of mental processes, and so on and so forth. However, the extent to which the model threatens to promise far too much is also being discussed.

In sum, the core question is how we can imagine human thinking beyond physically conceived information processing. An alternative model of human thinking is then put up for discussion, for which not only machine-like cognitive performance, but above all the intentional perception of qualitative information, i.e. of abstract entities, would be central, as well as the free, ultimately creative linking of patterns of quantitative information (signals, data) with such qualities (meanings).

Chapter 2

Information and Meaning

The digital revolution has been underway for several decades, so it should come as no surprise that the philosophy of information seems to be consolidating as an independent field in philosophy. However, no consensus has yet emerged on the concrete shape of this field, which is certainly also due to an unresolved conflict surrounding the central concept of information. [1] Our understanding of information is fed by two seemingly irreconcilable streams of thought: On the one hand - and especially in the humanities - we speak of qualitative, especially semantic information in the sense of meaning, [2] on the other hand – especially in the natural sciences – of quantitative information in the sense of a certain amount of raw data and the reduction of uncertainty associated with its receipt. In both directions, a whole series of concepts have been developed to define information: In qualitative terms, especially by Bar-Hillel/Carnap [1] and more recently Floridi. [2] In quantitative terms, especially by Shannon, in the sense of the above-mentioned reduction of uncertainty through the reception of physical signals, [3] as well as analogous concepts in thermodynamics. More recent approaches that propagate an agent-based, algorithmic understanding of information can also be understood via the quantitative approach, but attempt to build a bridge to the qualitative view.

Information and its processing seem to have a universal character, which furthermore seems to allow an abstraction from concrete physical circum-

[1] The following brief overview initially follows roughly Adriaans' presentation in his article 'Information' in the Stanford Encyclopedia of Philosophy (as of 01.11.23).

[2] It should be noted here that in the following, 'meaning' is used in the linguistic context corresponding to Frege's 'sense', while 'reference' is used here as usual instead of Frege's 'meaning'. This is discussed in more detail in chapter 8. The linguistic-philosophical problem of reference is assumed to be of secondary importance for now; it is, however, then also taken up in Chapter 8.

stances, as we find it realized, for example, in the concept of the Turing machine. [4] However, on the semantic level, this always applies only to a given context in which certain conventions assign a material symbol its immaterial meaning. Since Frege [5], and especially in the course of the 'linguistic turn', attempts have been made to quantify semantic information as sets of true sentences, but as we shall see later, this probably falls short. For purely qualitative approaches, on the other hand, it is largely unclear how semantic information can be derived from data at all. This 'symbol grounding problem' [6] of how signs obtain their meaning can be understood as a generalization or, alternatively, as a special case of arguments such as Searl's 'chinese room' [7] or 'Leibniz's mill' [8], namely that a human mind understood as a physical symbol system and its cognitive processes understood as physical symbol manipulation do not allow to build a bridge to meanings.

2.1 Floridi's Philosophy of Information

The above, fundamentally unresolved conflict does not, of course, prevent the digital revolution from progressing, except that the philosophies of information – which are only just emerging in this process – have to come to terms with it in some way. The persistence of two cultures in the sciences makes adaptation easier: The need to think of both concepts of information together simply does not exist in many areas. This is made use of by a school of thought in the philosophy of information that compartmentalizes its problems and wants to treat them on separate levels. Floridi's design of a philosophy of information can be seen as the spearhead of this development.

This should not be understood as a criticism of his intellectual honesty; Floridi does not ignore the symbol grounding problem and has not left it behind lightly, as a whole series of high-ranking publications on this topic and his own proposed solutions show. [9] Rather, his book 'Philosophy of Information' [10] is part of a larger project to explore the role of the concept of information with all its facets for our understanding of the world, especially also with regard to ethical questions. Consequentially, this was followed by two further books on ethics [11] and finally on the logic of information [12], in the latter of which he attempts to restructure the whole of philosophy starting from the concept of information: After Aristotelian metaphysics, Kantian epistemology and the 'linguistic turn' towards a theory of meaning based on logic, Floridi proposes not to follow the contemporary impulse towards a new metaphysics, but to abandon all kinds of representation in favor of 'conceptual design'. [13] The extent to which the associated relativism can provide helpful answers cannot be examined in detail here; however, the view presented in the following calls this into question, at least in parts.

With regard to every construction of a new field of research, it will inevitably be asked to what extent the project is a sensible refocusing or rather a career-enhancing PR stunt (it can also be both). In any case, Floridi provides very good arguments for a successful refocusing. First, he defines:

'The philosophy of information (PI) is the philosophical field concerned with (a) the critical investigation of the conceptual nature and basic principles of information, including its dynamics, utilization, and sciences; and (b) the elaboration and application of information-theoretic and computational methodologies to philosophical problems.' [10](S.14)

He then systematically demonstrates the content and methods of his philosophy of information. However, with his approach of examining the concept of information on different, largely independent levels of abstraction, he institutionalizes the dichotomous understanding of information in today's scientific community described above as a central conflict. And through his qualitative definition of information as well-formed, meaningful and true data, he avoids a number of further problems – including that of the information content of false information, which is now simply no longer taken to be information at all – but does not solve them either. (In a way, Floridi's understanding of information simply corresponds to the general understanding of successful symbol handling).

In my view, this is also the case with his solution to the symbol grounding problem: In his agent-based model, data is given meaning through the coupling of internal states with the outcome of actions, whereby practical success becomes the truth criterion. Such an understanding of the problem fits very well with the prevailing ideas in neuroscience and the idea of structured neural networks in AI research, where semantic information cannot be understood as a set of explicit sentences or rules, but 'sub-symbolically' as a set of implicitly (via training data) defined sentences. In chapter 4 I try to show in more detail why I do not consider such a solution to be expedient, and then propose an alternative solution in chapters 5 to 7. Here, therefore, only briefly: If the human interaction with an environment is thought of as a process of exchanging quantitative information, the problem only shifts from the gap between data and explicit semantic information to the gap between data and implicit semantic information. Meanings are abstractions for which it is just as unclear at the sub-symbolic level how they can be derived in a stable way from fluid (environmental) data streams in fluid (inner) networks, as it is unclear on a symbolic level how signs and meanings are linked.

Based on his fundamental assumptions of a non-digital, but information-based ontology (reality is not digital, but structured on the basis of information), Floridi develops a concept of knowledge not as 'justified true belief' (also because he considers Gettier problems to be unsolvable due to the necessary coordination of justification and truth), but as semantic information for which we can give an account, whereby the admissibility of the account depends on the network of existing questions and answers. Recognizing truth then simply means 'being informed' and Gettier problems correspond to a fundamental scepticism about this state of affairs.

It remains unclear what role the concept of consciousness should play in such a construction of human thinking as information processing. Floridi proposes a 'knowledge game' in which questions that require counterfactual reflection (of the context) can sort out AI systems as 'syntactic machines', and questions that require subjective (self-)reflection can sort out zombies (unconscious humans) as 'semantic machines'. But if meaning is really 'only' the coordination of quantitative information, then subjective reflection is only quantitative information processing of a higher order and it remains unclear why these zombies should not be possible.

Floridi's most impressive achievement is that, although he focuses on semantic information, he consistently thinks this concept through to the end in analogy to quantitative information in his 'informational structural realism'. This can be regarded as successful if his agent-based solution to the symbol grounding problem delivers what it promises. The agent-based solution in turn thinks the quantitative concept of information through to its logical conclusion; AI systems and zombies should therefore readily agree with this 'machine philosophy', but humans only if the symbol grounding problem is really so easy to solve – and this question can at least be regarded as open.

The danger posed by Floridi's design is the successful reinforcement of the compartmentalization of the two concepts of information and thus the concealment of the actual problem of the question of their interrelationship. Methodologically, it isolates the different levels, relativizes the different views and thereby rather dissuades us to take a closer look at the indeed existing interactions. His approach is thus progressive in a naive way that has a practically conservative effect: It is completely committed to our current understanding of information, which is strongly influenced by physics, and thus closes itself off to necessary corrections from philosophy itself, but also from biology, for example.

Consistently thought through further, the philosophy of mind thus becomes part of a more comprehensive philosophy of information, which in turn is then placed on an equal footing with a philosophy of nature as a theory of science and (independent of both) value theories, or, as already spun even further by Floridi himself, becomes the first foundation of all philosophy as a logic conceived in terms of information. Nevertheless, this is only possible at the expense of denying the fundamental difference between qualitative and quantitative information and buying into the associated assumption that human thinking is ultimately 'only' quantitative information processing.

2.2 Thinking as quantitative information processing

Whether this extension of a scientifically motivated materialism into the philosophy of mind is expedient will only be assessable in retrospect. Central to this is the question posed above as to what extent an action-based model in which information feedback leads to stable mental constructs is really possible. This has been the object of research in neuroscience for years and is now gaining further topicality with the successes of sub-symbolic AI models. Floridi's philosophy of information gives these activities a place, but does not manage to contribute fundamentally new ideas to neuroscience or AI research, nor to see the existing problems in a new light. Thus seen, the difference between natural and (sub-symbolic) artificial intelligence (AI) appears to be a purely gradual one, which in turn makes it rather incomprehensible that sub-symbolic AI systems are simultaneously so competent and yet still not intelligent in the human sense.

Based on the observation that our understanding of human intelligence in AI research, neuroscience, psychology and philosophy has led to a kind of consensus in which human thinking is modeled as 'purely' quantitative information processing, I try to develop an alternative model in this book; not as a ready-made solution, but as an invitation to develop solutions of this kind with the same commitment. The current successes of AI research must clearly be recognized as important achievements of the aforementioned consensus; however, in view of the major unresolved problems in neuroscience and the philosophy of mind, it seems legitimate to ask whether these successes do not mark the zenith of a research paradigm rather than the beginning of a golden age: After all, the revolutionary products of AI research do not only show us their great possibilities, but also make us marvel at their aberrations already inherent in their design. [14] The price for the successes achieved is the fundamental inability of these systems to explicitly manipulate symbol systems; thus the central innovation that allows the great advances in application does not really help us in modeling specifically human thought. And this in turn also raises doubts about the effectiveness of the central models in neuroscience. In this sense, this text is a book for the time after the hype, when the great progress that has been made with sub-symbolic AI has become established in society, has found its way into everyday life and it has become clear across the board what such AI systems can and cannot do.

In chapters 3 and 4, the core problem of the connection between quantitative and qualitative information as the sought-after relationship between scientifically understood information (data, signals) and semantically understood meaning is presented, as well as the research paradigm of symbolic AI, Dreyfuss' critique of that paradigm and finally the paradigm of sub-symbolic (modern) AI. [3] The core thesis is that our problem already arises from our world view of materialism (in the sense of a purely physical naturalism), which underlies the natural sciences. In the search for alternatives, the possibility of a scientifically tenable idealism is then considered in chapters 5 and 6, in order to show in chapter 7 how better fitting concepts of information and meaning could be constructed in such a model. The resulting class of models is still naturalistic in the sense that it is informed by the natural sciences as an inductive metaphysics and must not only adhere to all established scientific principles, but must also be able to make comprehensible proposals for solutions where open questions exist. This is the case, for example, with the measurement problem of quantum mechanics, which is made clear in chapter 6 by means of an excursus on the integration of physical theories. Chapters 8, 9 and 10 then deal with possible objections from philosophy, psychology and neuroscience, including the question of the causal closure of the physical world, the rejection of esoteric arguments especially in psychology, and the concrete implementation as a research project in neuroscience. Finally, in chapter 11 I attempt to illustrate the extent to which the proposed alternative has relevance to our lives beyond the question of natural or artificial intelligence, especially in the area of value theories, where it can shed new light on ethical issues, e.g. in the areas of diversity and sustainability.

[3] In the following, I will - where not specified in more detail - speak of scientifically and quantitatively understood *information* (data, signals) and humanistically and semantically understood *meaning*, since the term information fulfills a relatively clear and established function in science and technology and the term meaning clearly establishes the special features of the qualitative concept of information.

Chapter 3

Symbolic and sub-symbolic AI

The current, seemingly sudden successes of AI research have, as expected, a longer history, which is also and especially characterized by unfulfilled promises of philosophical theories. When the project of mathematical-logical positivism, developed in the wake of Frege and Mach in the Vienna Circle and then by Russell and North-Whitehead in the sense of the 'Principia Mathematica' [15], was shaken by Gödel on its own ground, space was opened up not only for Quine and American analytical philosophy, but also for thinkers such as Turing, Church and von Neumann to take the first steps towards the digital revolution: Even if not everything can be grasped logically and mathematically, the question remains as to where exactly the limits of such an approach to 'calculating the world' lie.

Dreyfus then took up the criticism of the late Wittgenstein, but also of Husserl, Merleau-Ponty and above all Heidegger, which went beyond Gödel, when an idea of human intelligence analogous to mathematical-logical positivism began to establish itself in early AI research with the concept of 'symbolic AI'. In both cases, the question is whether the world can be grasped as a system of logical propositions, or whether a – linguistic? social? sensual? – being-in-the-world is the basis of our human understanding of the world. (If one would want to go even further back in history, one could also start this overview with Leibniz, with his attempts at an alphabet of human thought and then his '*characteristica universalis*', and even then you would still find precursors in Cusanus and Lullus).

3.1 Dreyfus' criticism of symbolic AI

Dreyfus argued (in vain for a long time) that human intelligence was of a completely different nature than assumed in the concept of 'symbolic AI', which early AI researchers worked on in the 1960s. [16] For him, human intelligence is not characterized by the conscious manipulation of symbols, but by a whole range of unconscious competencies for which it seems extremely questionable whether they can be fully captured in lists of formal rules. He highlighted four theses which, in his opinion, underlie the concept of symbolic AI, but whose validity is by no means proven: The biological thesis is that the human brain is an organic equivalent of a calculating machine. The psychological thesis behind this is that the human mind processes symbols in the same way as a computer, using formal rules. The epistemological thesis, which goes even further, is that all knowledge can be formalized in symbol systems. And finally, the ontological thesis is that the world is structured in such a way that formalization in symbol systems is possible in the first place.

In contrast, Dreyfus presents an image of man that shows him integrated in a physical and social context, which he draws on in the form of unconscious background 'know-how' for his conscious thought processes. (Similar ideas have been taken up again and again in other contexts since Heidegger at the latest, for example by Polanyi with his concept of 'tacit knowledge'. [17]) Not only the physical, sensual participation in the world is fundamental to this, but also the existence of needs, which makes the intentionality of human thought a further central point. It seems highly doubtful to Dreyfus that this situatedness, this 'local context', can be realized via formal rules in a symbolic AI system. He takes up arguments from both Heidegger and Wittgenstein here.

But while 'continental' philosophers in particular were still striving for a 'richer' image of man, the idea of the human mind as a quantitative information processing system behind symbolic AI was already gaining ground in the natural sciences and philosophical models close to the natural sciences, from McCulloch/Pitts' [18] first considerations on neural computations, via Newell/Shaw/Simon's 'physical symbol system hypothesis' [19] to the 'computationalism' of Fodor [20] and Putnam [21], as well as in the models of neuroscience. In the philosophy of mind, on the other hand, arguments in the tradition of Leibniz's mill [8] were developed by Searle ('chinese room argument' [7]), and in a broader sense by Jackson ('knowledge argument' [22–24]), Nagel ('What is it like to be a bat?' [25]), Chalmers ('zombie argument' [26–28]) and many others [29, 30], who have nevertheless not yet been able to steer neuroscience or AI research in new directions, if only for lack of alternative approaches. (For an initial overview of the now very extensive and detailed discussion, see the corresponding pages on qualia etc. in the Stanford Encylopedia of Philosophy).

3.2 Sub-symbolic approaches

Friendly ignorance towards the critical arguments from philosophy of mind was certainly enabled by the fact that AI research was turning to focus on the 'sub-symbolic' approach of neural networks and thus on a development direction that seemed better suited to address Dreyfus' criticism. (Albeit only in the 1980s after symbolic AI was unable to keep its overblown promises – and like the 'connectionists' in neuroscience before.) Instead of teaching AI systems 'common sense' via a huge database of formal rules, networks working sub-symbolically (i.e. at a level below explicit symbolic formulation) learn the implicit rules that apply to the interpolation between data points in training data (i.e. concrete examples). Dreyfus' criticism of the ontological as well as the epistemological thesis is already invalidated here to the extent that with a sufficiently large amount of high-quality training data, the difference between a world that is structured completely accessible (and therefore implicitly formalizable) or only largely so disappears, at least in practice: It does not matter whether the system can capture all conceivable cases if it covers all relevant ones.

What is particularly crucial, however, is that Dreyfus' criticism of the psychological thesis appears at first glance to have been constructively implemented: By feeding it with concrete examples, the machine now appears to acquire its 'local context' in the form of implicit rules, i.e. in the form of the regularities between data points contained in the training data. Dreyfus' criticism of the biological thesis is then also invalidated to the extent that the biological computing machine can now model not only explicit but also implicit knowledge. The fact that AI systems thus become 'opaque' only makes them appear even more human, as of course not even we ourselves are always clear about all our thought processes.

At second glance, however, we have gained far less than the above explanations suggest – and have also given up something essential, namely the ability to explicitly manipulate symbol systems. Although invalidated in practice, Dreyfus' critique of the ontological and epistemological thesis plays a role also for sub-symbolic AI systems: Their situation is ultimately still similar to that of Mary in Jackson's knowledge argument; depending on their data set, they 'know' everything relevant to their function about red – including perhaps multiple color encodings – and yet have never seen it like a human. Every example fed in is just a set of physical symbols (data, signals, quantitative information) for which a human must initially establish the relationship to the color red. And no implicit laws learned in this way, even if there would be an infinite number of them, could make it possible to derive the red color impression from them. (The situation is not much different when it comes to the intentionality of the systems.) However, this means that Dreyfus' criticism of the psychological and biological thesis has not been sufficiently refuted either: Human thinking appears to operate not only with symbol systems (in the form of data, signals, quantitative information) but also at least with qualities; which in turn calls into question the role of the brain as a 'simple' calculating machine.

In a broader sense, Dreyfus' situatedness in a world is not only about sensory qualities, but about qualities in general. Even if we were to concede to sub-symbolic systems that they can model meaning in the sense of concepts (i.e. semantic information or qualities in the linguistic sense), for instance via the formation of stable network structures on the basis of action feedback as propagated by Floridi and many connectionists in neuroscience, the 'local context' of the human being would still include completely different types of meanings; in addition to linguistic concepts not only qualia, but also non-linguistic abstract entities such as numbers, and also ethical/aesthetic values, all of which can be understood as qualities (ideas, forms) with universal character.

In order to do full justice to Dreyfus' criticism, including what he took from Heidegger, we would have to broaden our initially 'narrow' concept of meaning as linguistic-semantic information to a 'broad' concept of meaning as general-qualitative information. In chapter 5 we will see that this happens 'naturally', so to speak, when we consider idealistic alternatives to our current, materialistic view of the world. For our critique of the thesis that human thought can be understood as purely quantitative information processing, however, we first want to make the counter-movement once again and understand the concept of meaning very narrowly in the sense of linguistic-semantic information, because neither the decision to broaden the concept of meaning, nor the departure from materialism can be formulated as a logical necessity, so that it must first be shown that not even a minimal model of meaning can be realized on the basis of sub-symbolic AI systems, or more accurately, that the possibility of such a realization seems at least very questionable. The minimum performance to be achieved is the 'stable' abstraction of concepts on the basis of quantitative information (data, signals). Our initial question of how information and meaning are connected, or more precisely, how information acquires meaning, thus shifts to the question of whether meaning can be understood at all as resulting from quantitative information (i.e. as directly derivable from a system of logical propositions or relationships - whether implicit or explicit). This will be examined in the next chapter.

Chapter 4

Artificial and natural intelligence

In the last chapter, we used Dreyfus to argue against classical, 'symbolic' AI. However, the question remains as to whether and to what extent our criticism derived from Dreyfus can also be applied to the 'sub-symbolic' AI approach that is currently on everyone's lips. This will be examined below before considering which paths could lead from here to an alternative description of natural intelligence.

4.1 Sub-symbolic AI

The neural networks behind sub-symbolic AI are also developments from the early days of machine learning in the 1940s and 50s. [18] Initially they were less in the focus of AI researchers, because the symbolic approaches from Chapter 3 seemed more suitable from a fundamental point of view for reproducing human intelligence, but they served as a central model for the developing neurosciences from the very beginning. It was only with the failure of the symbolic approaches in the 1970s (or rather the inability of these approaches to catch up with the promises associated with them), as well as a series of mutually reinforcing developments from the 1980s onwards, that sub-symbolic AI was able to assume its current prominent role.

The decisive factors for this were algorithmic developments as early as the 1980s/90s (especially by Rumelhart/Hinton/Williams [31] and Schmidhuber/Hochreiter [32]), the increasing availability of large amounts of data from the 2000s onwards (thanks to the Internet, mobile devices and cloud technologies), and the availability of very large, massively parallel computing power (thanks to graphics processors, GPUs), especially in the 2010s. This decade also saw the first breakthroughs, such as the winning of the ImageNet object recognition competition by the 'deep', i.e. multi-layered, neural network (DNN) AlexNet, and the successes of AlphaGo and later MuZero in games. The years 2015/16 are often seen as the 'AI turning point'. From 2020 on, this development continued to gather pace. In this year, AlphaFold won the CASP protein folding competition, making it clear to the wider scientific community that AI can contribute to solving 'serious' problems. (On the other hand, it is also became clear that the companies involved know how to functionalize the scientific enterprise for their advertising purposes.) Sub-symbolic AI then received widespread attention with the publication of 'large language models' (LLMs) such as GPT-3/4 and then ChatGPT, Bard, etc., which appear to be capable not only of natural language processing, but even of natural language understanding.

Largely independent of the discussion on the extent to which natural intelligence can be described analogously to sub-symbolic AI, it can be stated that Geoffrey Hinton's testimony that AI is 'able to do everything' [33] is certainly correct insofar as DNNs are suitable for solving arbitrary (practical, if not all theoretical) problems that can be defined implicitly via data, given sufficient availability of data and computing time. This will not be questioned also in the following. The fact that the successes of DNNs are primarily seen in the field of object recognition and speech processing is certainly due to the fact that vision and speech are our primary interfaces to the physical and social world and therefore the advances in these areas seem particularly intelligent to us.

Accordingly, the consequences of the 'AI revolution' will be felt above all in everyday life, where AI systems can take over simple tasks, making them automatable, highly reliable and highly available: A team of AI employees can not only be easily 'updated', but also 'scales', i.e. can be easily expanded by additional (practically unlimited) units. In many environments, these systems will work at least 'normally competent'; in addition, environments will certainly be adapted (and people will adapt) to reap even more automation gains. 'Low-hanging fruit' are all tasks that can be formulated as language processing or object recognition problems. There is therefore no question that the new AI technology will have a major influence on our societies and, above all, on the achievement of sustainable development goals [34], from individual work environments [35] to global geopolitics [36] – for better or for worse and for our discussion here even more important; regardless of whether these systems are intelligent in the human sense. And this also applies to the dangers posed by autonomous weapons, AI systems running amok, fake news and the problem of the (un-)fairness of purely data-based algorithms. [37] The latter seems to me to be the most urgent problem that our societies have to face; however, researchers are of course already addressing this problem, [38] which will therefore only be briefly touched on here:

In the ethics of artificial intelligence, we are not confronted with new questions of meta-ethics (e.g. about the conditions and possibility of ethical judgment) or normative ethics (within the frameworks of which general considerations about ethical judgments should be made), but with questions of applied ethics, which are not completely new, but do now arise in a completely new context. These questions concern, for example, the safety or fairness of this new technology. They are particularly interesting due to a lack of transparency of the algorithms, which amongst others require an explicit allocation of responsibility for decisions made: As those affected, we (sensibly) want to know on what basis decisions are made and who is responsible for them. Due to the nature of DNN AI algorithms outlined in more detail below, this raises a whole series of interesting questions around the concept of explainability, which can in addition benefit from the well-established reflection of the concept in the philosophy of science following Hempel. Accordingly, 'explainable AI' (XAI) has become an important, interdisciplinary research topic. Current work deals, for example, with the question of the extent to which the feasibility of XAI can be derived from underlying laws in the data (Eva Schmidt) or with the – so far rather unclear – reliability of XAI methods, which are in addition subject to fundamental limits on classical computing architectures (Gitta Kutyniok). Finally, the topic of sustainability of and with AI is establishing itself as the 'third wave' in AI ethics after the first wave with questions about safety and the second wave with questions about fairness.

To understand how these systems are so competent and yet not (in the human sense) intelligent, it is important to consider how they work: [39] In general, in this form of machine learning (supervised learning), a certain (set of) output(s) is linked to a certain (set of) input(s) via a mathematical function. This can be as simple as assigning a y-value to an x-value, or as difficult as assigning an animal name to a particular class of pixel patterns, or a word to a given incomplete sentence. In the easy case of the x,y-values, the function required may look very simple, but the functions needed to assign images or words will generally be much more complex. In principle, it should be fine to select any sufficiently flexible function, whereby its flexibility is determined by a (very) large number of customizable parameters, i.e. 'adjusting screws', so to speak. It should be noted that the form of the function can, on the one hand, facilitate adaptation to the task at hand if, for example, basic properties of the underlying relationships are already captured by the general form of the function. On the other hand, it can also make this more difficult or even impossible, since the given form represents a prejudice or 'bias' that may not be consistent with the modelling task. With deep neural networks (DNN [40]), we now have a (nested, non-linear) function that is not only largely bias-free, but which – thanks mainly to algorithmic developments in the 1980s/90s [32] – can still be optimized, i.e. adapted to the circumstances, even with the very large number of parameters required for object recognition and speech processing (hundreds of millions in number).

After the accelerated development in the 2010s, the computing time used to train new AI systems shows another leap in development from around 2016 onwards. [41] It was around this time that the companies involved began to fully explore the potential of DNNs and develop so-called 'foundational models' with maximum resources in terms of data and computing time. These included large language models (LLMs) for language processing, diffusion models for image processing and derived systems such as text, image, audio and video generators, as well co-pilots (AI helpers), e.g. for programming and office tasks. As impressive as the developments are – especially in the area of (e.g. text and image combining) multi-purpose models – it is already clear that the limits of this development direction will be reached within a few years: In addition to the extremely high costs (already in the order of 100 million Dollars in 2023), the end of the availability of further high-quality data seems imminent and another leap in available computing time, as was the case with GPUs, is not in sight. (Current models use hundreds of network layers, billions of nodes, hundreds of terabytes of data, corresponding to billions of words). The greatest influence on further development will probably come from the faster pace of legislation regarding the fairness of AI algorithms – particularly important here is the EU's AI Act –, but also with regard to copyright issues. It should also be seen as a positive development that open source models now seem to become competitive to commercial ones. In the future, model-related security concerns will become increasingly important, e.g. how models can be effectively protected against 'adversarial attacks' (the 'tricking' of algorithms) and 'data poisoning' (the 'poisoning' of the data on which the models are based).

With LLMs such as ChatGPT or Bard, the special mixture of high competence and lack of (human-like) intelligence can not only be experienced directly in the form of the frequently observed 'hallucinations' (of incorrect facts), but also well illustrated by sketching the way they work: First of all, language must be represented mathematically in order to make it accessible to DNNs, which after all combine numbers as input with numbers as output. For this purpose, language 'tokens' (words, suffixes, punctuation marks, etc.) are defined (e.g. approx. 50,000 for GPT-3), to which vectors are assigned according to a statistically determined relationship of 'meaning'. A certain number of tokens can be processed simultaneously (e.g. 2000 for GPT-3, corresponding to a newspaper article, 32000 for GPT-4, corresponding to a short book) in order to determine word probabilities over longer sections, whereby the computing time does not increase linearly with the enlargement of this 'context', but much more strongly. 'Attention' mechanisms allow the efficient coding of the essential relationships, as a kind of penalty against 'overfitting' to non-relevant ones. [42] The network can now be trained 'self-supervised', i.e. without further human intervention, in billions of parallelizable runs by assigning a part of a text to it as an input, for which the next token is to be found as output. The parameters of the model are gradually adjusted so that the word actually used in the text is assigned the highest possible probability. As a result, the parameters implicitly and globally encode the rules of language (or rather token) usage in the training data.

Successfully trained, the AI system can generate text by determining, based on a given input text, which language token (according to the training data) is most likely to follow the given text. This token can be appended and the expanded text generated in this way can be used again as input, so that a longer output is generated 'autoregressively' step by step. In practice, however, it has been shown that text generated in this way very quickly looks repetitive and generally uncreative to people. For this reason, the known models also use slightly less probable words with a certain frequency (depending on the so-called 'temperature' parameter), which significantly improves the results. With a view to a mathematical-theoretical explanation, this solution is less satisfactory, as the existence of this parameter clearly calls into question any real understanding on the part of the system. In practice, the above is followed by 'fine tuning' training runs for specific purposes, sometimes with the support of the language models themselves, but usually not entirely without human 'labeling' (of facts) and evaluation, e.g. in the form of 'reinforcement learning from human feedback' (RLHF). [43] It is also common to provide the models with certain content hidden in the prompt as 'domain context', or to allow the model to access predefined databases with content (in the form of vectors) for answers.

The great impact made by the publication of LLM-based AI systems is based to a large extent on the fact that users get the impression that the system can not only process language, but can actually understand it like a human. On the basis of the above mechanism of gradually attaching the most probable token, it appears at first glance that certain competencies emerge that are intelligent in the human sense. Just as it is hard to escape the individual amazement at the successful communication with the machine, the majority of commentators and scientists alike seem to be unable to avoid the temptation to interpret the purely statistical behavior of the system, as defined in the design, as an act of emergent, human-like understanding in each new individual case. And already for advertising purposes, the model designers are happy to jump on this bandwagon and fuel such speculations by means of the already proven functionalization of the scientific system. [44] (Which includes the identification of complex mental mechanisms such as attention or the emergence of mental representations with complicated, but ultimately sub-complex, mathematical algorithms such as 'attention' [42] or 'representation engineering'. [45])

However, the functionality of AI systems outlined above shows that such 'emergent' rules must always be implicitly contained in our training data, which can normally be proven in detail if the models are sufficiently documented. And this applies in particular and unfortunately also to all the prejudices that humanity has culturally codified in the course of its history, which then becomes the core problem for the (un)fair algorithms. (Even with best intentions, it can be extremely difficult to identify and treat all potentially discriminatory 'proxy variables' as such). [46]

Compared to other AI systems, LLMs benefit from the fact that their training data comprises a good part of the knowledge of mankind. All that is (humanly) intelligent about these models is that they work with meaningful language tokens (originally defined by humans) and an extremely large database meaningfully compiled by humans (including 'click-workers' in the Global South). This insight demystifies such models, but should not be understood as simply talking down their capabilities: Unlike the individual human being, the model does indeed have access (albeit of a different kind) to the entirety of the given data and – which should be seen as the real feat on the part of the developers – can make this knowledge available in a linguistically competent manner. There is also no question that the 'remixing' of given content will allow many conclusions to be drawn that humans have not yet drawn (as an explication of implicit rules whose existence in the training data we are not aware of), so that these systems also have a certain potential for innovation and, in an analogous sense, 'creativity'. The 'hallucination' of false facts is then not a bug, but a feature of such systems, as they refer to a lack of data for comparison with the 'real' world. Only that every option used as a 'truth maker' (other programs, databases, online access, etc.) provides relatively little and selective information that can solve the problem at this point, but not in general.

In line with this observation, data-driven, statistical approaches were initially conceived as a stopgap solution to simulate natural language *processing*, if not natural language *understanding*, and many researchers in this field still consider them unsuitable for enabling the latter for fundamental reasons: Natural language seems to be designed to encode the minimum that is necessary against a common background (also known as the 'missing text phenomenon'); teaching an AI system all the syntactic and semantic variations necessary for this – even if implicitly via training data – appears to require much larger amounts of data and computing capacity than those currently used. [47–49]

Further details on the analysis and criticism of sub-symbolic AI systems in this direction can be found, for example, in Marcus [50], Bender/Koller [51], Mahowald *et al.* [52] and others, but here our focus is on the question of whether we can understand human intelligence analogously to the current sub-symbolic AI systems, which we can probably answer in the negative at this point with a clear conscience after the above explanations. But what exactly are the differences? First of all, it seems clear that at least parts of our brain do not function completely different from DDNs, e.g. in the basic processing of sensory impressions. For this reason alone, the use of DNN-like models was and is central to neuroscience. [53] Beyond this, human intelligence is based on factors such as consciousness, intentionality, qualia and embodiedness, but differences can already be identified in the narrower sense: In contrast to sub-symbolic AI models, 'higher' thought processes appear to be organized more as conceived by symbolic AI, e.g. in the form of semantic networks. At the core of this other form of thought organization are stable abstractions in the form of terms, concepts or ideas. We do not find such stable abstractions in sub-symbolic AI systems and they also pose a major problem for the corresponding 'connectivist' models of neuroscience.

As mentioned above, compared to neuroscientific models, LLMs benefit from the fact that they already work in a space of defined concepts (the given language tokens), while it is still unclear for the brain how exactly such concepts could be realized as stable neural circuits (more on this in the next section). Furthermore, LLMs have access to a vast number of other abstractions as implicit regularities in the training data, which, however, are only stable for a given parameterization of an LLM. Simple DNNs that continuously adapt their parameterization according to their input (their 'experience') can easily be 'poisoned' by targeting the supply of data, as has already been observed in practice with racially derailed chat bots, for example. (And this can happen even more so to a correspondingly 're-trained' model by inverting what it has learned; Cleo Nardo called this the 'Waluigi effect'.) The only way to give such a chat bot a truly stable idea of politeness is to prefix it with filters, i.e. firmly implemented (behavioral) rules.

Any 'emergent' rules that we observe for DNNs are ultimately based on the global parameterization of a nested function that is continuously in a state of flux through each additional data input without any qualitative inhibition threshold. Here, too, it is clear that parts of the human brain must function very similarly, but also that we are additionally capable of forming 'real' abstractions, which seem to be characterized by an unexpected (quantum leap-like) stability, as well as the fact that they – unlike in the DNN – do not seem to be completely defined by the sum of the underlying individual cases. It is precisely the 'broadness' of an abstraction such as 'truth' that seems to be a central advantage of human thought. LLMs also have a concept of truth, but this is predetermined by a language token and associated data-implied rules; it is never complete (final would probably be too strong for the human condition), always narrowly abstracted (on the basis of given cases), and also extremely vulnerable (much more so than in humans) to poisoning by manipulated data.

In contrast, humans seem to have an astonishing ability (and tendency) to abstract broadly and to resist the described erosion of abstraction. (This does not call into question the fact that, on the one hand, the brain works with 'unstable abstractions' in the form of sums of quantitative information at the level of quantitative information processing and, on the other hand, is not to be understood as an assertion that our mind cannot abstract false assumptions; only that these are still stable and broad even in their falsehood).

4.2 Neuroscientific models

Now one could assume that sub-symbolic AI in the above sense only has the wrong 'endpoints' to model human intelligence. [54] Instead of predetermined language tokens and a fixed set of training data, we should assume 'raw' sensory data as input and the possibility of experiential learning, and additionally assume the neural circuitry to be malleable enough to allow for 'real' abstractions in the form of stable neural structures or processes, so that abstraction erosion does not occur. This is the general thrust of the 'connectionist' approach in neuroscience and the cognitive sciences, about which we must at least gain an initial overview for the following (my short summary is based on an overview volume by Maurer [55]). [1]

The connectionist approach comprises a class of neuroinformatics models that attempt to represent cognition as parallel, distributed information processing in neural networks. The basic program for this was already presented in 1986 by Rumelhart and McClelland. [56] The underlying arith-

[1] Initially, very abstract theories such as functionalism, with its notion of a mutiple, i.e. essentially biology-independent, realizability of thought, were in the foreground in neuroscience in the 1970s/80s. Then neurophilosophy argued against explicit representation and for sub-symbolic theories of meaning on the basis of partitions of vector spaces, with the brain as a vector-to-vector transformer. Since the 90s, theoretical neuroscience has been primarily concerned with the above mentioned connectionist models, but also for instance differential equation systems for the description of ionic currents in the underlying structures.

metic operations correspond to activation forwarding between neurons, depending on continuously updating synaptic connections. In addition to the mathematical-physical theory of dynamic systems, the paradigm of self-organization plays a central role here. Thus cognition is understood on the basis of self-organizing, sub-symbolic processes that have to manage without external reference or control functions such as symbolic structures. The various models are then classified according to their degree of localization, depending on how locally or globally patterns can be assigned to network locations. A particularly interesting research question is then how neuronal patterns synchronize and thus, for example, how different properties of an object can be understood as properties of this one object (an example of the so-called binding problem [57]). In these models, abstractions such as terms or concepts must be understood as dynamic processes that filter out certain statistical prototypes from the neuronal input and for which it must be assumed that, despite their fluidity, they are able to establish themselves as stable 'attractors' in the neuronal activity patterns, e.g. via resonances. (Since such self-organizing processes are correlative and not causal in nature, the extent to which they can be causal for conscious phenomena remains in any case unclear.)

The advantages of the connectionist approach are first of all the dynamic, self-organizing, adaptive learning, which is also capable of processing incomplete, incorrect and contradictory input, as well as the distributed, parallel and active information processing. Somewhat less obvious, but also central, is that in this approach memory locations are addressed in a content-dependent manner, whereas this is generally not the case with symbolic approaches.

Disadvantages of the connectionist approach for the description of human cognition are, first of all, the lack of reference to the manipulation of symbolic structures, the limited generalization of what has been learned, and the relatively 'slow' (i.e. data- and computing time-intensive) learning, especially in the case of high-dimensional problems. We can relate these points to the human ability for 'real' abstraction assumed above. Another central problem of connectivism is the stability/plasticity dilemma, also known as 'catastrophic forgetting', namely how new patterns can be learned without too much change to existing patterns. This in turn can be related to the erosion of abstractions assumed above. In the competition between symbolism and connectionism, the suggestion of a middle way has emerged in recent decades, especially with regard to Kahneman's observation of two modes of human thought (more on this in a moment), but here the problem of integrating sub-symbolic and symbolic models arises.

Many current developments in the field of AI research can be seen as attempts in this direction, such as Le Cun's Joint Embedding Predictive

Architecture (JEPA) or 'causal inference' based approaches in the wake of Pearl [58], where data (distributions) are used to infer underlying processes and objects in order to build 'world models' that could in the longer term also serve as a context for counterfactual reasoning or 'value alignment' (the alignment of AI systems with ethical principles). As 'neuro-AI', however, this approach naturally also has a longer history. [59–61] Corresponding approaches nevertheless suffer from the lack of ideas for a generalization of the integration of the two levels, which is why the existing models are essentially dependent on extensive contextual specifications by humans. Here too we then find the hope that the problem could be solved using specific hardware structures, in analogy to specific neuronal structures. Hinton's idea of 'mortal computing' (in which the computing processes are no longer 'immortal', i.e. independent of the underlying hardware) can be interpreted in this direction. [62]

In essence, this corresponds to the attempt to solve the problem 'agent-based' as with Floridi, i.e. by deriving the meaning and stability of mental objects from the interaction with an environment, but this has so far failed both practically and theoretically: In practice, reinforcement learning by means of quantitative information from the interaction with an environment corresponds to classical condition; higher-dimensional (complex) tasks are learned only slowly and an exploration/exploitation problem arises, namely that it is unclear at what point 'enough' has been learned. How much (sensorimotor, neuronal, psychological, ...) structure formation helps with this depends on the structure of the respective problem: The more local the problem, the more helpful is the localization of structures. Setting a leg, for example, is in the very first approximation a rather local problem; here it is accordingly of great help to carry out certain motor calculations locally, e.g. through the intrinsic reflexes of muscles. For higher thought processes, however, it remains unclear how structure formation could completely solve the problem. From a theoretical point of view, we can also use well-known psychosemantic considerations in the philosophy of neuroscience to point out that a causal covariance is not yet a representation, i.e. that causal relationships are only a prerequisite for meaning. [2]

[2] Finally, it should be mentioned that research into 'classical' symbolic AI has not stood still either: The technical 'frame problem', derivable also from Dreyfus' criticism, i.e. how symbolic AI systems can make practical or context-dependent decisions without always having to explicitly consider a large number of obviously unimportant rules, is today regarded as solved in the narrow technical sense. Accordingly, there are theories for the formalization of common sense (background) knowledge (McCarthy, Lifschitz), especially also on the basis of physical and social world models (e.g. by Davis). However, the creation and evaluation of sprawling formalization models remains a challenge and, more

Neuroscience thus seems to rather support the assertion that human thinking cannot (yet?) be modeled as information processing, at least not in a simple – be it symbolic or sub-symbolic – way. This observation can now be made more specific in so far as that there is an inherent contradiction between basic, presumably sub-symbolic, and higher, symbolic modes of thinking, which does not seem to be resolvable within or by simply combining the existing models. More specifically, the resolution of the contradiction seems to require mediation between stable, broad abstractions and the fundamental, quantitative information processing that can certainly be modeled via neural networks. (We return to neuroscience in chapter 10.)

4.3 Psychological models

As briefly alluded to above, this view of the problem (albeit not explicitly formulated) is also supported by more recent research findings in psychology. Kahneman has proposed that cognition has two modes; a fast, instinctive-emotional mode and a slow, reasoned-logical mode. [63] The important significance of the first mode for everyday life, in combination with the existence of a whole series of cognitive biases of this mode, leads to a substantial irrationality of human thinking, which is particularly evident in statistical reasoning and in the self-assessment of one's own rationality. Kahnemann's thesis is not uncontroversial in detail (if only because some of the studies cited were called into question in the course of the 'replication crisis' in psychology), but it at least seems clear that human cognition covers the entire spectrum of thinking styles spanned by the two hypothetical modes. As mentioned above, with regard to Kahnemann, AI researchers are also considering (and have been considering for some time in the context of 'neuro-symbolic' approaches) a possibly necessary combination of sub-symbolic and symbolic approaches, whereby the focus here is on a mathematically-technically conceived link for which it would remain open from a neurobiological perspective, of how symbolic structures could be neurologically implemented and stabilized.

This leads us to the question of whether there can be psychological factors that could have a stabilizing effect on abstractions such as concepts. For this, it must first be noted that non-applied psychology, without a grand unifying theory like evolution in biology, is understood to be divided into more or less overlapping areas. For our investigation here, these can be summarized into three groups, covering firstly cognitive, secondly affective and

generally, it remains unclear how practically computable decision-making processes could be delimited in completely open contexts.

thirdly fields that extend beyond these areas, such as consciousness, language or psychomotor skills. It is probably not too much of an injustice to assume, with regard to our research question, that two paradigms in particular can be regarded as central to modern psychology: Firstly, the thesis of the modularity of the mind, which is to be understood as composed of different interacting units, and secondly the information processing paradigm, namely that these units and their interaction can be explained by information processing in the nervous system. Both paradigms appear to be compatible with the neurobiological considerations above, but this also means that the puzzle of abstractions and their stability remains.

Beyond the cognitive and probably also the affective group of phenomena, the complex interactions between people, the environment and society then become the focus of attention. Are there stabilizing elements in the bodily or social nature of human beings? Unlike Solm, for example, who sees nothing special in the fact that categories of sensory perceptions are differentiated via qualities, Fuchs at least also recognizes the hard problem of qualia in psychology, but sees the problem definition as somewhat misleading, since consciousness should be understood as a process of 'enaction' and 'eninteraction' of the subject in its physical and social environment. [64] Would a physically and/or socially situated neural network be able to escape the difficulties discussed? Here we again approach Dreyfus' own arguments, who sees the central difference in the situatedness, the having-of-a-world, of natural intelligence. However, this does not appear to be compatible with the scientifically understood information processing paradigm: If being physically or socially located only means exchanging quantitative information with an environment, then this environment cannot achieve the hoped-for stabilizing effects according to the above considerations. (Although it is quite conceivable that the practical problem of the hallucination of facts by LLMs can be mitigated this way).

On the contrary, the gap between sub-symbolic and symbolic approaches in cognition only seems to widen for the 'whole' person: The difference now seems to merge into the previously discussed one between quantitative information and meaning. In addition to the cognitive abstraction of concepts, the human world appears to be populated by a whole series of other 'abstractions' or meanings; colors and smells, but also emotions, numbers, ethical annd aesthetical values, etc., which also seem to clash with a naively conceived information processing paradigm. (We return to psychology in chapter 9.)

4.4 Philosophical models

This brings us back to philosophy. We have already discussed the *Lebenswelt* (lifeworld) that, according to Dreyfus, distinguishes humans from machines in chapter 2. But even in the narrower sense of this chapter, i.e. in the restriction to cognitive processes, we run into the problem outlined above not only in AI research, neurobiology and psychology, but also in philosophy:

Since McCulloch/Pitts [18], Putnam [21] and Fodor [20], and despite arguments by Putnam himself, Searle and others, the 'computational theory of mind' (see e.g. Chalmers [65] for an overview) has persisted among a number of modern, mainly analytic philosophers in the English-speaking world. They seek to describe the human mind as a 'pure' information processing system, mostly in the sense of a neurally implemented Turing machine – much like Floridi does. As we have seen in the previous two sub-chapters, their work had and continues to have a great influence on neurobiology and cognitive psychology, and here in particular the idea that the quantitative processing of information itself, independent of the physical circumstances, constitutes mental phenomena. In order to be manipulated in computational processes, these must then be understood as 'mental representations'. This, however, opens up again the above-mentioned gap: On the basis of fluid, syntactically defined information processing procedures, the mind manipulates stable mental objects that serve as the basis for a symbolically/semantically conceived 'language of thought'.

This problem can be further complicated if we distinguish between a 'computational language of thought' (in the wake of Fodor) and a 'representational' one (in the wake of Ockham), as did e.g. Chalmers in response to Quilty-Dunn/Porot/Mandelbaum. [66] Since sub-symbolic information processing takes place at a level below the 'atomic' (not further decomposable) representations, sub-symbolic AI systems and naively conceived connectionist approaches in neurobiology are in conflict with the (symbol-level) 'computational' variant. However, structured connectionist approaches and analogous sub-symbolic AI systems could be compatible with the 'representational' variant. Here again, the relevant idea for us would be that certain stably emerging network structures and/or processes could be the basis of stable mental objects. According to the above discussion, this seems unlikely to me; however, further investigations in this direction should of course by all means be carried out both in philosophy and in neuroscience, since the hope that future approaches in the field of dynamic systems theory might be able to show how the sought-after spatial or temporal superstructures could be stably implemented, must not yet be given up. (We return to philosophy in Chapter 8.)

4.5 Abstract entities

If we now want to further investigate the finding of stable and broad abstractions, we encounter the problem of having to specify what exactly is supposed to constitute the process of abstracting and what kind of products then emerge from it. The problem with this is that the existence and nature of abstract entities are the subject of a largely open debate in academic philosophy. As Falguera/Martinez-Vidal/Rosen write in their overview article on Abstract Objects for the Stanford Encyclopedia of Philosophy (as of 09.08.2021):

The abstract/concrete distinction has a curious status in contemporary philosophy. It is widely agreed that the ontological distinction is of fundamental importance, but as yet, there is no standard account of how it should be drawn.

The debate is not only about what can be considered an abstract entity (properties? concepts? possible worlds?), but also whether such entities exist (Platonism/Platonic realism) or not (nominalism), or in what attenuated form this may be the case in between. The different positions are often closely linked to fundamental ontological and epistemological assumptions of their proponents, which of course further complicates the discussion. It should also be noted that while the terms used refer to historic predecessor debates, their meaning has changed significantly in some cases. Thus Plato's ideas had causal relevance, whereas the modern Platonist usually regards causal inertness as part of the definition of abstracta, and in the controversy about universals in the Middle Ages other criteria were again in the foreground. The modern debate about abstract entities was initiated by Goodman and especially Quine, [67, 68] the latter with the thesis that our best scientific theories oblige us to assume the existence of mathematical entities, which was in turn questioned by amongst others Field [69] and Benacerraf, [70], from the latter with regard to the epistemic accessibility of abstracta. Even before that, from Locke's idea that we form abstractions by omitting empirical details to Kant's attack on the classical distinction between empiricism and rationalism, abstractions have played an important role, especially with Frege, [71] who assumed a third realm of objective (non-mental), non-physical entities, which was later taken up by Popper as a 'third world'. [72]

In the modern discussion, non-physical is then usually defined as nonspatial and causally inert (i.e. without a distinct, 'direct', causal contribution) – certainly with a view to the general problematic nature of the concept of causality. Further developments of Locke's approach to possible algorithmically definable processes of abstraction are now viewed rather critically; no approach seems to cover all cases. As indicated above, the apparent lack of epistemic access can even be developed with Benacerraf into an argument against the existence of mathematical entities, albeit one that is not entirely free of contradictions. Against a simple identification of abstracta with universals (for which it is impossible to be indistinguishable but distinct), 'immanent' universals are brought up, which are anchored in the particular and are thus distinguishable, e.g. 'this one color in all bodies of this color'. (Cowling [73] or Künne, [74], amongst others, provide an overview of the diverse discussions on abstract entities.)

Strictly speaking, for our considerations on how an alternative understanding of natural intelligence could be formulated, we would now have to go through all the options individually, but because the different possibilites are, as noted above, mostly intertwined with other ontological and epistemological options, this would go beyond the scope of the text almost immediately. We therefore have no choice but to follow the path that seems most suitable for our project (fully aware that this is an invitation to *postmortem* criticism). With regard to the – in addition not only weak – existence of abstract entities, the answer will therefore be that this is taken for granted in the following, since an answer to the contrary leads us almost immediately to the model, for which we want to develop an alternative possibility.

It should therefore be assumed that there are objective, non-spatial and causally inert building blocks of reality. (And thus abstract objects are real in a very concrete sense, similar to Meinong, for example). And since we also want to consider models in which space may be an emergent phenomenon, we do not regard the possible existence of immanent universals as an argument against the abstract building blocks being universal in nature; but also not as a conclusive argument for their existence, which is simply assumed here for the sake of exploration. The particular is then initially only possible as a specific combination of universal properties. With regard to epistemic access to abstract entities, the idea (advocated in mathematics by Gödel, among others) of a fundamentally given, intuitive-mental access to them [75] appears to be the only way to circumvent the problems formulated by Benacerraf.

In the academic discussion of abstract entities, these are very specific commitments, but they can perhaps be credited with the fact that – especially as far as mathematical entities are concerned – they seem to correspond very well with the everyday intuitions of philosophical laymen. Well-known academic arguments against Platonic-realist models are aimed at logical paradoxes or recourses as in Plato's Parmenides or Bradley's work, [76] but are not widely accepted as inevitable and are sometimes regarded as self-refuting; so there is no irrefutable veto against our choice either. Natural scientists, on the other hand, might be particularly annoyed by a supposed lack of minimalism in ontology. However, it should be noted that those natural scientists who are not unconditionally materialistic, but think in terms of real physics, also assume the existence not only of universal mathematical properties, but of universal physical properties such as charge or spin as well. (Otherwise the laws of nature would have to be thought of as sums of practically infinitely many individual types of interaction.) But what is the relationship between the universal property and the specific particle? Should we imagine it as a (quasi-)local excitation of a universe-wide quantum field? Analogously, we could then imagine further fields for non-physical properties. In a certain sense – although possibly not in the original Platonic sense – we cannot completely avoid the fact that physical entities seem to participate in universal properties. (Though we will clearly have to discuss the 'ontological parsimony' of possible models later on again.)

So as not to narrow our view prematurely, we will furthermore make no distinction between the (then congruent) entities abstracta and universals, but also not between these and (now always universal) qualities, and will use these terms interchangeably in the following. (And as long as Platonic ideas are understood as causally inert, they are additionally included.) This family of terms then includes a whole range of entities, such as colors, numbers, terms, concepts, values, etc. Later, a further qualification may seem necessary. The one distinction that should be made already here is that between basic abstract entitites as building blocks and complex abstracta, which should be understood as being composed of the more basic building blocks. The core hypothesis is therefore that abstract entities exist in the form of universal qualities or their possibly individuated combination and that these are intuitively accessible to natural intelligence. (Later on, of course, it still has to be discussed, what exactly this intuitive-mental grasping of ideas is supposed to mean). We have arrived at this hypothesis here via the finding of stable and broad abstractions that cannot be easily explained in another way, but we will see that going down this road will allow us to explore further possible peculiarities of human thought like qualia, intentionality, etc. more easily. In contrast, rejecting this core hypothesis seems to lead inevitably back to the current consensus of purely quantitative information processing, with the associated problem that the practical implementation of this program in science and technology does not really seem to work out. (Though we will also not succeed in formulating compelling arguments for simply rejecting the latter position altogether).

On the basis of the above considerations, we can now make the distinction between quantitative and qualitative information within the proposed framework: Quantitative information always appears as a change in qualities (e.g. a signal as a change in the flow of charges), qualities themselves as abstract entities. Since only the change is essential for quantitative information (and the translation is always carried out by a subject, both initially and at the end), it can be realized independently of specific qualities, in the way we can model its processing with Turing machines. The difference to be assumed between purely physical computing and human intelligence would then be that the former can only operate with quantitative information (so that a human must always assign the context of meaning in the first and last step), while humans resemble such machines in body, but only process qualitative information, i.e. meanings, in mind. The purpose of the machine part would then be to provide reliably available, appropriately logically structured causal relationships between qualities.

This distinction seems most likely to be correct for sub-symbolic AI systems, but it is also correct (under the assumptions made about the existence of abstracta) for symbolic AI systems, which do not process symbols as meaningful signs, but merely as changes in given qualities (e.g. numbers, or ultimately the flow of charges), quite analogous to Searle's chinese room example. (Mary, on the other hand, works with meaningful symbols, but the one for the unseen color is only formally accessible to her; with Nida-Rümelin, one would have to argue here that the recognition of the concept of a property is to be distinguished from the recognition of the property.) And this would also still be the case for hybrid sub-symbolic/symbolic AI systems, which does not seem entirely unlikely in view of the problems with causal-inference (Pearl) or multi-layered (Le-Cun) hybrid systems: The generation of qualitative information on the basis of quantitative information is generally extremely complex and ultimately the abstraction step that is fundamentally unclear as explained above. (Although the technology can and will certainly provide context-specific solutions, i.e. solutions for selected symbol systems.)

All of this would ultimately also apply to physical models in neuroscience. And as long as embedding in physical and social contexts is understood as no more than the exchange of quantitative information, agent-based, embodied cognition, or similar approaches would not provide a solution either. Our above working hypothesis on abstract entities thus almost automatically throws us back from the 'minimal' problem of stable and broad abstractions to the 'maximal' problem of natural intelligence, which is discussed under the terms consciousness, qualia, intentionality, etc.

To summarize, it should be noted that by assuming the independent existence of abstract entities and an intuitive-mental access to them, we have already turned our backs on the established consensus in technology and neuroscience. Perhaps the most important insight up to this point is that our search for artificial general intelligence leads us to very fundamental philosophical questions.

4.6 Two paths, one goal

If we really want to understand the gap between neuronal activity and mental representation as a special case of the gap between information (in the scientific sense) and meaning, we have to ask ourselves what possibilities are open to us to bridge this gap. The basic problem seems to be so deeply anchored in our scientific understanding of the world that 'Moving forward!' (in the sense of a – possibly also dynamically – structured connectionism) seems to be the only respectable alternative for the scientific community. However, at least two further paths are conceivable from philosophy: First of all, thinking about the structure of the world from the bottom up, a fundamental expansion of our scientific world view would be possible, to grant meanings or at least mental content a matter-independent objective existence, as it is increasingly being done by panpsychists, for example. Or, alternatively, we could retreat to our limited possibilities of recognizing an objective world, as we find it, for example, in the succession of Kant. (Relativism in the sense of simply incompatible cognitive worlds is not considered here as an alternative, since it cannot contribute anything substantial to the actually interesting problem of the interaction of these worlds). [3]

On the first ('metaphysical', more scientifically conceived) path, we would be able to achieve the most consistent renewal of our scientific world view if we first go back completely behind our current state of knowledge and look for the ineluctable building blocks of our models: These would be subjects and qualities of some kind. None of these building blocks seems to be further reducible, because even a natural science that knows how to reduce the subject from the outside entirely to qualities would still have to explain why the subject adopts its subjective perspective – and this quite independently of what exactly constitutes this subjective perspective, be it consciousness,

[3] With regard to relativist, e.g. 'postmodern' positions in philosophy, we can only conclude that these should not be seen as a kind of insight into the functioning of our world, but that the possibility of such essentially skeptical (indeed always self-contradictory) arguments must be understood as observations in need of explanation. With the model developed later on, it is easy to understand why we are necessarily susceptible to sceptical and relativistic arguments; why our 'world map', built against the background of a historically located society, allows for manifold interpretations and thus a myriad of 'grand narratives', which are in addition extremely difficult to separate from the prevailing power relations. The only difference is that in this model, the coupling of our world maps via the causal network of the physical world allows for an inescapable correction of those worldview designs that are less coherent, helpful and/or aesthetically attractive.

qualia, intentionality, etc. (In line with Searle's statement that the ontology of the mental is an irreducible first person ontology. [77]) Starting from nothing more than subjects and qualities, what science has found as regularities of the material world would then either have to emerge or be added. In any case, however, we would have gained the opportunity to completely rethink the relationship between (quantitative) information and meaning. This leads us to dualistic, panpsychistic or (so to speak as the extreme case) objective-idealistic conceptions of the world. The first path would thus require a movement of science towards the subject in order to bridge the gap between information and meaning by modifying our understanding of meaning.

On the second ('epistemological', more philosophically conceived) path, we would have to start with Kant, but we should not ignore the criticism of his idea of a completely autonomous cognitive subject, expressed from very different directions (e.g. by Heidegger, Wittgenstein, Rorty, etc.) and taken up by Dreyfus again. The active role of the subject in cognition focuses on meaning (as opposed to information) as the natural 'currency' of human thought, but the things themselves are not merely passive preconditions ('dispositional structural properties') for cognition, but are actively integrated into subject-independent causal relationships, which binds also the subject itself into a world (through the exchange of information, among other things) and thus makes the continued compatibility of thought and world plausible. This connection between subject and object in cognition calls for a shift from epistemological skepticism towards the objects. In order to bridge the gap between information and meaning here, we would have to underpin a Kantian idealism with a science of objective causal relationships, from which a concept of information could be derived.

Both paths therefore lead us to one goal: To a naturalism without materialism, to a metaphysics that is also inductive and to a scientifically understandable mind that is not limited to physical information processing. In such a model, we should then be able to take into account that natural intelligence functions not only gradually, but substantially differently from our previous and current AI models: Conscious, linked to experiential features, intentional, partly subconscious, implicit, capable of abstraction and creativity to an astonishing degree, but also of ethical action. In the next chapters I will attempt to outline such an alternative model by way of example.

Chapter 5

Objective idealism and other alternatives

So far, we have moved from the question of how data acquires meaning to the question of whether meaning can be understood as a system of logical propositions or relations at all – even if defined implicitly via data. In the last chapter, I then argued that what we can observe about human thinking rather speaks against an understanding of meaning as such a system on the basis of purely physical information processing, but rather in favor of an additional intuitive-mental processing of abstract entities. None of these steps seems logically binding; however, the respective counter-movements appear equally unattractive at best. And also the next step I would like to take here can hardly be described as inevitable, but at least it is likely to bring some movement into a deadlocked discussion.

In response to the question of why human thinking cannot be understood on the basis of logical propositions or relationships alone, one might want to give the classic answer that our world is more than just material reality that can be grasped with logical propositions. Our inability to determine the relationship between quantitative information and meaning would then be inherent already in the materialism that underlies our current scientific world view: Where there are only material processes, non-material meaning must arise from quantitative information, realized via changing constellations of material building blocks. Concepts, for example, are then no more than stable neuronal structures that take into account the regularities between actions and feedback, i.e. they would have to be thought of in purely functional terms. However, if I move away from materialism and recognize meanings as going beyond it, then the question is no longer how meaning is derived from data, but rather how data relates to meaning. This necessarily results in a much more complex picture of human thought. The working hypothesis here is that we have to pay with the symbol grounding problem if we do not want to move away from materialism and that we should therefore investigate alternatives to materialism, especially with regard to the explanation of human thought.

The core of the argument against materialism (with the additional assumption that abstract entities cannot simply supervene on material entities in their existence) would look somehow like this: If materialism is the case, then meaning is also explainable in material terms. But if this is the case, then it can be realized via finite constellations of material building blocks, and if this is the case, then it can be grasped with a finite sum of propositions. (And consequently it can also be realized in the same way via other constellations of material building blocks.) Dreyfus now questions this, but ultimately also materialism. One does not have to agree to this argument necessarily if, for example, one accepts strong emergence as an explanatory device; however, the latter does not seem especially attractive for further model building. (The parallels to the ancient and then medieval discussion of ideas, forms and universals are interesting, especially the extent to which these exist independently, only realized via bodies – ultimately emergent? –, or only as linguistic concepts.)

One is easily tempted here to think that a stringent argument against materialism could be constructed along the lines of the above considerations: That on the basis of either *cogito*- or Gödel-like arguments it could be proven that human thought accomplishes feats that go beyond the possibilities of material systems. For example, that self-knowledge in the *cogito* corresponds to an infinite regress of material information processing, which would have to confirm its own correctness again and again. Or that, for example, the human ability to be able to deduce even unprovable statements in non-contradictory logical systems (more generally; to step out of contexts) lies beyond the possibilities of material information processing systems. The latter idea is derived from Gödel and is known as Penrose-Lucas argument, [78] but suffers from the same problem as the above considerations about the *cogito*, namely that it is assumed in each case that human thought proceeds in the form of a consistent formal system and not, for example, on the basis of much simpler heuristics. The central question here is not whether the respective proof has been carried out logically correctly, but whether the 'metaphysical contact points', i.e. the underlying assumptions that make a translation of metaphysical concepts in formulaic language possible, are correct. More convincing than such arm chair philosophy arguments would be ideas for specific experiments that could be used to demonstrate in detail the claimed superiority of human thought. In the next but one chapter we explore this problem further, but for now we should acknowledge the inadequacy of formal proofs for the pecularity of human thought (as well as the phenomenon of skepticism) as part of the puzzle we face in the philosophy of mind.

Strong emergence would be a way out, as described above, but the unattractive aspect of strong emergence is of course the lack of a comprehensible mechanism that places the material and non-material phenomena in a causal relationship with one another. Emergence would have to be understood as an intrinsic function of material constellations, but in the same way that qualia, for example, elude a functional explanation, they ultimately also elude the emergent variant. Red is difficult to explain purely functionally, even if the function contains an emergence step. Ultimately, the problem is not the connection between the material constellation and the non-material content of consciousness, but the qualitative nature of the content. I will return to this discussion below, as I believe it already implies the main difference between idealist and today's typical panpsychist solutions.

It seems less questionable to me that emergence would help refute the argument against materialism, if it would be able to properly account for qualities. In order to pick out red from the conceivably infinite number of possible conscious qualities, a finite number of sentences is not enough, and if I want to withdraw to a presumably only finite number of colors, for instance, I still have the task of delimiting color impressions as conscious contents from an infinite number of other possible ones, but we already see the realization of practically infinite possibilities on the always equal building blocks in materially conceived information.

Nevertheless, if we continue to follow the line of argumentation outlined above and leave materialism behind us, we must first show that we are still in a position to include the natural sciences in our arguments (which I attempt in this chapter and the excursus in the next one) before we can develop an alternative understanding of human thought (in the chapter after next). For both steps, it is essential to engage in a more detailed definition of the non-materialistic model we have in mind here. Without a model fleshed out in this way, we would have gained an idea of meaning, but we would have left our idea of what constitutes physical information behind us, so that there would still be no bridge between the two. (In any case, even courageous sceptics [79] will not want to deny themselves a pragmatically conceived science. [80, 81] And even without a direct claim to coherence between theory and reality, the aim will be to develop helpful ontologies on the basis of good reasons).

At the end of the last chapter, we already suggested that the design must take at least qualities and subjects into account. In principle, three paths are open to us here: Dualism, panpsychism and idealism. Due to the abundance of existing literature and the diversity of proposed approaches, these three alternatives can only be defined very vaguely here. I will attempt to identify three trains of thought away from materialism as characteristic of the three paths, as outlined in Figure 5.1. (It should go without saying that this will not always be appropriate and will even more often not correspond to the authors' own understanding.) The first train of thought aims for a complete reversal and sees non-material building blocks not only as real, but also as fundamental to the material world. This is referred to here as idealism. The second school of thought recognizes non-material building blocks (or for the sake of simplification also aspects) as real, but not fundamental to the material world and thus standing alongside the material building blocks. This is referred to here as dualism. (Aspect-dualistic models seem less suitable for our purposes, but they play a not insignificant role in the academic literature as a kind of subsystem optimization within philosophy). The third line of thought levels out the difference between material and non-material building blocks; both types come from the same building set. This is referred to here as panpsychism, although it should be noted that this covers a very large field of possible theories: With Meixner [82] we can distinguish dualistic or idealistic, as well as atomistic or holistic variants (but rather as extreme cases), some of which would be categorized as dualism or idealism on the basis of the classification attempted above. (Chalmers makes a similar, but altogether less convincing classification. [83,84]) For the discussion here, I am referring to panpsychisms when I mean positions that operate with the same understanding of space-time, matter and 'weak' subjects as materialism. According to them, mind is formed on the basis of psychophysical laws that are to be understood as additive to the known laws of nature. Panpsychisms that assume space, time, matter or subjects to be emergent (weak or strong) are addressed below under the term idealism.

In my view, this distinction makes sense because the former models, like materialism, are concerned with a 'narrow' mind/matter problem – ultimately how the human brain works –, while the latter ones, like dualism and idealism, aim to solve the 'broad' mind/matter problem, namely how the material and nonmaterial parts of our world can be integrated, i.e. amongst others also how universals can be understood. Here, the objective existence of meaning plays just as important a role as that of strong subjects. (On the historical relationship between idealism and panpsychism, see e.g. Skrbina, [85] on the current view of their relationship e.g. Seager; [86] Brüntrup among others gives an overview of panpsychist theory development. [87])

None of the three possible paths is without obstacles: The dualist encounters the 'interaction problem' between mind and matter, the panpsychist encounters the 'combination problem' (how a human mind could be combined from building blocks, or how it could be detached from a world soul) and the idealist encounters a difficulty that can be described as the 'emanation

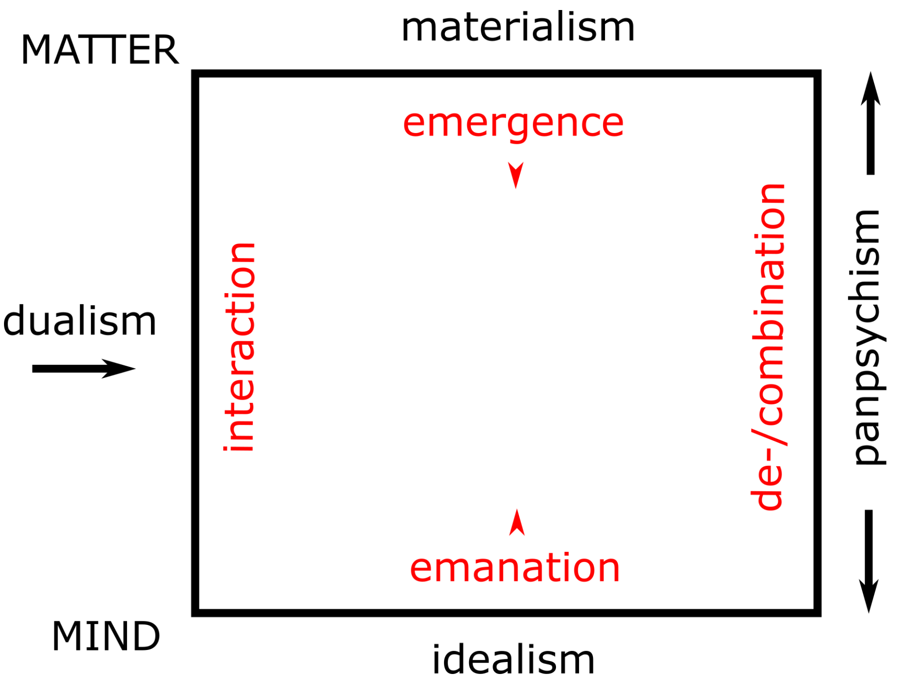

Figure 5.1: Alternatives to materialism and their problems.

problem' of how exactly the material world should be constructed from immaterial building blocks.

In the following, I argue in favor of a scientifically tenable, objective idealism, which in my opinion could also play an interesting role for the natural sciences, e.g. in the interpretation of quantum theory; however, the arguments presented should be helpful for any alternative to materialism. In contrast, dualistic approaches seem to me to be helpful only in their panpsychistic form in order to bridge the mind/matter gap, and panpsychistic approaches are all the more promising for solving the broad mind/matter problem the more idealistic they are. Idealism, as an extreme case of panpsychist theorizing, can thus be understood as a solution to the combination problem; the strong subject of idealism serves as a focus for the disparate contents of the mind. Once the extreme case has been shown to be manageable, the way back is then open again. (No further arguments will be presented against materialism, as these can be found in large numbers in the literature. Meixner, for example, presents the arguments developed so far against materialist/physicalist models with particular commitment, even if not everyone will find his proposed dualism equally attractive. [88])

A central difference between idealism and panpsychism (as it is understood here) has already been briefly mentioned above: Unlike materialism, panpsychism can indeed provide objective qualities that must remain completely inexplicable the former. But like emergence in materialism, the coming together of a mind in such a panpsychism would have to be an intrinsic function (also) of material constellations. (Though in materialism one would need 'super-strong' emergence.) In my opinion such an intrinsic coupling would, on the one hand, require further, purely non-material – but nevertheless causal – laws and, on the other hand, link material constellations and non-material contents of consciousness too rigidly. The paradigm here remains a materially conceived causality. In idealism, on the other hand, material constellations can be understood as starting points on the basis of which subjects then act as flexibly as their evolutionary and historically developed constitution allows. The paradigm here is mental causality. The evolution of psychophysical laws must be imagined in a correspondingly different way: In panpsychism, the result is a weak subject whose mental structure is the result of a material causality that extends into the non-material, whereas in idealism, the subject, which is always already sovereign but initially completely incapable, works on its own mental structure, stitch by stitch growing its freedom, in the happy case all the way to the exit from its (then actually perhaps not quite so) self-inflicted immaturity.

The aim is thus an idealistic model of human thought, which must, however, be based on an idealistic model of emanation. In the remainder of this chapter, I will therefore attempt to outline the basic features of such a model of emanation as a fundamental defense of the extended model of human thought that we will encounter in the next but one chapter. The aim must be the integration of idealism and modern science, including a mathematically consistent reinterpretation of the physical world as a borderline case of a world that appears to be material in some parts, but is essentially non-material.

5.1 A scientifically tenable idealism

Idealism has been in a state of retreat for some time, which now offers few opportunities for a constructive exchange with modern science. (Important counter-movements, however, came from the Marburg neo-Kantians, for example.) In this way, it was and is still able to act as a kind of background story for considerations especially in ethical and aesthetic matters, but only in a very abstract sense for the natural sciences. The idea that idealism clarifies foundations of reality that are in principle inaccessible to other scientific endeavors, so that these endeavors are limited to mere preliminary work on fundamental questions, is an empty assertion if it prevents productive interactions. From the point of view of the natural sciences, such background stories also inevitably have something pseudoreligious about them, and what is even more important: Such a view of idealism is just as incapable as materialism of productively discussing the mind/body or mind/matter problem. A scientifically tenable idealism would therefore be a formulation of idealism that specifies how we are to imagine the emergence of matter from non-matter, so that philosophical and scientific investigations can be combined in order to make predictions about rationally accessible consequences, which in turn make it possible to evaluate the assumptions made. (To have a chance of success, most likely many false proposals will have to go down in history beforehand - so here we go ...)

The test of whether idealism has anything to contribute at this point would then be whether it can help us to draw a more coherent picture of reality, for example in relation to the functioning of our brain, the measurement problem in quantum theory, etc. Chalmers and McQueen have made an excellent (albeit so far unsuccessful) attempt in this direction, in which they experimentally determined whether the measurement problem of quantum mechanics can be explained by means of a certain panpsychistic idea of mind. [89] A direct transfer of this idea to idealistic ventures is unfortunately not possible, since for the panpsychist the measurement problem shows a place for the interaction of consciousness with the physical world, while for the objective idealist it is only a measurement, i.e. the creation of a section of our reality, appears to be material in parts, but is essentially non-material. In order to be successful, however, a scientifically tenable idealism will have to produce similar, above all interdisciplinary ideas.

5.2 The emanation problem

The core problem – and thus the most interesting construction site – of a scientifically tenable idealism certainly has similarities with the interaction problem of dualism and the combination problem of panpsychism, but ultimately differs significantly from them. With recourse to Plotinus and especially Proclus, one could call it the 'problem of emanation', namely how exactly the emergence of matter from non-matter is to be conceived, including the phenomenon of material causality, and in accordance with modern science. It is only surprising at first glance that this problem has not been at the center of attention in the history of idealist thought for a long time, since the overwhelmingly successful unification of the theories of modern physics, which is at the heart of the problem, has only taken place in the last two centuries. And although idealists could still claim that the bridge between mind and matter exists but is in principle not amenable to rational investigation, they would have to counter the argument that this is an effectively dualist position that has little to offer for modern science and the mind/body or mind/matter problem.

In the following, I will formulate my considerations from the standpoint of objective idealism, which assumes the objective existence of non-material building blocks, in contrast to subjective idealism, i.e. the assumption that the world is only the product of one or more interacting minds, since the unavailability of objectively existing entities in the latter approach has no advantages, but poses additional problems for the formulation of our model. (Also not considered are epistemic or 'a-centric' idealisms – an overview including these variants can be found in Tse [90] – as they do not appear to be very helpful for our project here, possibly with the exception of the 'a-centered' approach of Kodaj. [91]) Objective idealism, unlike subjective idealism, is thus a – not only 'Platonic'! – realism. The classification of possible objective idealisms can then be done on the basis of their answers to a series of 'design questions', which I will demonstrate below. Incidentally, this is not entirely out of date; a whole series of collections of essays on the topic of idealism [92–95] show that not only panpsychist approaches such as Goff's [96] are on the rise, but also that the early suggestions of Foster, [97] Springe [98] and others on idealist theory formation are increasingly being taken up again. [99, 100]

5.3 What is the role of the subjects?

As we have seen above, objective idealism assumes mind-independently existing non-material building blocks and at least one agent of some kind as its foundation. (The ancient view that ideas can have dispositions seemed to have largely gone out of fashion, but is currently experiencing a renaissance; more on this later). Different views of objective idealism can therefore be distinguished primarily by the role of the agents. (Neo-)Platonic thought assumes active human agents and needs at least one god-like agent who takes care of the 'maintenance', i.e. the causal upkeep, of the world. Leibniz considers agents across all scales, but their effectively passive nature necessitates a God as caretaker in his system as well. Hegel's absolute idealism dissolves the subject as agent into a sum of non-material elements of a larger world-soul. Holistically conceived idealistic panpsychism shows a certain similarity to this, but positions everything in the space of the material world, while atomistically conceived idealistic panpsychism rejects the idea of a cosmic soul.

This brief outline shows that the central decisions we have to face when designing a scientifically tenable idealism are primarily characterized by the following two questions:

1. Who is responsible for the causal maintenance of the material world? A population of singular agents - which could be of a very simple nature, as 'physical microbes' somewhat like cellular automata, or a god-like mind, or even both? I will speak of 'population', 'god' or 'mixed' theories in the following. Material causality, on the other hand, is not an option, since in 'true' idealism the dispositive forces of matter must arise from the non-material world; material causality is no longer an explanatory instrument, but is itself in need of explanation. In line with this, we also do not observe any 'mechanics' of thought that could propagate into the material world as material mechanics. Additional psychophysical laws could represent an option that should not be disregarded, except that this is then a panpsychist rather than an idealist project, the results of which could nevertheless be transferable to the latter. In contrast to atomistically conceived panpsychist theories, population theories can avoid the combination problem (how human subjectivity emerges from mere 'mind dust') if their agents can act at all scales (simple ones at the micro-scale, more complex ones at our meso-scale), which is possible in idealism as long as space is understood as emergent (more on this below). In contrast, God-theories – just like holistically conceived panpsychist theories – run into the opposite of the combination problem and must now explain why we experience ourselves as singular subjects and not as part of a larger mind. Finally, combined theories, which operate with a population of singular agents and a god-like mind, have no direct argumentative advantages over pure population theories, but they are not unconditionally worse either. (Our experiences as singular subjects are not an argument against an additional larger mind, just as in population theories the agents' experiences at the micro-level do not provide an argument against our meso-scale existence.)

For the design of a scientifically tenable idealism, we should therefore start with a population theory, but we are in principle free to add a God-like spirit later. However, whether we want to take this step or even believe we have to take it will most likely remain a question of faith, i.e. with Alvin Platinga and Thomas Aquinas, whether we are willing to recognize the *sensus divinitatis* as a 'proper basic belief'.

2. If they exist, are singular agents active or passive? Are they just a bundle of non-material building blocks, likesensations, thoughts, and so on, or do they have a unique – albeit possibly in essence very limited – agency? In general, active agents are to be preferred, since they can explain not only subjectivity but also 'real' agency. Such agents are available in population or combined theories. The latter, like God theories, also allow for passive agents, which in turn allows for holistic unification; simpler agents could then be part of a cosmic mind. There is a strong tradition in idealism for such a holistic unification, but from an argumentative point of view it is more of a quasi-religious idea, as it requires further arguments that would have to go well beyond the discussion so far. Quantum holism can be used as an argument here, bearing in mind that it is a purely theoretical concept to begin with, but biological evolution could just as well be seen as an – ultimately stronger – argument for population theories.

To summarize, I think that in idealism, unlike in materialism, there is no strong argument for passive agents, but a very strong one for active agents, namely that humans experience 'real' agency. (While the hard problem of qualia is nowadays mostly accepted as a severe issue, the situation of the question of agency is much less clear, although it seems to me that there is ultimately little that distinguishes this problem from that of qualia.) Building genuine agency on the basis of passive agents seems to be a controversial option to say the least. [101–104] (Linked to this is the question of whether qualities can be assigned dispositions, which is not considered here; such a link already seems too rigid for the physical world, but above all for the mental world, where, moreover, such a link does not correspond to our intuition; we do not experience our thoughts as necessarily causally linked, they do not 'push' themselves into causal chains. However, the model of human thought presented later can be extended quite easily in this direction, as a dualism, in which all physical properties have dispositions, or a panpsychism, in which all properties can have dispositions, and in which the 'core subjects' would then be no more than bundles of properties, too.)

5.4 Space-time or space and time?

If one assumes a non-material, objective existence beyond space, then an existence beyond time is also an implicit characteristic of idealism. Physically measurable time would then arise from the clockwork of the material world, in which the result of the actions of other agents must be awaited for genuine material causation. In contrast, due to their partially non-material nature, living beings could have a very subjective experience of time, which would not always have to correspond to physical time. I examine the question of whether a such a concept of time can give rise to a scientific theory such as the (general) theory of relativity in the excursus in the next chapter.

For the following considerations, I will assume that an integration of the phenomena behind the general theory of relativity is indeed possible, but

I will not adopt the idea of a unified spacetime instead of space and time. Modern science itself has two different notions of time in general relativity and quantum theory, the latter being based on the assumption of universal time, so that relativistic corrections to quantum theory are often made *ad hoc*. The inability to perform 'laboratory experiments' on the cosmic scale makes it advisable to start with a quantum mechanical, i.e. universal, conception of time and assume that the effects of relativity can be transferred to the structuring of space alone. Recent scientific attempts to unify general relativity with quantum theory often follow the same path [105] (more on this in the excursus in the next chapter). Apart from this basic decision, however, the idealist can then be agnostic in most other questions of the philosophy of (physical) time, I think.

5.5 How does space arise?

In contrast to panpsychism, which in its current variants mostly refers to the 'narrow' mind-body problem of integrating consciousness into existing science, idealism is usually conceived as an explanation for the postulated objective existence of non-material entities such as numbers or ethical and aesthetic values, i.e. the 'broad' mind-matter problem of integrating the material with a non-material world. Other than panpsychist theories, a scientifically justifiable idealism can therefore not simply be constructed on the basis of additional non-material building blocks that are positioned in space based on the known constellations of material particles (or particle-like field excitations), because this would leave it unclear how the objective existence of entities such as numbers or values could be located in space.

In response to Chalmer's classification of idealist models, I have argued at greater length that space should be understood as an emergent rather than a fundamental feature of idealist worlds. [84] Beyond this, science seems to tell us that the realization of spatial relations, as opposed to at least some non-spatial relationships, must arise on the resulting micro-scale; in idealism then in the form of a 'grounded' relationalism, which should, however, essentially function like substantial space. A central question already for Leibniz and then Fechner, namely how space can arise from relationships between spaceless entities, must be answered. And again, only mental causation by agents can help the 'true' idealist: Space must be the consequence of the actions of agents that work on certain properties of points in a network of relations, so that the causal function of space emerges. (Below and then in the excurses in the next chapter I make a first suggestion of how this might work). This approach corresponds to physical ideas of emergent space, [106] where we take whatever plays the functional role of space as space. A philosophical formulation of such 'spatial functionalism' can be found, for example, in Chalmers. [107, 108]

5.6 How does material causality arise?

As already outlined above, a second fundamental challenge for the necessary reinterpretation of the natural sciences in the light of a scientifically tenable idealism concerns the role of causality and natural laws. While at least some panpsychists can build their theories on the existing scientific conception of causality and natural laws, the idealist is forced from the outset to make any 'interaction problem' between mind and matter impossible:

In proposing the non-material as fundamental, matter, like space, becomes an emergent feature of the world. This leaves no room for a real problem of interaction between mind and matter, but also excludes material causality: Our laws of nature do not function in the non-material realm (loosely based on Schiller, 'thoughts do not collide in space') and are therefore not fundamental, so that we now have to explain how and why they come about in the material world. If the idealist wants to avoid the scenario of a 'pre-stabilized harmony', i.e. a choreographed change without real interaction that requires no further causal explanation, only mental causation can be considered for an answer. After all, in idealism proper, it is the only known initiator of change in the non-material world and thus also in the emergent material one.

As a result, idealism can certainly invoke material causality and natural laws, but only as a consequence of subject actions in a fundamentally non-material world. This in turn presupposes at least one subject that has the ability to perceive the non-material and to manipulate it. (Which means unlike in the discussion following Kant in German idealism, where the subject ultimately dissolves into the world-soul, it is reclaimed here as a necessary nexus of action, which in turn presupposes the prior perception of the non-material by this very subject.) Linked to this is the problem that idealists, unlike modern science, cannot fall back on a causal development on the basis of chance and natural laws to explain the structure of the actually observed world, especially in the case of the Big Bang theory and biological evolution. Since material causality must emerge from mental causation, the only remaining 'scientific' explanation is an evolution of either a population of subjects or a god-like mind that generates the emanating material world, in line with, for example, Goff's idea of a self-designing universe. [109]

5.7 Which ontological model to choose?

With all this in mind, and leaving behind our usual scientific picture of particles (or particle-like field excitations) in space (or spacetime), the question naturally arises as to how we can then still manage to speak of objects and changes in the world at all. (The choice of 'objects' as referentially fundamental could probably be justified with arguments from Strawson's *Individuals*. [110]) The philosophical tradition has developed two main ways of answering this question: Substance theories and bundle theories.

In substance theory it is postulated that objects are constituted by a substance that bears properties, whereas in bundle theory the object is no more than the bundle of its properties, without a so-called 'bare particular' as a core, by which its essence, its being under change, could be identified. Although the idealist can in principle remain agnostic about the distinction between objects as bundles of non-material building blocks of qualitative nature or as bundles with an additional core, I argue in the excursus in the next chapter that modern physics seems to support the bundle-theoretic view of objects.

One of the main problems of bundle theories is that the so-called 'compresence' relation, which constitutes the bundling of qualities, leads to a series of logical puzzles. However, the idealist is free to accept a para-logical nature of the non-material world as long as he can show that 'material consistency', i.e. the sum of strict rules maintained for the material world, prevents any spillover of 'strangeness' across the mind-matter divide. It is nevertheless commonly assumed that this can still not work out, since in what is probably the most important argument against bundle theories, it can be shown that they make the identification of indistinguishable objects in the material world impossible: [111] If positioning in space is not available as a feature, objects with exactly the same bundle of universal properties become essentially the same object. And if positioning in space is to be used to solve this problem, it remains unclear how the bundling relation can accomplish this without infinite regress, since the same form of linkage seems to be required again and again to make the relations of indistinguishable objects consistent (see Hawthorne/Sider [111] for further explanation).

Although this is usually considered a knockout argument against bundle theories, I will argue in the above-mentioned excursus in the next chapter that this flaw might actually be a core feature of a scientifically tenable objective idealism, since it allows to shed new light on quantum theory. Apart from the bundle-theoretic one, other ontologies are possible in principle, but none seems to fit equally well into the overall project, with the notable exception of North Whitehead's process philosophy perhaps.

5.8 Bundle theories, merology and space

Another problem of the compresence or bundling relation, already discussed between Armstrong [112] and Lewis, [113] still needs to be solved, namely that of structured universals. (A genealogy of the modern concept of universals in analytic philosophy is attempted by MacBride [114], an overview of the problem of structured universals is given by Fisher. [115]) The idea of objects as bundles of properties is based in its naive interpretation on an unclear concept of space and consequently also on an unclear merology. Physical objects are not simple, but spatially structured bundles of properties; a person is a bundle of properties, but also the sum of sub-bundles of properties, at least some of which, such as an arm, are spatially situated.

Physics imposes the additional requirement that the consistent structuring of space must take place at the micro level, as space seems to be fundamentally defined at the level of elementary particles already. This structuring cannot be adequately taken into account with a naive concept of bundling, as the bundling of bundles must also be permitted and a spatial arrangement within bundles must be made possible. Fisher shows that the problem is not a simple one, but that solutions are certainly available. [115] The idealist has the additional constraint (or, if one thinks longer about it, actually the additional possibility) of emergent space, so that bundling, including sub-bundling, can be understood as fundamentally constituted beyond or 'before' space. Spatial situation and thus individuation would then come into play through the inclusion of sub-bundles with spatial properties, i.e. bundles to which the stricter rules of the material world apply. The problem of structured universals is thus closely linked here with the problem of positioning in space, which initially does not appear to be sensibly realizable in bundle theories; more on this in the excursus in the next chapter.

We can look at a person as an example of this: She is a bundle of a mind and a body, partly material and partly non-material. If we look at the 'partial bundle', which is just their arm, we can see other sub-bundles like the skin on their arm and so on. These bundles are largely material, but not necessarily completely so. If we continue to 'unbundle' the bundles, we arrive at more and more materially composed building blocks, but only at the very end do we arrive at bundles that do no longer have any non-material properties at all, but are only defined by their material – i.e. above all spatial – relationships to each other (quite analogous to Plato's basic bodies of geometric nature). However, the fact that we call these relationships material is only due to the fact that they obey certain requirements regarding their consistency in space and time; if we were to disentangle them further, they would simply dissolve into their non-material building blocks.

5.9 The model: A bundle-theoretical view of objective idealism

Based on the above arguments (on holism, space, matter, material causality, causal development and the problem of bundling), the resulting model can be outlined as follows:

1. The world consists of non-material building blocks beyond space (which is not yet formed for our purpose here). These are of qualitative nature, which here means that they are qualia, fundamental concepts, mathematical entities or 'core' values, etc. The exact nature of these classes of building blocks requires further investigation, but this need not be completed here in order to proceed. (From the perspective of the material world, these building blocks can be understood as possibilities that can be actualized in that world as material reality; but in the overall view, non-material 'possibilities' are also actualized as part of abstract entities in the mental world.)

2. Furthermore, the world is inhabited by very simple, non-material subjects who are able to perceive and influence the non-material building blocks. (Be it colors, numbers, ideas, etc.; the concept of ideaesthesia is not entirely inappropriate here.) These subjects are much simpler than what we generally understand by subjects, souls, or similar, which is why they will be called core subjects in the following. [1]

3. Non-material building blocks can be (re-)bundled by core subjects, whereby they can bring new entities, including objects, into the world and change it. The bundling of building blocks then produces new building blocks, e.g. in the form of a red cat as opposed to a 'catty red', whereby the building block 'cat' itself is already 'packed' from simpler building blocks. This packing of qualities will occupy us in Chapter 11, as it means that complex qualities can be analyzed and, for example, concepts must indeed be understood as being integrated into contexts.

4. The core subjects follow very strict rules for the manipulation of certain bundles. The totality of these bundles represents the material world. The rules that are followed lead to the laws of nature. Since space is emergent, core subjects can manipulate the world at very different spatial scales, but their ability to perceive and act depends on what material and non-material 'machinery' is available to assist them in these tasks: Very simple subjects consist only of the core subject and a few acquired non-material

[1] The term 'monad' should be avoided here, because in Leibniz's sense it would imply that each core subject would have access to the whole of reality, even if only from their particular view point; in the proposed model, however, the core subjects can only perceive and manipulate one (super-)bundle at a time.

properties and are responsible for the 'maintenance' of the physical world at the microscale, by means of rules that have been evolutionarily acquired as nonmaterial building blocks and which then underlie our laws of nature. Such agents are not life forms as we know them, but rather like cellular automatons (cellular in the informational, not biological sense), or 'physical microbes'. (In physics, Stephen Wolfram has previously suggested that our natural laws could be explained by the activity of cellular automata. [116]) Subjects can, however, grow into living beings of immense complexity, depending on which non-material and material properties are bundled with the core subject.

5. Not only the biological world, but also the physical one – including space and matter, as well as the gap between mind and matter – developed as a product of the evolution of a population of core subjects. [2] Worlds with untenable rules sort themselves out; if we hadn't been lucky, we wouldn't be here to wonder about it. The material world functions as an anchor of the non-material world, via the creation of identity through positioning in space and the possibility of consistent change and therefore growth through the movement of matter(-properties). More on this below and in the next chapter.

5.10 The model: The material world

In order to make this approach accessible to scientific investigation, we must now specify more precisely how we are to understand space, material objects, properties and forces:

1. Space is understood functionally in the sense that micro-subjects consistently organize the positioning and movement of material objects in relation to each other on the basis of evolutionarily learned rules and 'marker properties'. As in quantum theory, physical time is initially regarded as independent of space and ultimately results from the sequence of actions of the micro-subjects.

2. Objects are bundles of non-material building blocks of qualitative nature. However, when we infer the existence of objects on the basis of rational investigation, we may mistakenly come across pseudo-objects that are irrel-

[2] The idea of interacting agents as the basis of an evolution of the physical world, in which natural laws can then be understood as 'habits' of this world, can also be found, for example, in James Ward, Charles Sanders Pierce and, more recently, Galen Strawson. The problem for all of them, as here too, is to explain the interaction of agents with each other and, above all, within a person (why are we not aware of our neuronal processes?) and then to build a bridge to the natural sciences.

evant to the material world. (Think of Phlogiston, virtual particles, etc., which exist in the idealistic world as nonmaterial 'ideas', but not as functional objects in the material world.) A point in space can be transformed into an elementary particle if the bundling relation is extended beyond spatial relationships to additional properties such as mass, charge, or spin.

3. When we talk about the scientifically relevant properties of objects, we should generally assume that they are fundamental non-material building blocks. As with pseudo-objects, however, when we infer the existence of properties on the basis of rational investigation, we may mistakenly come across pseudo-properties without objective relevance in the material world; we must discuss this in particular with relational properties such as velocity, acceleration, and so on. But at least properties such as mass or charge should be assumed to be non-material qualities of particle bundles, which subsequently function materially.

4. All forces must be understood as pseudo-forces, since material causation is only a consequence of the actions of subjects. Accordingly, properties such as mass or charge cannot themselves be the direct cause of attraction or repulsion. The realization of physical phenomena is due to the action of micro-subjects, which, however, adhere to their evolutionarily acquired rules as to how they act on the bundles surrounding them, so that we can observe the laws of nature as a result: Particles with mass or opposite charge are transported to each other on the micro-scale by subjects, whereby mass or charge are not the cause of their attraction, but provide the framework conditions for micro-scale action; properties do not cause, but enable systematic causation. As with Lego bricks, structures are created according to certain rules, but these rules are not the direct consequence of the properties of the bricks, but of the actions of a subject that acts on them and that takes into account the properties of the bricks, and must do so in order to act successfully. Especially in non-equilibrium cases, the situation on the micro-level is wide open with regard to the exact course of a certain physical process, i.e. for fluctuations. Also the movement of particles occurs via the splitting of existing and the formation of new bundling relations, so that ultimately all interactions arise from such formation or splitting events. Interestingly, such a view of physical forces has parallels with process philosophy: Material effects do not arise directly from properties or substance, but from the process of the agents' actions.

5. The simplest scenario for the structuring of the (sub-)microscale then appears to be a 'game of particles' of micro-subjects in which bundles of particle properties are bundled and unbundled according to the material consistency rules with changing 'marker properties', whereby the latter determine the functionally conceived spatial positioning of the bundles. (More on this

in the next chapter.)

6. The 'evolutionary purpose' of the rules learned by the micro-subjects for the handling of physical properties would be the growth of the structures they are building, but this requires that objects can be individuated, which is not simply a given for a world in which initially only universals are available as building blocks. The central principle behind the rules for the physical world, and as a consequence also behind our laws of nature, would therefore be to enable and maintain the identity of objects. Fundamental physical conditions and, above all, conservation laws would have to be understood against this background, as well as even the symmetries that are related to these laws. Our model would here turn Noether's [117] considerations on their head, in the sense that conservation laws do not follow from symmetries, but symmetries from identity conserving laws. In contrast, the evolution of conceivable alternative worlds without comparable rule sets would stagnate or disintegrate. Worlds without the second law of thermodynamics, for example, would show no biological growth. The stability of the physical world would ultimately be due to the fact that 'successful' individuals in it never die, but also have no offspring. Only on the basis of physical evolution would biological evolution (initially to be explained almost entirely in physical terms) and finally our cultural evolution be possible.

7. In the course of this 'extended evolution', some subjects would become bundles of increasingly complex physical and mental structures; from elementary particles, cells, plants and animals to humans. The transition from the inanimate to the animate, from animals to humans, and towards ever greater freedom of will would be a continuous one. (More on points 6 and 7 can be found in chapter 10.)

5.11 The mind-matter problem; an interim assessment

At this point, we can now return to the mind-matter problem to take stock: In the above model, a person as a complex whole on the meso-scale can act by re-bundling some relationships of the bundle that it is, which also influences the material part of the person and here initially its brain, so that material chains of causality can cascade from there to a desired result in the material world. However, it should be clear that this can only happen to the extent that the framework conditions allow for it, which presupposes the existence of corresponding physical states as well as mental facilities, including a 'rich' subconscious that mediates between the core subject and its body. The brain would first of all have to offer as many different physical states as possible with the same energetic (and entropic?) properties in order to enable as many different material results of mental causation as possible. The following chapters will attempt to shed further light on the implications of the outlined bundle-theoretical view of objective idealism for neuroscience and the philosophy of information, as well as the functioning of minds and brains.

5.12 Scientific questions to be answered

The work of course only begins with the above model. First of all, the idealist must now show that the said model reproduces modern science not only on the conceptual level, but right down to its mathematical machinery. To do this, the following questions must be answered in detail:

1. Does the model allow a mathematical representation of time and space in accordance with the (general) theory of relativity? The first steps towards answering this question are taken in the excursus in the next chapter.

2. Does the model design allow for a mathematically consistent reinterpretation of quantum theory? The first steps towards answering this question are also taken in the excursus in the next chapter. It seems very promising that a suitable mathematical formulation of these ideas is already available with an approach by Lombardi *et al.*

3. Does the model allow the integration of statistical thermodynamics (which still lacks an overarching theoretical framework) and thus the adoption of our scientific concepts of energy, entropy and information?

4. Does the model allow integration with the standard models of particle physics and cosmology? The fine-tuning and arbitrary parameters of the standard models would then be the result of an evolutionary emergence of the cosmos; but can this idea be reconciled with the mathematical machinery of the models?

The outlined research program is in any case a 'moonshot project' and can easily fail at several points, yet in this fact it does not differ in principle from established projects such as the string-theoretical reinterpretation of quantum theory and general relativity. This is ultimately the 'trick' of the chosen approach; by positing subjects and qualities, we now 'only' have to develop suitable reinterpretations of scientific theories, which is a comparatively well-tended field. The most important point will therefore be to make the model as precise as possible and thus accessible to interested scientists; first and foremost with the aim of being able to predict and evaluate the consequences of the new model. Whether the final model will actually allow us to solve open problems in physics, improve our understanding of how the brain works or make the difference between human and machine intelligence clearer remains to be seen. In the following chapters, I will attempt to take the first steps in this direction.

5.13 Preliminary conclusions

In order to remain relevant and to realize its potential for further growth, idealism must be more than one of many possible 'background stories' for science. To do so, the idealist must give a mathematically consistent reinterpretation of the physical world as a limiting case of an essentially non-material world. In this chapter I have made a first attempt at such a reinterpretation, with a model based on a bundle-theoretic view of objective idealism. Any theoretical construct that hopes to shed light on the gap between mind and matter will ultimately be judged by what explanatory opportunities it can offer and how useful it is subsequently for the study of our reality. In this discussion, however, decisions for or against theories will not be made on the basis of individual arguments, but only on the basis of an overall better fit. The hypothesis here is that idealism has the potential not only to bridge the gap, but also to make an important contribution to our understanding of modern science. To illustrate the latter in particular, a first attempt in this direction is made in the next chapter, including a new interpretation of quantum theory based on the above model. With proposals like this, idealism makes itself vulnerable, especially to the scientific refutation of certain parts of the theory. But as in personal relationships, there can be no deeper connection without vulnerability. In this sense, traditional idealism was probably too invulnerable to remain as relevant as it once was.

5.14 Mental and physical objects

Of particular relevance to the model presented here is the distinction between mental and physical objects, which are all bundles of non-material building blocks, but are nevertheless of a fundamentally different nature in practice. To clarify this fact, also as a basis for the following, this should be emphasized once again:

In the model presented, the material part of a green triangle is nothing more than a bundle of atoms; it is not inherently green or related in any way to a perfect triangular shape, although it must roughly occupy a certain pattern of points in space and reflect a certain wavelength. Atoms, on the other hand, consist of bundles of properties such as mass, charge, or spin, for which – unlike for properties such as color or shape – very strict rules of manipulation apply. Ultimately, therefore, the material parts of objects are also nothing more than bundles of non-material, universal properties, which then leads to the 'oddities' of quantum theory.

When causal information about the material part of the green triangle reaches my body and is then processed by my brain, the neural activity in the higher brain regions will evoke a bundle of the properties green and triangle in my mind, as part of my 'world map', i.e. the non-material bundle of universals that is my representation of the world. The rules for associating neural activity with qualia would have been evolutionarily acquired by agents; more on this in the following chapters. My mind is to be thought of neither solely as a core subject, nor solely as a 'map of the world', but as a combination of both.

A 'whole' material object is only ever complete through the combination of a realization in the causal network of the physical world and a representation in my world map. It derives its material functionality from the first fact, but its existence as a separate object from the second. Material objects and our cognitive processes thus always bridge the mind/matter gap.

If I am colorblind, then this chain is interrupted at some point and the property green is not invoked. If the triangle, for instance because it is painted on a saddle) leads to neuronal signals that correspond to a face rather than a triangle, then a face rather than a triangle is evoked in my mind. My mind is therefore prone to errors, illusions, and serious distortions such as schizophrenia, but that is the price we pay for the advantage of using universals in identifying and contextualizing objects, as well as the ability to manipulate purely mental objects independently of material causality.

The model thus largely agrees with modern physics, in so far as for it material objects are no more than collections of elementary particles that are difficult to delimit, and it agrees with modern psychology, in so far as we seem to be working with a mental representation of the physical world that can exhibit extreme deviations from the actual physical world, think for instance of phantom limbs.

Finally, we are now in a position to examine the consequences of the previous considerations for the philosophy of information and, in particular, for human thought, but this will only be undertaken in the next but one chapter. If you are interested in or have doubts about the science sketched out above, you should first go on to the excursus on physical theories for the outlined 'Model A' in the next chapter; if you don't want to do this, you can skip that chapter and then have to make do with my assurance that a meaningful understanding of our basic physical theories (quantum theory and relativity) seems quite possible in this model.

Chapter 6

Excursus: Physical theories for Model A

As explained in the previous chapter, an essential touchstone for Model A is whether a theory of emanation can be formulated on its basis that catches up with modern science not only at the conceptual level, but also with regard to its fundamental mathematical models. Above all, this requires providing comprehensible explanations as to why we find the physical theories that underlie our current scientific world view for the claimed fundamentally non-materially structured world. These are the (general) theory of relativity and quantum theory. The standard models of particle physics and cosmology based on these theories would also in A-world have to be traced back to a partly contingent development; the 'physical evolution'. The relationship to quantum theory and the underlying question of how individuated structures could emerge from universals at all seems even less clear than the relationship to relativity, which is why the first subchapter will examine the problem of quantum theory.

The central 'design principle' or 'evolutionary purpose' of the physical world in Model A is the foundation of identity that allows core subjects to grow their 'projects', i.e. the bundles they perceive and manipulate, in a stable manner: Analogous to the idea of 'spatial functionalism', namely that space is simply whatever plays the functional role(s) of space, [108] in Model A we assume that whatever plays the functional role of matter and material causality is the material world. Concretely, these are then the bundles of universals that are manipulated by microsubjects according to evolutionarily learned consistency rules and only appear functionally physical for this reason. The evolutionary 'purpose' of these rules is the reliable growth of the structures that the subjects build, which requires a (more or less) stable identity of these structures. And this identity is then guaranteed by adherence to the consistency rules for 'material' properties; all growth of individual structures is thus anchored in the material world (and must be).

There is a rich philosophical discussion on the subject of identity and individuation, especially around the concepts of *haecceity*, 'thisness', and of *quidditas*, 'whatness', which have been discussed since the Middle Ages. All of this need not be retraced in detail for what is attempted here; however, the centrally important position of Hawthorne and Sider on the (im)possibility of individuation from universals [111] will have to be discussed further below.

Likewise, the philosophical discussion about natural kinds can be largely ignored, although these 'naturally' play a role in A-world: In the model, one could characterize as natural kinds all objects that are intended in their kind by subjects. This would then primarily include elementary particles, but also centrally controlled organisms and even their artifacts, whose hybrid material/non-material nature we discussed in the previous chapter. The level at which we can speak of centrally controlled organisms (cells, plants, animals, humans, etc.) would be a primarily empirical question, as will be discussed in chapter 10. It would also have to be discussed whether more complex inorganic structures (hadrons, atoms, molecules, etc.) are the result of subject intentions or rather the consequence of material consistency rules on top of them. Here one could still speak of derived natural kinds.

The talk of natural kinds is not without significance for physical theories in Model A: If we grasp a causal connection by recognizing the natural kinds involved in it, this is a substantial gain in understanding, which will normally allow numerous further conclusions. But this is not necessarily the case; due to evolution, development or socialization, a causal connection can also be assigned an idea that is alien to the original intention or consistent development; then, no understanding of the species itself is gained, but only parts of its causal functionalities are subsumed, which allows less powerful conclusions. So our idea of an entity is not necessarily the same as the one the structuring subject has. To take it to the extreme as an example: We do not know whether our idea of a giraffe is the idea that the giraffe has of itself.

So far we have assumed for the (sub-)microscale a 'game of particles' by micro-subjects as the most probable scenario, which is why, at least for the time being, only spatial points and elementary particles, i.e. bundles of 'traveling' properties, will be assumed as natural species – and thus atoms, molecules, and so on as only derivative natural species originating due to consistency rules. After these introductory considerations, we can now turn to quantum theory.

6.1 Quantum theory

The structure and change of matter (including radiation) is described in modern physics by quantum theory (QT) at the most fundamental level in the form of quantum field theory (QFT) [118]), which is another cornerstone of modern science alongside the theory of evolution. It states that our reality can be described by a universal 'wave function' and a series of (quantized, operator-valued) fields in space-time, in which quantized excitations act as elementary particles, from which radiation, atoms, molecules, crystals, amorphous materials, etc. are then composed. The fundamental objects of QFT are thus infinitely extended fields, but their interaction takes place at specific locations in space-time: The field 'quanta' are namely discrete and countable, are carriers of energy and momentum and therefore 'hit each other like particles', as Robert D. Klauber puts it. [119] (Here the concepts of atomism and the originally opposite pole of energeticism, with the assumption of a universal energy field, have largely converged.)

Unlike the earlier theory of quantum mechanics (QM), QFT allows the 'transmutation' of particles, i.e. the mutual transformation of particles or groups of particles into one another. It nevertheless shares with other quantum theories their statistical, non-deterministic nature; like these, it exhibits strong non-local, holistic elements on the (sub-)atomic scale, where interactions can no longer be assigned to clearly delimited locations in space; and finally, it also suffers from the conceptual problem of quantum mechanical measurement (more on this further below).

In order to bring together the seemingly disparate theories of general relativity (GR) and QFT (although one could naively imagine otherwise, they do not fit together mathematically in their current formulation), a number of even more fundamental theories have already been proposed, such as string theories, which assume again much smaller vibrating 'strings' as the basic building blocks of the physical world. [120] Parts of the scientific community doubt, however, whether we will ever be able to evaluate any of these new theories as right or wrong, if only because of the extremely high energies that would be required for the corresponding experiments.

Such efforts towards a 'theory of everything' nevertheless appear very attractive from the point of view of physics, since in the past it has repeatedly been possible with great – and above all practical! – success to bring seemingly disparate parts of physics under uniform mathematical descriptions, e.g. the merging of theories of falling bodies on earth with those for the motion of celestial bodies; of mechanical energy and heat; of magnetism, electricity and light in the form of electromagnetism; as well as space and time, energy and mass, right up to our current models of GR and QFT.

After the Higgs boson, theoretically predicted on the basis of the Standard Model, was actually found experimentally, the ongoing work on QFT and the Standard Model of particle physics has reached a state in which our ideas appear so flawless that the lack of stumbling blocks is perceived as downright an obstacle to finding explanations, for example, for why the postulated 'particle zoo' appears so arbitrary. [121] As a result, we hear with a certain regularity about experimental indications of 'new physics', which have so far always turned out to be measurement errors, inaccuracies, etc.

We will examine the topic of 'grand unification' further in the third subchapter; at this point we can state that if we want to catch up with the current state of modern physics at the (sub-)microscale, we should attempt a reconstruction of quantum theory (and not its alternatives) with Model A. However, the ontological assumption of universal fields should rather be seen as a necessary tool for a mathematical summary of experimental observations; we can discard it if we can realize the basic properties of the theory without fields; we need a quantum theory, but not necessarily a quantum field theory. (Fields themselves can of course be realized in Model A; for instance as properties of – however realized – spatial points, for whose propagation certain consistency rules apply, or also via non-spatially mediated relations between elementary particles). In this sense, a new interpretation of quantum theory will be outlined below, based on a bundle-theoretical view of objective idealism; i.e. for Model A, but in principle also for other models that share the basic assumption that reality is made up of bundles of objectively-existing, non-material universals.

6.1.1 Open questions in the interpretation of quantum theory

Even a century after Bohr's first experiments, the interpretation of quantum theory is still a field with many open questions. [122] The following interpretation assumes that the 'strangeness' of quantum theory can be understood as a consequence of a vanishing distinguishability of (according to their properties) indiscernible particles and makes use of the observation that a similar vanishing distinguishability is also found for bundle theories in philosophical ontology.

In quantum theory, when a system is formed from parts, e.g. a molecule from elementary particles, the individual parts are incorporated into the overall system in such a way that they lose their independent identity; in philosophy this is called 'quantum non-individuality'. As a result, the system requires a global description, for instance via a wave function and its evolution, until a measurement is made that irreversibly collapses this construction and leads to statistically distributed and quantized (i.e. not continuously distributed) results. Although the correctness and usefulness of quantum theory is beyond doubt, most scientists agree that it is still rather unclear how the basic assumptions of this theory can be reconciled with our current, still essentially materialistic, view of the world. [123]

The open questions can be divided into four categories, with the first category dealing with the fundamental observations of quantized, particle-like interactions and the possibility of transmutation, i.e. the transformation of elementary particles into one another. [124]

The second category deals with questions about the 'holistic' nature of quantum systems: Why is there no identity of indiscernible particles at the micro-scale, but non-localization according to Bell's inequality [125] and entanglement, i.e. the non-local coupling of properties? Why is there complementarity and Heisenberg uncertainty, i.e. that certain properties cannot be measured simultaneously with arbitrary precision? The latter can be summarized under the concept of contextuality, [126] whereby non-locality can then also be understood as part of the same problem. [127]

The third category deals with what is probably the most central question: Why does the act of measurement play a special role in an objective scientific theory? And why is the measurement not simply a revelation of already existing values, but depends on the context of the measurement? [128] Finally, why are the results statistical in nature?

In the fourth category we can collect all the more 'technical' questions: Why does the mathematical machinery need complex numbers? Where do the parameters of the standard model come from? And so on.

Using Model A, we could now argue that the vanishing distinctness of indiscernible parts is at the heart of quantum theory: Consider a system in which the properties of the indistinguishable parts are actually absorbed by the system until the parts are forced to separate again, causing the system to lose the properties that must be assigned to allow the re-formation of distinguishable parts. In philosophy, not only Morganti has argued similarly that we should regard quantum-theoretical properties as belonging to the whole system. [129] Such a system would clearly show non-locality and could exhibit entanglement, since the assignment of properties is system-wide. Depending on how properties are taken up and released by the system, it could also show complementarity and uncertainty if, for example, it is not possible for the system to update certain properties at exactly the same time.

Finally, the measurement process could be understood as a, indeed context-dependent, way to enforce a distinction at the (sub-)micro level, but the enforced distinction would not be bound to a human observer but to a given physical context, eliminating the need for a special role of the observer. Even the more fundamental observations of quantization (properties come and go 'as a whole'), transmutation (only the total balance of properties counts) and indeterminism (properties of the parts are initially 'lost' to the system and then manifest themselves statistically in the case of renewed part formation) could be understood in this way.

The proposed inclusion of parts in a whole implies the at least temporary disappearance of their substance, though not of all their properties, from the material world, which would represent a major break with our everyday worldview in which the substance of objects, rather than just some of their properties, is assumed to be materially persistent. However, it is a much smaller break with the ephemeral notion of matter in QFT. It is our *materialistically conceived* everyday worldview against which quantum theory appears 'strange' to us, and which Model A may allow us to overcome.

6.1.2 An interpretation of quantum theory based on a bundle-theoretical view of objective idealism

But why should Model A imply a vanishing distinguishability of (according to their properties) indiscernible particles? As a bundle-theoretic version of objective idealism, Model A assumes that reality is composed of bundles of objectively existing, non-material universals. It is precisely for such bundle theories of objects, however, that philosophical ontology finds a vanishing distinctness of the indiscernible: Objects that are equal in terms of their universal properties are not individuated from one another, i.e. they are the same object. Even a spatial relational positioning to each other does not solve this problem as long as the object positions are not characterized in some external way, e.g. by the existence of a substantial space, since otherwise the relations would again require differentiation. As discussed in the previous chapter, however, for A-world we actually want to assume that space is not substantial but emergent, so that it would itself have to be made up of bundles of universals or relations between such bundles, which then throws us back to the above problem. Since Hawthorne and Sider [111] at the latest, objections of this kind have been considered a knockout argument against bundle theories, but here it will be argued that the lack of distinguishability is not an insurmountable problem, but part of the solution to another problem, namely that of the intelligibility of quantum theory.

So to argue with Model A for a 'physical evolution' of the material world, we would first have to explain the origin of space and then elementary particles, or alternatively of elementary particles with spatial relations to each other. But Hawthorn and Sider have shown that spatial relations between bundles of universals cannot simply be thought of as external relations between these bundles. Could then physical evolution have begun with the emergence of a network of spatial points, i.e. an effectively substantial space? Let us first assume that we wanted to establish a lattice of spatial points that could be built up by the subjects, with nothing but very simple properties shared by the spatial points in pairs and thus practically corresponding to 'next' markers. These spatial points would be indistinguishable, but all participating micro-subjects would now 'work' on the resulting singular spatial point. Since, according to our theory, the subjects are concerned with the individuation of their products, they maintain the multiplication of properties for this point in space: If this piece of space is then to be traversed by an elementary particle, i.e. a bundle of properties is to be passed on, it does not have to work only one time through our one 'next' marker, but micro-subject by micro-subject again and again. In order to objectify this between the micro-subjects, genuine 'double markers', 'triple markers', etc. would have to be assigned to the spatial point, at least if we do not want to believe that micro-subjects could already have understood the concept of counting. In this way, what has already been considered in the previous section is realized: Space, like material causality, not as a disposition of substance or property, but as a consequence of the rules according to which micro-subjects work with (in themselves disposition-free) properties. As a result, the extended arises from the non-extended; but neither as a continuum nor as a 'granular' space, but as a network of points that merge into a continuum if they are not individuated by the assignment of a bundle of elementary particle properties.

The problem with this approach is that the high symmetry of the state of the world at the beginning of the production of spatial points immediately requires the production of further properties that characterize the respective new points, which would then already have to be understood as particle properties. This would most likely still result in very many symmetrical states in which the 'individual' points would immediately coincide with other points, and even in a much later state of the world this could easily occur again and again. The only positive aspect is that the merger of several particles, i.e. bundles of properties that are bundled with a point in space in the form of 'next markers', would show non-local effects. (In addition to parallels with Spinoza's notion of the effective motion of properties in an all-encompassing whole, [130] there are also parallels here to the fundamental notion of a dynamic space in 'loop quantum gravity' theory as perhaps the most important alternative to string theories. [131]) In any case, Hawthorne and Sider are not wrong that the positioning of bundles of universals is not easily possible; micro-subjects and evolutionarily learned rules must be added, and even then the situation cannot apparently be solved satisfactorily as envisaged above.

The following approach appears more promising: In line with the already mentioned spatial functionalism, [107,108] we could assign one or more further universals to individual objects, which would allow the realization of a 'grounded-relational' positioning instead of a positioning on 'substantial' spatial points as above or the purely relational positioning rejected by Hawthorne/Sider. This would be possible on the basis of properties that come across as a spectrum; imagine a color with many different shades. Let us now assume that there are in fact three such spectrum properties on the basis of which the functionality of our three-dimensional space is realized: Each object would always be assigned exactly one shade of each of the three 'colors out of space'. [132] (The fact that we find three spatial dimensions realized would be the result of an evolutionary optimization process; we need at least three dimensions for the implementation of arbitrary relation graphs. The possible existence of higher dimensions not perceived by us would then have to be explained with additional evolutionary benefits).

A movement would now mean that the object would be transported a number of 'shades' corresponding to the time interval and its speed on the respective 'color scale', so that the property bundle that is the object would lose the property of the initial shade, but would gain the property of the final shade. The resulting positioning would be relational, because only the relative position of the objects on the color scale would play a functional role, i.e. there is no substance that is space, but these would be 'grounded' relations, because there is a fixed reference that is principally inaccessible, but nevertheless exists, so that we could do without external relations between objects. [133] (What must be added in the sense of modern physics is a relativistic metric.)

In any case, even this approach cannot do without subjects and rules; the mere 'positioning' on the color scales does not yet have any spatial functionality; only when subjects begin to manipulate objects according to the above rules for using the color scales do they suddenly behave 'spatially' and thus individuated and at least to some extent already physically. The designation of the three spectrum properties as colors is of course a purely metaphorical one; we only perceive the effects generated on their basis, and this only after a translation into the qualia, i.e. the possibly completely different properties, that convey spatiality to us.

In this approach, elementary particles would be individuated if their bundles differ in that they are assigned to different points in the coordinate system of the color scales, i.e. in that their bundles differ by at least one shade of color. On the basis of this individuation of simple objects through functional-spatial positioning there are various conceivable growth processes. In principle, we would now have to try to systematically list these possibilities and consider why they do or do not play a role in our actual physical world. More generally, we should probably expect a certain balance to be struck between individuation and interaction: More individuation is easiest to achieve by indiscriminately adding more and more properties; individual growth is thus easy to achieve, but the objects that have grown in this way still only participate to a very small extent in the shared, material world. More complex objects will occupy a larger space in this world, if they are created from simpler building blocks through mutual interactions, for which the user of fewer properties is advantageous, since they are meant to be shared by as many objects as possible. The physical world we observe suggests that collaborative growth has overtaken individual growth at some point. Our actual physical world would then appear to be a sensible compromise; as individually structured as necessary, but as universally structured as possible: A few basic physical properties underlie the existence of elementary particles, whose interaction then gives rise to the entire complexity of the material world. Already here, both individual and collaborative rules for the actions of (micro-)subjects would have developed in an evolutionary way. That micro-subjects could not simply create more complex entities directly at the lowest level would ultimately be the result of the practical non-existence of complex mental structures at this level; micro-subjects could only adopt the simplest rules of action and therefore could simply not want anything more complex at all. (We would really have to imagine them as cellular automata and not as living beings; Model A is not an animism.)

The quantum nature of structures on the (sub-)microscale would then come from the fact that as soon as structures interact, micro-subjects try to integrate them into 'their' bundles as sub-bundles; parts become 'entangled'. All properties are absorbed into the new overall bundle, which is now 'managed' by more than one subject. And this includes the spatial properties of all parts, so that we obtain a non-local overall system. Particularly over short distances, moving particles could easily lose their individuality and become part of a larger quantum system, for which multiple properties like charges, spin, and so one would then have to be maintained globally by the subjects. (Mathematically speaking, we would have quasi-sets instead of sets of properties.) By 'balancing' the spatial properties as well, quantum systems would subsequently acquire non-local spatial structures. Once 'captured', they would not automatically individuate again. Only circumstances brought to the system, such as a measurement or limiting structures, i.e. generally a symmetry break, could cause this. Every interaction with other systems, every scattering, would give the subjects the opportunity to break out with 'their' bundles of properties from the overall system. Finally, complementary properties would be those whose actualization in the material world would require actions of the micro-subjects that cannot be performed simultaneously, e.g. moving the system and letting it interact locally, or possibly also adjusting the position on different color scales at exactly the same time. (Further considerations would certainly also be needed here on the 'weak' complementarity of energy and time. [134])

Partial bundles that are indiscernible in terms of their properties, such as elementary particles, would thus be indistinguishable within larger bundles, such as molecular systems, and would 'pass on' their properties to the overall system; just as we find for a physical system that must be described by a wave function. The only difference is that in the world view of objective idealism, the temporary disappearance of entities at the micro level would not be problematic because the resulting bundle can retain the constituent properties without having to be constituted by separate sub-bundles. A wave function could then be understood as an inventory of the materially relevant properties of such a bundle. And since dispositions do not lie in the properties but in the micro-subject agency, Model A could also make practically inconceivable properties such as spin understandable: It does not need a physically plausible representation as 'quasi-rotation' or the like to understand its role as a 'marker' for the rule-based action of microsubjects.

6.1.3 Comparison with current debates: Quantum reconstruction and *haecceitas* in classical statistical mechanics.

The last two decades have seen a great deal of interest in the project of 'quantum reconstruction' (QR), i.e. the derivation of quantum theory from the simplest conceivable assumptions, so that it is possible for us here to compare the conclusions of Model A with the findings in this field: Interest in QR was sparked primarily by Hardy's work, [135] who argued that the core of quantum theory is its inherently probabilistic nature, and then showed that under this assumption the simplest possible theory is our quantum mechanics. More recently, Masanes, Galley and Müller [136] found that, starting from the assumptions we must make to consider the case of measuring unique values from unitary, i.e. uniformly evolving, quantum states, we automatically arrive at Born's rule, which links the mathematical mechanism of quantum theory to the interpretations of its results. Cabello [137] proposed that there is no underlying physical law for measurement results, but only a set of consistency requirements that must be met; Born's rule is then ultimately the result of these requirements. Recent work on quantum reconstruction thus seems to be in agreement with Model A, insofar as quantum theory is understood as fundamentally probabilistic (parts manifest themselves in averaging over all micro-subject actions in a practically chance-based manner), whereby Born's rule is derived from the fact that certain consistency rules are observed (the overall accounting of the properties in the bundle is correct).

Another interesting debate is that of *haecceitas* in classical statistical mechanics. Unlike quantum-theoretical approaches, 'classical' statistical mechanics usually assumes the distinguishability of particles with the same properties. Such classical approaches play a very important role in the simulation of physical, chemical and biological systems from the atomic scale upwards: The atoms of these systems are often regarded as classical particles, whereby the forces between them, which can actually only be described correctly by quantum mechanics, are approximated classically using various theoretical constructions and often with the help of empirical input. Alternatively, density-based theories can be formulated, with which the difference between the assumption of distinguishable and indistinguishable particles can then be investigated; [138] if only the density is considered, then neither distinguishable nor indistinguishable particles can freely penetrate each other, but only distinguishable particles will 'get stuck' in certain environmental structures, as is actually observed experimentally in the formation of glass, for example.

If we take these results at face value (there are, of course, a number of ways to avoid this), then they would mean that somewhere on the micro-scale we find a transition from indistinguishability to distinguishability. This would be consistent with Model A, insofar as objects gain their identity only at the level of molecular structures, since here the systems are forced, on the basis of the statistical relations, i.e. consistency rules between their physical properties, and through interaction with their environment, to exhibit the effects that make them appear to us as stable material structures: The continued 'flashing' of material properties across the entire system would result in the structure of the system. This manifestation of physical properties would not require quantum mechanical measurement by an observer, but would always occur when systems would be forced to do so by circumstances, e.g. in the formation of glass. Here, as there, distinguishability is seen as an emergent rather than fundamental property of the world.

6.1.4 Connection of Model A to an existing interpretation and mathematical formulation of quantum theory

The above explanations do not yet offer any direct connection to quantum theory, which is primarily understood as its mathematical formulation. The most sensible next step towards such a connection would be to relate what has been said so far to an interpretation of quantum theory, which for its part has already formulated the connection to the mathematic apparatus. In principle, a whole series of elaborated interpretations in the literature could be used for this purpose. However, the modal hamiltonian interpretation (MHI) developed by Lombardi *et al.* appears to be particularly suitable. [139] Modal interpretations go back to van Fraassen, who proposed that it is not a 'collapse' of the wave function that determines which properties of the system are updated, e.g. in a measurement, but that this is determined by the respective physical context, as 'modal' condition. In its wake, various 'actualization rules' were proposed to specify which properties are actualized, i.e. in principle also measurable, in given physical contexts. Unlike a whole series of older proposals, the model suggested by Lombardi *et al.* has so far been able to hold its own.

All MHI approaches have a number of basic assumptions in common: Quantum theory describes individual systems (no relations or the like); measurements are normal physical processes (without 'collapse' of the wave function); and quantum systems are represented by operators (roughly; calculation rules) that stand for observables (properties that can be observed). The model of Lombardi *et al.* is then characterized by its actualization rule, namely that the observables/properties actualized in each case are those of the Hamilton operator (standing for the energy), as well as of all operators that commutate with the Hamilton operator (i.e. stand in a certain mathematical relation to it) and which have the same or a higher symmetry. A measurement or physical interaction can then be understood as symmetry breaking that forces the actualization/'measurability' of specific properties. Lombardi *et al.* were able to show the meaningfulness of this rule in a whole series of applications, were able to use it also in a relativistic context and illustrate its applicability in the case of repeated measurements, for whose identical results normally the wave function collapse is used as an explanation. [140] (A more detailed first overview of MHI approaches, including the one by Lombardi and co-workers, can be found on the corresponding pages of the Stanford Encyclopedia of Philosophy.)

What is interesting for our considerations here is that Lombardi describes quantum systems in her interpretation as bundles of properties whose statistical relationships are described by the respective quantum state, i.e. not as any kind of 'set' of substantial elementary particles. [141] Lombardi herself no longer sees any principle of individuation here, but from the point of view of Model A one would have to object that the bundling relation assumes such a function. (Quite in the sense of Falkenburg, who describes elementary particles as a 'metaphysically glued' bundle of physical properties in her overview of 'particle metaphysics'. [142]) In any case, we have a possibility here to find a connection to an established interpretation of quantum mechanics for Model A: Lombardi's approach describes the bundle of physical properties of elementary particles; in Model A, however, the bundle would in addition have a non-material part of simple mental building blocks, namely the rules according to which the micro-subject(s) act, when perceiving and manipulating the whole bundle.

If one wanted to take the thread even further, then it would indeed be conceivable that the combination outlined above could solve a still open problem of quantum theory, namely that of molecular symmetry: While the Hamiltonian operator describing molecular systems is always 'fully' symmetric to begin with, this is not the case for a whole series of molecules, which is why quantum theory does not seem to specify which of, for example, two chiral, i.e. 'handed' or mirror-symmetric, molecules is actually described by the corresponding Hamilton operator. In practice, the so-called Born-Oppenheimer approximation is usually used here, in which the structure of the molecule is simply given by the pre-positioning of the 'classically-mechanically' understood atomic nuclei, and only the electrons, which are then distributed over this 'framework', are treated as a quantum system. The meaningfulness of this approximation can now also be demonstrated in simulations, from which the assumed molecular structure can be seen as a manifestation of strong statistical correlations between the positions of the atomic nuclei. [143] The problem of molecular structure then appears to be the problem of quantum mechanical measurement, namely that the structure is only unclear to the extent that it only emerges in the measurement or interaction. Lombardi *et al.* point out, however, that this is not entirely correct, as we are still not able to unequivocally distinguish between certain symmetrical cases like the one mentioned above. [144, 145]

In the above MHI approach there is now the possibility that further properties are added to determine which of several symmetric states are realized in the specific molecule, [146] although it remains unclear in Lombardi *et al.* whether this property should result from the respective context (why is it then stable over time?) or whether it is really an additional property that does not simply result from the properties of the parts (unlike, for example, the total charge) or is determined by the coming together of the parts as codified in the Hamilton operator. Here, Model A offers a way out: It does not seem unlikely (if one is willing to follow the previous argumentation) that the micro-subjects, who are fixated on the most consistent individuation possible, have acquired rules for the reliable formation of indeterminate (or rather underdetermined) structures, by including additional properties in their bundle that fix the resulting objects to one possibility. Model A could thus solve two problems of the reduction from macro- to micro-processes: The emergence of particle identities and of irreversibility would both be due to the rule-governed action of micro-subjects. [1]

To summarize, however, one should be much more cautious at this point in formulating that the possibility of connecting Model A to an existing interpretation of quantum theory and its mathematical formulation certainly seems to exist.

6.2 Relativity

Our universe emerged from nothing around 14 billion years ago and has been expanding ever since without a center or boundary at an increasing rate. We do not know how large it currently is, but the observable part has a diameter of around 100 billion light years. The universe as a whole appears to be very homogeneous and isotropic (not directionally structured) and shows interesting features such as black holes (singularities in space-time), as well as gravitational waves (ripples in the fabric of space-time itself). To a first approximation it is simply empty and only a tiny part of it is filled with 70% 'dark' energy, 25% 'dark' and about 5% 'normal' matter, the latter gravitationally bound in the form of hundreds of billions of galaxies, each with hundreds of billions of stars, but also gas, as well as diffuse radiation.

We know very little about the 'dark' components, the existence of dark

[1] Drossel, [147] shows that we must assume that the second law of thermodynamics cannot be completely traced back to the underlying laws of quantum theory, since it requires at least an additional 'rule of equal probabilities'. Thanks to this rule, systems then behave as required 'typically', with the result that the most probable development is indeed the case to be observed. The emergence of such additional laws beyond the micro-scale would generally be unproblematic in Model A, as already envisaged above for molecular symmetry. Ultimately as a consequence of the idealistic approach of a purely functional space, micro-subjects would act on different length scales and could thus have acquired evolutionary rules which could be found as laws of nature emerging on higher levels. In this sense, the model also offers a 'deflationary' explanation of 'downward causation': Once an overarching bundle structure has been formed on the basis of fundamental processes, it can subsequently serve as a target value for feedback loops and thus act 'downwards'.

energy being suggested by the fact that the observed expansion of the whole requires a driving force, and that of dark matter by the observation that more than just the known bodies appear to exert gravitational forces on galaxies and stars. (Ultimately, however, these hypothetical components are probably just an expression of our ignorance of the real processes). The number of planets in the universe is most probably of the same order of magnitude as the number of stars, which – given the large number of planets and the age of the universe – raises the question of whether we should not already have seen signs of other forms of life in the universe (a question known as the Fermi paradox). Despite the abundance of phenomena in our universe, researchers have found a very small number of rules for their physical-mathematical description, consisting of not much more than a list of about two dozen parameters (the exact number depends on how one wants to count) and four fundamental forces (three of which are described by quantum theory, leaving gravity).

Combining our knowledge of the night sky into a consistent theoretical model is undoubtedly one of the greatest intellectual achievements of mankind, from antiquity, Copernicus, Kepler, Newton and many more right up to modern physics. Our cosmological worldview needs little more than space and time, matter including radiation, and a recipe for how gravity, the yet-to-be-explained one of the four fundamental forces of our universe, acts on the other ingredients. (The situation becomes more complicated when we look at the early history of our universe; in the first phases of its expansion, even the fundamental forces were 'unified' and only later did subatomic particles form and then atoms, etc.).

According to classical understanding, gravity is a long-range effect, depending on the mass of the bodies involved and the square of the distance between them. Our current best theory to describe gravity is General Relativity (GR), which describes gravity as a distortion of 'spacetime': It explains the observed gravitational force on objects by the deformation of spacetime itself by other such objects, resulting in the actual observable change in motion near massive objects and at very high speeds. [148]

Not so much with regard to GR, but with regard to our standard model of cosmology, there are certainly a number of open questions. [149] In addition, cosmology is not accessible to experimental verification to the same extent as particle physics, where time and length scales allow direct experimentation; much must therefore necessarily be more speculative. Ultimately, the question must be how the remarkable accuracy of GR's experimentally verified predictions fits in with the fact that we know practically nothing about 95% of the energy and matter in the universe. Regardless of the cosmological uncertainties, however, GR is basically undisputed, should accordingly also be seen as a cornerstone of our scientific world view and must therefore be integrated in some way by Model A. (Important successes of GR were the passing of the 'classical' tests proposed by Einstein himself on details of Mercury's orbit, the diffraction of light by massive objects such as the sun and the gravitational redshift of light waves, but also numerous other tests including the prediction and subsequent experimental confirmation of phenomena such as gravitational time dilation and gravitational waves).

GR was the main work of Einstein, which was based on his earlier discovery of the theory of Special Relativity (SR). The latter redefined our thinking about time and space. Based on the assumptions that the laws of physics should not depend on the state of the observer and the experimentally supported assumption that the speed of light is constant even for moving light sources or observers, he stated that time and space must be understood as coupled, as a uniform 'spacetime': If I move a light source, but the speed of light remains constant, i.e. the two speeds do not add up, space and time must somehow take this into account. In order to restore consistency for both the traveler and the observer, the traveler and the observer must experience length contraction and time dilation relative to each other. As a result, 'simultaneity' becomes dependent on relative movement. By correctly taking the movement of light into account, Einstein was able to unite classical mechanics and electromagnetism.

Only after the development of GR was SR described as 'special', since it did not take into account the 'curving' effect of mass on spacetime and thus only 'special' cases with 'flat' spacetime could be treated. In thinking about what happens with not only moving but also accelerating objects, Einstein realized that one cannot tell whether one is falling free of forces or under the effect of a gravitational force; that gravity, like centrifugal force, could therefore be a pseudo-force that appears to act on us while we are actually moving in a curved space; which would then also explain the equivalence of inertial and heavy mass. The mathematical core of GR, Einstein's field equations, relate the geometry of spacetime to the distribution of matter in it. One (not unchallenged) consequence of the concept of spacetime is the resulting idea of a 'block universe'; an unchanging spacetime 'block', which is at odds with our common concept of time, which Einstein therefore described as a 'stubbornly persistent illusion'.

6.2.1 Relativity and Model A

In order to begin a reconstruction of the 'phenomenological content' of SR and GR, I will first try to show that Model A agrees with the basic assumptions of these theories. The ultimate goal, which clearly is a long way off here, would be to derive a largely equivalent alternative to Einstein's field equations from the model. One problem with this is that time in Model A is initially only defined on a relative basis via the sequence of subject actions, which are nevertheless to a large extent interdependent. (Physical time would then ultimately have to result from quantum events. [150]) As already noted in Chapter 5, for A-world, we will therefore have to move away from the concept of a uniform spacetime and propagate a locally classical understanding of time, as is also common in quantum theory, which only receives the appearance of a coupling to space through the finite speed of light. However, this is not a problem in principle, as some modern alternatives to SR and GR do not proceed totally differently. [105] Reichenbach had pointed out early on that SR 'chooses' certain consequences for the structure of space and time through the experimentally supported assumption of the speed of light as a constant, but that an alternative distribution of the consequences of the basic phenomenon of coupling would also be conceivable. [151] With Model A, one can then add here that the nature of the coupling itself can also be understood differently.

Apart from this, SR/GR and Model A agree on the following basic assumptions: First, both SR/GR and Model A formulate a relational view of space and time; time and space are not absolutely measurable, but only as relations between events. Secondly, in both approaches time and space are coupled and consequently relative to motion. Thirdly, we can understand the speed of light as a finite conversion factor for this coupling. With regard to GR, we must add a fourth point: Space and time are also coupled to the distribution of matter. However, it is not obvious that these points are fulfilled in Model A; we must therefore go into more detail now.

In Model A, positioning in space via elementary particles and spatial points (on the 'color scales') alone is not sufficient for the foundation of identity as the evolutionary purpose of the physical world. A 'causal positioning' through the (semi-)local limitation of interaction possibilities must also be added. If otherwise instantaneous interactions across the entire universe would be permitted, positioning in space could hardly fulfill its purpose. One conceivable limitation – and with Model A the result of the physical evolution of our universe – is a consistency rule that specifies a maximum speed of interaction, i.e. the speed of light. (The exact value of which would only be significant in conjunction with the other fundamental constants of physical evolution.)

However, once we have introduced our speed limit, a subsequent problem arises: Without further modifications, processes then run differently in moving systems than in stationary ones (i.e. not moving relative to the network of spatial points); the identity of these processes would no longer be preserved across a change of system state. The successful creation of identity therefore requires that the physical laws are not 'processed' relative to the coordinate system of the 'color scales', but relative to the movement; this requires a modified balancing of dynamic properties. In combination with the finite speed of light, the first to third of the basic assumptions listed above are thus taken into account. Together with the assumption of a time resulting from the succession of micro-subject actions, we thus obtain the observed relativistic effects; not because there is a uniform space-time, but because the consistency rules acquired by the micro-subjects in physical evolution would result in the Lorentz equations, which give a mathematical description of the observed effects.

But even the 'causal positioning' is not yet sufficient for a fully consistent identity foundation, which brings us to GR: Just as the finite speed of light is central to SR, the equivalence of inertial and gravitational mass, as a result of which all bodies fall equally in a gravitational field and the free-falling body cannot distinguish its motion from an uniform one, is central to GR. (It should not go unmentioned that there is still an extensive discussion in physics about the equivalence principle(s).) The fact that this equivalence appears to exist – it has now been experimentally confirmed to the Xth decimal place – is seen by Einstein as an indication that gravity must be understood as the curvature of space-time by matter, i.e. as an essentially purely geometric phenomenon. With Samaroo [152], however, equivalence can also be understood (following Frege and then Demopoulos in mathematics) as an important criterion of identity: Free-falling coordinate systems are only equivalent to Lorentz coordinate systems if all non-gravitational experiments yield the same results, i.e. if the equivalence principle holds.

With Model A, we would now turn this consideration on its head and assume that the micro-subjects fall back on the (then exactly) same property for inertia and gravity precisely in order to preserve the identity of processes between uniform and gravitationally accelerated moving systems. In A-world, this would ultimately be seen as a necessary evolutionary adaptation to the initially accidental introduction of a longrange force (gravity), which on the other hand brings with it the evolutionary advantage of keeping things 'close at hand', but still locally flexible due to its relative 'weakness', and thus available for further growth processes. In accordance with the abandonment of a uniform space-time, we would not be able to understand gravity as a purely geometric phenomenon within Model A, but would still be able to arrive at the 'results' of GR.

Finally, it should be noted that Model A can also accomodate some of the 'wilder' speculations of cosmology, such as the idea of natural constants that change over very long time scales, [153] and invites and even encourages new ones, such as rumminations on whether perhaps the same (evolutionarily 'grown') rules do not apply to all regions of the universe. The requirements of identity preservation will often put a stop to this; in other galaxies, for example, very basic consistency rules will ensure that there are not differently functioning uranium atoms, while conversely the purely random addition of non-material properties will simply remain without consequences. In Model A, the Big Bang is 'only' the exploration of the 'color scales' by the microsubjects, with universal properties practically as 'fields' over these color scales. It is also conceivable to ask whether perhaps the step towards life could first require a 'pre-biological' evolution, in which material bundles and micro-subjects would first have to have acquired some additional rules before the step towards the simplest living beings can succeed. If this 'pre-biological evolution' had so far only succeeded on Earth, the 'Fermi paradox' of the otherwise lifeless universe could be explained in this way. In any case, it should always be borne in mind that also for A-world two things hold: The advocacy of theories such as the above require further arguments. And only on the basis of careful empirical investigations can it be decided whether such theories can have any claim to truth.

6.2.2 Possible connections of Model A to mathematical formulations of alternatives to the general theory of relativity

As with quantum theory, we can use existing alternatives in the literature for the mathematical formulation of gravitational theories on the basis of Model A. Relativistic variants of models that are based on a modification of Newtonian theory for large distances (i.e. relativistic modified Newtonian dynamics (MOND) [154] approaches such as tensor vector scalar gravity (TeVeS) [155]) appear suitable here. With a view to a unification of quantum theory and relativity, modifications of the 'loop quantum gravity' approach could also be considered. Nevertheless, it has to be kept in mind that all these alternatives to GR already entail their own problems.

Especially long-range deviations from laws could be understood relatively easily with Model A as a result of the evolutionary development of the rules behind (or rather under?) our laws of nature. Model A always offers the somewhat unfortunate possibility of understanding empirically found relationships as consistency rules, which as contingent evolutionary artifacts would not always require further explanation. However, as in our established theory of biological evolution, this should only ever be seen as a last resort. For example, the well-known elementary particle 'zoo' contains three, but no more 'generations' of quarks, of which only the first is observed in nature. Instead of dismissing this directly as an evolutionary and thus in itself meaningless artifact, proponents of Model A should first consider whether the other generations could not play an important role in maintaining the identity of systems under extreme conditions, such as those we do not only find in particle accelerators, but must also assume for the very early universe. Other properties such as particle masses should be examined analogously for their function in the *interplay* of particles.

6.3 A 'strange' unification

The question now arises as to whether we can also think the above together. Could a unification of our 'great' physical theories QFT and GR be possible on the basis of an interpretation of these theories in the light of a bundle-theoretical view of objective idealism? Here it should first be noted that the reduction of the natural sciences is to a large extent still only an ideal and that even where unification has succeeded, it may be asked whether not only a mathematical unification has been achieved, as it is discussed for the 'collection' of effective field theories in QFT. [156])

Unification of theories in physics is usually striven for by unifying mathematical models. One reason for this is certainly the great success of this approach in the past, especially in the case of Maxwell's equations. However, there are also cases, with Newton's mechanics certainly being the most prominent, in which unification was first achieved at a deeper, conceptual level, which then required the subsequent introduction of new mathematical approaches. The first type of unification seems to work best in cases where the phenomena in question can be understood as different sides of the same coin. The second type of unification seems necessary in cases where the relationships between phenomena result from a deeper conceptual connection between them.

The latter type of unification can hardly be pursued in a general, systematic way, which makes it unsuitable for academic research in physics, and attempts to proceed in this way could therefore justifiably be described as 'strange' unifications, if only because of their much higher probability of failure. If we now want to use Model A to think QFT and GR together, this is certainly to be regarded as such an attempt at a 'strange' unification and should therefore be treated with the utmost caution.

The conceptual commonalities underlying both QFT and GR according to Model A are, firstly, the construction of the physical world from universals and, secondly, the foundation of identity as the evolutionary purpose of this physical world. The evolutionary acquisition of material consistency rules required by the micro-subjects to convey these concepts would lead, on the one hand, to the balancing of properties via wave functions (and thus to quantum theory) and, on the other hand, to the greatest possible preservation of identity under movement (and thus, as explained above, to relativity). Even beyond this, Model A would offer the possibility of creating a common framework for dealing with problems in physics and the philosophy of mind.

Nevertheless, one of the most important goals of unification would be to obtain a closed mathematical description even for extreme cases of coupling effects, such as those found in the early universe or in black holes. In these cases, the above 'strange' unification is obviously far too woodcut-like to be evaluated, so that we must again exercise caution. To summarize, it can perhaps be said that speculations in the field of modern physics can also be meaningfully formulated on the basis of idealistic models.

Chapter 7

Is thinking more than quantitative information processing?

This book began with an overview of the philosophy of information as conceived by Floridi: As a new field of philosophy, with its own catalog of open questions, its own method (the method of levels of abstraction), and an old, new problem at its center, the ’symbol grounding problem’ of how data acquire meanings. Floridi’s definition of information as well-formed, meaningful and true data circumvents the problem rather than solving it. The way out indicated by his agent based semantics adds little to the familiar, functionalist understanding of meaning and is accordingly just as susceptible to the typical counter-arguments of, for example, Searle. While the philosophy of information may initially seem to promise to be able to incorporate metaphysics-related sub-areas such as the philosophy of mind or the philosophy of language in a restructured way as sub-categories, so that it could stand alongside value theories as an equal counterpart to natural philosophy in a common metaphysics, it ultimately lacks a supporting foundation that would have to consist of a conclusive insight into the relationship between data or information in the scientific sense and meaning.

Floridi’s draft falls particularly short with regard to a topic that one would expect to be at the center of a possible philosophy of information: The task of gaining a deeper understanding of the relationship between natural and artificial intelligence. We approached this topic, which has now become more relevant than ever also to society, in the third chapter by looking at Dreyfus’ criticism of research in the field of classical artificial intelligence in the 1960s and 70s. We have seen that in the (non-)discussion between Dreyfus and the early AI researchers the even earlier discussion between mathematical-logical positivism in the sense of Russell/North-Whitehead's 'Principia Mathematica' and 'continental philosophy' thinkers such as Husserl and Heidegger was repeated: The former wanted to grasp the world as a system of logical propositions, while the latter insisted on some kind of 'being-in-the-world' as the basis of all thought. It is interesting to note that the parallel 'refutations' of Russell by Gödel (in mathematics), the late Wittgenstein and Quine (as a starting point of American analytical philosophy) were important impulses for Turing, Church and von Neumann, who then explored theoretically and practically what exactly – if not everything – can be grasped mathematically-logically, i.e. can be calculated. In this sense, Dreyfus tried to remind the early AI researchers of a lesson that was taught at the very beginning of their field. The fact that he drew on Heidegger for this was therefore only logical, but certainly not very helpful to his cause. Dreyfus thus intensifies our initial question of how data or information in the scientific sense acquires meaning to the question of whether meaning can be understood at all as a system of logical propositions or relationships.

In the following, fourth chapter, I argued that Dreyfus' criticism ultimately remains largely unchallenged by the current triumph of neural networks. Just like the 'good old-fashioned' symbolic AI of the past, also the current wave of 'sub-symbolic' AI based on deep neural networks is based on the idea of thinking as purely quantitative information processing: In the case of symbolic AI, the name explicitly refers to its nature based on formal systems, i.e. essentially rules, but sub-symbolic AI is also rule-based, only now implicitly via globally parameterized, nested functions. In contrast, a long catalog of characteristics of natural intelligence indicates that it functions not only gradually, but substantially differently: Natural intelligence is closely linked to consciousness, intentionality and experiential features such as qualia, i.e. the subjective contents of mental states. It enables understanding, e.g. insight into causal relationships rather than 'blind' reliance on correlations, as well as aethetical and ethical judgments that go beyond what we can translate into explicit or data-induced implicit rules for programming or training machines. Furthermore, according to psychologists, natural intelligence ranges from unconscious psychological processes to targeted information processing and from embodied and implicit cognition to 'real', i.e. fundamentally free, agency, as well as creativity. Natural Intelligence thus seems to go beyond any neurobiological functionalism by dealing with meanings instead of information in the sense of data.

In the fifth chapter I tried to emphasize what I consider to be the core problem of the above conflict: Our inability to determine the relationship between data and meaning is already inherent in the materialism underlying our scientific worldview (in the sense of a complete determination of the mental by the material world): If qualia, concepts, values etc. are understood as the result of material processes that are realized through changing constellations of material building blocks, then quantitative information processing is the only remaining way to conceptualize natural intelligence. A theory for the formation of concepts, or more generally of 'meaning' from data – in the most materialistic sense as stable neural structures that take into account regularities between actions and feedback – then becomes the central problem to be addressed. But although materialism is a very powerful model, it does not seem likely, given the current state of natural intelligence research, that the complex philosophical and psychological phenomena mentioned above can be easily understood within this framework.

It therefore seems legitimate to examine whether alternatives to materialism could fare better here, whereby (at least) dualistic, panpsychistic and idealistic positions are available, although, as is well known, these are not without conceptual problems either: While materialism has the emergence problem outlined above, the dualist must explain the interaction between its two worlds, and while panpsychism (based on the assumption that there are non-material basic building blocks on top of the material world) would have to make sense of the (de)combination of a single mind either from 'mind dust' or a 'cosmic mind', the idealist must be able to explain the 'emanation' of the material world from only non-material building blocks. In chapters 5 and 6 I have argued for a scientifically tenable, objective idealism, which in my view could also play an interesting role in the interpretation of quantum theory; however, the following should be of importance for any alternative to materialism.

In this chapter, I will now address the core question of this work: Assuming we follow our argument above and, as a consequence, want to evaluate alternatives to materialism in this respect, how can we imagine the functioning of human thought? The following three chapters are then devoted to possible counter-arguments from neurobiology, psychology and philosophy.

7.1 The manipulation of universals as a common feature of alternatives to materialism

The assumption that alternatives to materialism could form a more helpful or more correct foundation not only for the humanities, but also for the natural sciences, does not necessarily go hand in hand with the idea that the relationship between data and meanings is automatically clarified on the basis of such a foundation. However, the alternatives mentioned offer the possibility of postulating the subject-independent existence of meanings, so that the relationship sought can be understood as one of a special type of reference. How exactly certain data or signals are able to evoke certain meanings in the minds of humans can then presumably be the subject of any number of dualistic, panpsychistic and idealistic theories, so that in any case further scientific investigations into the overall viability of such theories are necessary.

What all such alternative theories will have in common, however, is that a non-material part of the human mind will emerge alongside the material part of the brain. And it can be assumed that the majority of viable alternatives will posit that the human mind is able to manipulate universals (here: universal, non-material 'building blocks' of meaning), and that it is precisely this possibility of manipulating universals that constitutes the peculiarity of the human mind, i.e. the non-material counterpart or rather co-player to the material brain. (In principle, the manipulation of tropes is also conceivable here, think, for example, of Donald Cary Wiliams' ontology of tropes; then however, it would remain unclear to what extent an intersubjective understanding of meanings could be possible.) In any case, the question then arises as to whether conclusions can be drawn from such an idea of human thought processes, which could be subjected to further, also scientific investigatons.

7.2 An alternative view: Human thinking as quantitative information processing plus the manipulation of universals

For the non-materialist, the key question regarding natural intelligence must therefore be how exactly we can understand human thinking not (only) as quantitative information processing while remaining true to modern science. Neurobiology has worked out that large parts of human thinking can actually be understood as quantitative information processing: This begins with the reception of material signals, the conversion into neuronal activity, the neuronal processing of sensorimotor data at lower and middle levels, but becomes increasingly 'fuzzy' at higher levels, where we can normally correlate brain activity with mental activity, but not explain it causally. [157, 158]

Here the non-materialist can now propose that brains serve as 'anchors' for higher level processes, but that parts of these processes are non-material in nature. A first simple hypothesis would be, for example, that certain, for instance cortical brain regions serve as a kind of 'memory register' for mental entities: Activity in a particular brain region evokes a particular mental entity and vice versa, but the brain region acts only as a material placeholder for the actual non-material content. (More on this in chapter 10.)

If the mental content is not completely defined by the material constellation, the question naturally arises of how to describe what needs to be added. It is not too surprising that science and philosophy have comparatively few ideas to offer here. In the context of idealism, it seems obvious to turn to the 'bundle theories' of objects in philosophical ontology, which assume that objects are nothing more than the bundle of their properties, which in turn are traditionally understood as universals. But because in this case objects with exactly the same bundle of universal properties become the same object, such theories have a very nonmaterialist problem with the vanishing distinguishability of indiscernible entities. [111] I have already argued in the previous chapters that this core problem should actually be seen as a core advantage, since it allows an intelligible interpretation of quantum theory within a scientifically tenable, bundle-theoretical view of objective idealism.

If we follow this approach (even if only for lack of good alternatives for our question), we would also have to understand the human mind as a bundle of universals, anchored in a brain, which for the objective idealist would be just another sub-bundle of a whole person, with certain special restrictions on the manipulation of the 'brain-bundle' due to material consistency rules. (As a dualist or panpsychist, one could probably get by with a simpler construction, but could probably still understand the mind as a bundle of universals.) This 'mind bundle' would serve as a kind of 'world map' for the individual, representing the agent's entire world, not necessarily in strict accordance with objective measures of time and space, through which he can interact with the material world only indirectly. For such an individual, reality would not mean understanding an 'ideal network of propositions', as Heidegger criticizes it, but 'having a world' that would in fact provide a local, partly subjective context in universal, non-physical terms, as required for a human-like intelligence according to Dreyfus. [16]

7.3 Arguments for the manipulation of universals

What have we gained at this point? We have formulated an alternative view of human thinking, as not only quantitative information processing, but closely related to it: While in the 'lower' part of the model we find abstractions as signals of signals, i.e. information, further 'up' we encounter bundles of universals or nested sums of qualities and/or meanings as mental units. Conscious and most probably also some 'higher' parts of unconscious thinking then correspond to the manipulation or (un)bundling of universals, which can thus go beyond the underlying information processing, i.e. the manipulation of material constellations.

This structure avoids the 'materialist trap' of having to explain the emergence of meaning from data: Information/data is transmitted and shared across the material world, according to the consistency rules for that part of the world, but meaning is not transmitted or shared directly, it is created by the individual, through linking certain non-material building blocks with certain material signals or ('higher above') other non-material building blocks. The transfer of meaning would depend on common, evolutionarily but also historically acquired rules for the transformation of information into meaning. Simple universals or 'ideas' would thus serve as 'atoms' of complex bundles or abstract objects in our thinking.

It should be noted that the explanations in the last two chapters have left crucial questions unanswered: For instance, how exactly do subjects move from given qualities to new ones? Here we must postulate a creative element: Given qualities merely suggest, through their meaning that can be intuitively grasped by the subject, possible steps towards further qualities. The relationship will mostly be underdetermined, so that the progression from one quality to the next is essentially a free, creative act of the subject – but one that is steered along certain paths by the context and here above all by the physical and mental structures available to the subject. Both the initial conditions and the possible transformation function should be understood as (bundles of) universals. Ultimately , we might require a 'semantic logic' of the non-material building blocks; this will be discussed in chapters 8 and 11.

Whether this alternative view is of value should, in my opinion, be evaluated in a two-stage process: First, we should consider how the model fits with the list of relevant phenomena that require an explanation. Second, as with any (pre-)scientific model, we can attempt to derive predictions that lend themselves to further investigation, including scientific research. For the first step, we can acknowledge that the model does indeed provide (by design, so to speak) a way to explain both the conscious, intentional, experiential, higher-order thinking-related features of natural intelligence, as well as the unconscious, embodied, implicit cognitive ones.

We can then postulate additional mechanisms in relation to human learning behavior, to explain, for example, why we have to draw on existing knowledge when learning, but still have to take the final step towards a new meaning ourselves: For an individual with a certain material and non-material structure, a suitable set of material inputs could be a strong incentive to take the right mental 'step'.

And in terms of people's creative abilities, we can imagine moves for manipulating universals that would enable more complex steps than (sums of) inductive or deductive inferences, e.g. whole sub-bundles could be shifted according to less consistent or even 'divergent' rules, maybe even in the sense of Pierce's concept of abduction as a step that expands knowledge.

On the philosophical side, the model would fit well with the observation that our thinking easily falls prey to skeptical arguments, also in the epistemological sense of critical idealism. On the psychological side, the model would allow (or even require) a complex mental structure with strong subconscious forces; direct access to the material part of our body, for example, would be found at the level of our subconsciousness, which would make us susceptible to numerous forms of somatization, as can indeed be observed. Finally, the difference between the predominantly information-based implicit and the predominantly meaning-based explicit execution of rules could explain observations such as Moravec's paradox [159] and Kahneman's two types of thinking. [63] The issues outlined in this paragraph are further explored in the following three chapters using counterarguments from neurobiology, psychology and philosophy

7.4 The purpose of qualia: How the manipulation of universals could cut short quantitative information processing

The most important question at this point is probably how such a complex mind-brain construct could have developed in line with our biological theory of evolution. This is of particular interest in view of the findings of neurobiology that the neuronal functioning of simple organisms can be fully explained as quantitative information processing. [157, 158] So what would be an achievable(!) evolutionary advantage for more complex organisms if they could assign mental properties to material signals? (Or why else would life have ventured into the non-material after eons in the material world?)

Here too, as with quantum theory, the core problem of bundle theories of universals could come to our aid as a core advantage: Imagine, for example, the effect of using universals in object recognition (the first task in which the current AI wave was able to achieve a breakthrough) when, say, a bear approaches. In the symbolic AI model, correctly identifying the bear as a bear would be the result of repeated steps of gathering information and comparing it to a list of existing candidate objects. (To avoid false early matches or flickering between results, certain thresholds would be useful for the assessment). In sub-symbolic AI, the process would move from explicit rule-following to implicit 'recognition', with the 'rules' for identifying a bear being data-induced and distributed opaquely across multiple nodes of the neural network.

It seems clear that something like this sub-symbolic 'thinking' must also take place at the lower, information processing levels of human thought, but unlike AI, human intelligence is able to learn things from extremely little data, maybe amongst others through 'insight' into underlying contexts.

Our new view of natural intelligence would suggest that neural networks are a reasonable model for human thinking up to the generation of qualia, but everything that comes after would then be understood as the manipulation of universals: If the bear approaches, a previously unidentified sub-bundle is generated and further qualia are added according to the additional information. The trick is that this addition of qualities not only takes into account the collection of additional information, but also replaces the need for repeated comparison of property lists or the previous definition of built-in implicit rules: The more bear-like properties the bundle collects, the more identical – in the literal sense! – it becomes with the existing 'bear bundle' in the individual's world map, because the properties that are added are universal, i.e. literally the same for all entities that participate in them. Depending on the existing context, even at a very early stage the existing and the newly added properties can more or less suddenly imply a certain known bundle, at which point the 'bear bundle' can simply switch to the current context, importing as a side effect its entire 'bear context', i.e. the sum of all other meanings already assigned to it.

Such a mechanism would not only explain why context as the sum of preexisting relationships plays a crucial role in human understanding, but also why humans are amazingly good at 'zooming in' on content. The model further implies that when the correct object is identified, we experience a 'holistic' import of additional 'common sense knowledge', which can indeed be observed in humans, but is a major challenge for (sub-)symbolic AI. [16](On the other hand, when things go wrong for us, they often go very wrong: Illusions, hallucinations, etc. are completely real for us until they can be corrected.)

To answer the question posed at the beginning, we come to the conclusion that human thinking as a manipulation of universals could offer certain possibilities to shorten informational processes and that this could have been an evolutionary driving force for the development of complex non-material mind/brain structures. Similar 'shortcuts' for informational tasks through recourse to universal properties could of course also play a role for other complex mental activities, such as understanding natural language. In relation to sensorimotor tasks, mental entities could enable a kind of dimensionality reduction in complex optimization problems; a simple movement in a low-dimensional 'qualia space' could correspond to complex movements in higher-dimensional (also informational) spaces. At a lower level, mental entities could serve as 'stable targets' for feedback processes (more on this below).

7.5 The next steps: Initial ideas for possible scientific investigations

As mentioned above, the second step within our evaluation process would be to see if we can derive predictions that are suitable for further, if possible, scientific investigation. It should be clear that actual experimental investigations would require a much more detailed theory of the underlying processes than has been developed here so far, but the rough sketch of a model given above already implies a number of questions to be investigated: Do the 'upper' circuits of the brain serve only as 'memory registers'? If so, the length of the 'entries' in the brain should, for example, be independent of the actual content. Or can we influence the supposedly contingent connection between information and meaning, e.g. signals and qualia, and thus show their contingency? Moreover, in our model the mind would have to be built in step with the brain; can we find evidence for such a complex, integrated development? (Later we could ask: Does our model provide helpful insights into psychopathology?)

If we turn from neurobiology to information theory, we could ask, for example, whether or to what extent we can directly show that informational processes can be shortened by manipulating universals. Unlike in the discussion of qualia, we would not ask: 'Does Mary learn something new?' (when she sees a color for the first time), [22, 23] but 'How much does Mary learn? Does she learn a lot? An unfinite amount? (Wouldn't we need an infinite number of statements to 'define' red, e.g. by exclusion?) But can we use what we intuitively assume here for a proof? To show the advantages of manipulating universals? Probably not in a direct way: As outlined above, the transfer, or rather mutual generation of meanings from information would be limited to the respective subject, whereby the material signals are always limited to a finite information content, since material consistency also means informational consistency.

Can we construct a human-solvable mental task that demonstrably exceeds the maximum possible computational capacity of the human brain in a material-informational sense? Understanding as opposed to processing natural language might indeed already be such a task, but it seems practically impossible to correctly specify the context-dependent computational power required for a language task that would be complex enough. And how could argumentative gaps be closed that link human (over)performance to a prior evolutionary, biological, or social adaptation, e.g. as algorithmic optimization? The same applies to the comparison of the numerical and analytical solution of problems in mathematics (more on this below).

More far-reaching implications of the proposed model could probably be explored in a more distant future: Consciousness, qualia, mental causality, etc. would not be strange remnants, but central features of any mind. Accordingly, machines based only on quantitative information processing would never be able to fully emulate natural intelligence. However, general AI might still be possible if we were able to fully understand the coupled non-material mind/brain evolution to construct not only an information-processing artificial brain, but also a universals-manipulating artificial mind. The simple adoption of purely materially suitable structures by subjects 'as if by magic', on the other hand, is not to be expected, since any control of more complex structures would also require certain non-material structures, which would normally have to develop in step with the construction of the material structures.

7.6 A possible mathematical model for the manipulation of universals at the mind/matter interface

Above, I made the claim that alternatives to materialism allow us to propose models of human thought that go beyond the current standard model of (purely quantitative) information processing. A simple example of object recognition has already been discussed to illustrate how such models of information processing plus manipulation of universals could work, and the ways in which the latter might shorten quantitative information processing by relying on universal qualities. The original evolutionary advantage of using qualities, among others, could then have been to optimize informational processes, e.g. in connection with the recognition of complex patterns.

To examine this idea in detail, we would need a mathematical model for the proposed new mode of human thought, preferably in analogy to existing mathematical models of abstract machines. Probably the best known of such models is the Turing machine, which in each processing step starts from an internal state and then transitions to a new internal and memory state based on given transformation functions and input symbols taken from its memory tape.

It should be noted here that the model we are looking for is for the interface between quantitative and qualitative information processing; we can already simulate purely quantitative information processing with the Turing machine, and we can describe purely qualitative information processing analytically via its symbolic representation – and where this is not possible, at least to a large extent via language. The interesting case occurs when quantitative information is to be translated into qualitative information; for example, when a color is assigned to the optical signal for improved pattern recognition, which can then in turn be used universally across contexts.

What would change in the Turing machine now that we manipulate universals? Based on the assertion that universals-based bundle theories allow an intelligible interpretation of quantum theory, one could intuitively assume that there should be a connection to quantum information processing ('quantum computing') and therefore that, with a modification of the Turing machine known as the 'quantum Turing machine', the model we are looking for could already be available.

As explained in the excursus in Chapter 6, we observe in the context of quantum theory, when a system is formed from parts, e.g. a molecule from elementary particles, that the individual parts are incorporated into the overall system in such a way that they lose their independent identity. As a result the system requires a global description, e.g. via a wave function and its evolution, until a measurement occurs that causes this construction to collapse irreversibly and that leads to statistically distributed and quantized (i.e. not continuously distributed) measurement results for only context-dependently available properties. As explained in chapter 5, a similar vanishing distinguishability of indiscernible entitites is also discussed for bundle theories in philosophical ontology. Within the framework of such bundle theories, it would therefore not be inconceivable that entities at the micro level could temporarily disappear or merge into supersystems. The mathematical tool of the quantum mechanical wave function would then be a tool for accounting for the materially relevant properties of a system.

With regard to quantum information processing, in the above view, the entanglement of subsystems corresponds to the merging of properties in the overall bundle. By manipulating the overall bundle, all possible system states can then be addressed simultaneously, so that in the Grover algorithm, for example, a desired state can be projected out step by step through the skillful (multiple) manipulation of the overall system. (The non-locality of quantum systems resulting from contextuality is accordingly central to quantum computing. [160]) This can be used, among other things, for searching in a list, where the computational effort then only increases linearly with the square root of the number of entries in the list, which can represent an extreme saving for very large lists.

The essential point, however, is that here a certain number of indistinguishable particles merge into an overall system, whereas we want to investigate the opposite case, namely mental constructs in which otherwise indistinguishable instantiations of universals first acquire their identity as part of the identity of an overall system. We are therefore not interested in constructing systems from many parts that are always the same, but in understanding the connection between systems that are constructed from the same pool of completely different but universal parts. Unlike in the material world, in which entanglement or bundling is only possible when the identity of the parts can merge into the overall system due to their indistinguishability, in the mental world under discussion we have the situation that all parts are indistinguishable as universals and only receive their share of identity from being bundled into an overall entity, which derives its identity from the material world.

The analogy we are looking for is therefore not between universals/bundles and parts/supersystems, but between universals and the states of the supersystems, which are always superpositions of the states of the subsystems. As soon as the subsystems are merged into the supersystem through their entanglement in quantum information processing, the possible states of the supersystem behave in the same way as we have envisaged for universals: The manipulation of one state can have a direct effect on all other states, e.g. when picking out a property acts as a 'universal deletion' of all states without this property. We could therefore assume that the manipulation of universals works analogously to the quantum Turing machine, but without the need for a separate preliminary entanglement of subsystems, which is physically necessary for material systems.

The modification of the Turing machine known as the quantum Turing machine could therefore be the mathematical model we are looking for. For such a machine, we would replace the alphabet of possible input symbols with one that offers with each input a weighted superposition of the entire set and also allows for superpositions of the internal states, so that transformation matrices must be applied instead of transition functions. A typical algorithm then consists of the repeated application of a suitable transformation matrix in order to pick out the desired result step by step from the superposition of all possible states and to amplify it in order to make it measurable. The repetition of the application is necessary because the results of each measurement are only obtained with a certain statistical probability, so that the amplification must be sufficiently strong in order to obtain the correct result with a high enough probability. This would be different for the 'universals machine': A single application of the appropriate transformation matrix directly picks out the correct solution. (In the language of quantum information processing, only the application of the 'oracle' is required, but not the diffusion operator for amplitude amplification, since the 'mental state' picked out does not correspond to an entanglement of physical particle states.)

In any physical implementation of such a machine, the necessary adjustments would require further steps; however, this would not be the case for human thinking if it really manipulated universals.

At this higher level of abstraction, the difference in our example above with the bear could then be illustrated as follows: In simple quantitative information processing analogous to the Turing machine, the comparison of the observed properties with the properties of possible objects corresponds to the search for the object 'bear' in a list with N entries, whereby n properties must be compared in each case (total effort O(n x N)). In quantum information processing with the Grover algorithm, I could match my list of n properties for all N objects simultaneously, but I would need a number of $\sqrt{N}$ steps resulting from the statistics of the measurement until the result is statistically significant (total effort O(n x $\sqrt{N}$). When manipulating universals, the task of matching the properties is eliminated, as these are universal. We simply add properties until the constructed bundle is indistinguishable from the target object. (If this is not possible, all approaches reach their limits; I am then correctly unsure of what exactly I am looking at.) Just as with the quantum algorithm, the entire list can be matched in one step, but there is no need to amplify the selected state for the quantum mechanical measurement via $\sqrt{N}$ steps. The total effort is thus independent of the number of possible objects N (the length of the list) and only increases with the number of properties n to be taken into account (total effort O(n). It must remembered here, however, that the effort of 'delivering' the qualia via the information-processing brain will scale less favorably.

From the perspective of Model A, the key point would be that in quantitative information processing, symbols are uniquely determined only by their relations to each other, i.e. context-dependent, while universals are uniquely determined as such, i.e. across contexts. In the example above, my knowledge is coded as a list of objects with properties, so that I have to search in the list if I want to infer the associated object from a set of properties – and the reverse structuring would make the search even more difficult. In any case, only the pairwise relations between qualities are coded, and their universally unambiguous and thus context-connecting nature plays no role.

However, if I can handle the universal qualities themselves, then not only the existing but also the non-existing relations can be used: If brown is part of the unknown object, not only are all brown objects picked out instantaneously, but all non-brown ones are also discarded. And unlike quantum information processing, the direct manipulation of universals does not require the summation of statistical results. Universals 'entangle' all objects that participate in them; a process that we mimic in quantum information processing by entangling particle states with each other and then assigning universals to the resulting superpositions.

On the basis of these considerations, it would now be necessary to devise experiments that provide clues as to whether natural intelligence can be better represented as pure quantitative information processing or as a combination of information processing and manipulation of universals. Further evidence for this can already be found in the field of 'quantum cognition', [161, 162] which does explicitly not assume that brains are quantum computers, but attempts to take into account that at least some cognitive processes can be better represented by means of quantum information theoretic algorithms than by their classical counterparts. The contradiction inherent in this could be explained by the possibility of the human mind manipulating universals; both phenomena would then derive from the fact that the world is structured on the basis of universals.

7.7 Central aspects of human thinking in Model A

Human thinking would thus have to be thought of as quantitative information processing, quantum information processing-analogous and qualitative information processing in at least three layers. In view of the criticism leveled at the image of purely quantitative information processing in chapter 4, the central difference would be the possibility of linking material signals with mental entities. This would also explain the capabilities of natural intelligence for 'stable' and 'broad' abstraction, discussed in chapter 4: Stable abstractions conceived in this way leave their empirical basis behind and can then be manipulated on a new level and other than deterministically, quite independently of the empirical basis. Broad abstractions conceived in this way open up superordinate facts (qualia, ideas, forms, universals, ...) and thus cover an infinite number of cases. Such stable and broad abstractions can then be used also across contexts. All this is only the case for the realization of abstractions in sub-symbolic AI systems if these abstractions are already given as context, e.g. in the form of basic elements such as language tokens.

The question already posed above, of how much quantitative information processing can be saved through stable and broad abstraction, really is the key. When investigating the direct link between quantitative and qualitative information, as we find it with optical signals and color impressions, the material and thus quantitative informational consistency of the physical world always gets in our way, as also already mentioned above. Only if we could be sure that we are working with exactly the same qualitative information, we could use this to demonstrate the short-cutting of quantitative information processes: Imagine a very large number of color plates from which I can select a desired plate on the basis of a color impression communicated to me; can I show that detecting the color as quality offers informational advantages over receiving the physical signal? The problem is that the quality of color cannot be communicated to me directly, but only mediated via physical signals, so that the separation of signal and meaning necessary for the argument is impossible in the envisaged experiment. Also, the respective algorithm on the basis of which purely physical, neuronal information processing would solve a given task is generally unknown, so that it would in addition remain unclear which specific computing efforts one would want to use for comparison here.

At the level of the direct linking of signals with meanings, the already discussed access to 'superphysically' stable and broad attractors for feedback processes seems to be a direct advantage, which, however, can again not be easily quantified. The situation is different when using abstractions between contexts; here there would be the observable and quantifiable advantages of being able to classify patterns across contexts more quickly like in the bear example above. But even in this case, the processing of neural signals sits at the beginning and the neural algorithms at a higher level are again unknown. In this sense, one could assume, for example, that the human abilities in object recognition and especially in language understanding [48, 49] represent such 'overachievements', but in order to quantify them, one would first have to find out which conceivable neuronally implemented algorithms are to be compared with the use of universals. It is not enough to observe that current neural networks require much more computing power, as we already know that our brain is 'wired' in a much more complex way and therefore also the purely physical information processing is certainly already based on more efficient algorithms. Taking the connection between signals and meanings even further, it may be possible to construct advantageous examples if one assumes that abstractions, unlike quantitative information, could be more easily inverted as a whole and across contexts, e.g. when used counterfactually. However, the problems of quantifying this are again the same.

If we break away from the direct link between quantitative and qualitative information, stable and broad abstractions seem to enable us to save practically any amount of quantitative information processing: In mathematics, quantitative-numerical information processing is finite, local and discrete, while qualitative.analytical information processing allows the manipulation of infinite, non-local and continuous abstractions such as the realm of real numbers. [1]

In human language, on the other hand, we find stable and broad abstractions in phenomena such as broad (especially ethical) terms, 'open texture' in the wake of Wittgenstein and Waisman (so far mainly in the philosophy of law), or even when we try to describe the functioning of language via quantum theoretical models in 'quantum cognition'. In psychology, the idea is found in the form of 'Gestalt' theories, in the natural sciences with the call for insight into causalities instead of just making use of correlations, in art we encounter stable and broad abstractions when finite material constellations, i.e. arranged materials, evoke 'dense' symbolism, i.e. the work of art. [163]

One consequence of the above considerations, which will be discussed in more detail in Chapter 11, should be mentioned here: With the further bundling of abstractions/universals, we 'pack' them into new bundles, parts of which subsequently appear to consist of seperate parts only upon targeted inspection. We operate most efficiently with (higher) abstractions when the details are hidden from us; the world according to our perception and thinking appears to us as a whole. When we trace the details, they become our 'whole' (conscious, because focused) world; we 'forget' everything around us; our thinking is thus always already intentional.

7.8 Is meaning based on universals?

In the above model, universals play a central role in the understanding of meaning also beyond the 'elementary' meaning of the universals themselves. In the informational processing of the sentence 'Xanthippe is a woman', the term woman initially has no retrievable meaning for the processing system, since the assignment does not automatically result in further assignments that give the term meaning in the sense that it makes explicit what is associated with the term woman in other contexts. The system can now be given further sentences with such assignments, but here we end up back at

[1] Conversely, Model A could be of interest for problems in the philosophy of mathematics that are concerned with the fact that deterministic information or deduction processes cannot generate new information; in Model A, subjects generate new information through the 'creative' assignment of qualities to signals.

Dreyfus' criticism, that the contexts on which human thought can fall back can in general not be summarized in such lists of sentences. However, if we understand meaning as such a cross-contextual mediation of relationships, then it is precisely the existence of universals which make meaningful thinking possible in the first place, by retaining universality in every context and connecting contexts via this property. [2]

7.9 Subjects

In Chapter 5, we have already seen that subjects play a central role in objective idealism. It is important to distinguish between two conceptions of objective idealism here: The 'narrower' understanding of the term demands the objective existence of non-material things such as qualia, numbers, values, but could be reconciled as dualism with a materialism limited to the material world or a correspondingly derived panpsychism. The 'broader' understanding of objective idealism sees the material world as derived from the non-material world. In Chapter 5, I argued for the latter, since in my view dualism and panpsychism (understood in this way) are not equally helpful in bridging the perceived gap between mind and matter.

The essential point is that a 'broadly' understood objective idealism, as explained in chapter 5, goes hand in hand with the fact that causality must always be understood as mental causality. At least one subject is then necessary to guarantee the 'maintenance' of the world, i.e. the continuation of causal relationships. In the traditional sense, this can be a world soul or a god; alternatively, a (gigantic) ensemble of simple subjects can be postulated. With Leibniz, one can also propose a combination in which the supreme monad God guides the ensemble of themselves essentially passive monads through a pre-determined 'harmony'. Finally, a combination in which a multitude of subjects act on different levels of existence would be conceivable. In chapter 5, I have argued for an ensemble of simple 'core subjects', since this can be understood as a minimal model for the aim of my investigations, namely to assess whether alternatives to materialism can provide a more helpful understanding of natural intelligence. Whether further subjects are

[2] Here we can think of a connection to Cassirer's philosophy of symbolic forms [164], in which symbols – understood not in the mathematical sense of AI research as formal, but as meaningful signs – are universal in the sense that they have an independent existence through the possibility of using them in different contexts. The above considerations would like to contribute to explaining the ontological status of such symbols, especially with regard to their relationship to the scientific idea of the concept of information; here too, the connection between meaning and information (symbols and signals) must be explained.

perhaps even necessary remains an open question here, but would in any case require additional arguments, e.g. from the philosophy of religion in the sense of Plantinga. [165] (We have discussed the alternative, not yet mentioned here, that properties themselves could have dispositions in Chapter 5).

In view of the alternative view of human thought outlined above, the resulting idea of what a subject is will therefore seem far too much to one person and far too little to the other: Even if we leave out materialism, which cannot muster any understanding of the idea that a non-material subject-core could exists and could reach out to an objectively existing mental world, a core-subject in the bundle of its qualia will be far too powerful for the deflationary panpsychist, while for the classical idealist this subject is far too weakly conceived if it appears only as a 'balance-sum' of universals, even if it comes along with an 'engine' at its core.

In the dispute between Marx and Fromm as to whether materialism is able to grant the subject a center in the sense of a unifying consciousness, the argumentation proposed here clearly takes Fromm's side that this is not the case. However, in the field of tension between Hume (later Mach) and Kant as to the extent to which the subject is no more than a bundle as opposed to the sovereign center of the contents of consciousness, Model A is positioned halfway: Again, this is a matter of constructing a minimal model in order to answer the question of whether we can think of alternatives to thinking as quantitative information processing. For this minimal model, a clear positioning on the side of the sovereign subject would not be helpful: Freud's 'discovery' of the unconscious and more than 100 years of psychology in its succession conceived in the broadest sense speak *against* a monolithic subject whose thinking could be idealized with or rather against Rorty [166] as a mirror of nature. And it speaks *for* one that can be diagnosed with Kahneman [63] as mostly far removed from such ideal rationality, and should better be understood with philosophers such as Wittgenstein and Heidegger as always already situated. Our subject, as well as its thinking, *must* therefore be seen to a large extent as a product of its intellectual as well as material history; it must to a large extent be a bundle. A pure bundle would now have the advantage, which deflationary panpsychism makes use of, that a unification with the natural sciences would appear easier. However, this would not be a real bridge to philosophy, since the core problem that Kant saw would no longer inform our deliberations; even the core function of the subject would have been 'discussed away'; instead of a minimal model, we would get one that is too simple.

The model proposed here therefore assumes minimal core subjects that must fulfill at least two functions: To be able to perceive and freely manipulate a bundle of universals. The former allows for a consciousness to be the unifying element of the subject bundle (neither the universals nor the non-perceiving core subject can do this), the latter is necessary for the maintenance of the world (i.e. ultimately the development of tropes in a world of universals). In my opinion, the accusation that such a subject can only ever be a ’balance sheet total’ does not apply. The core subject is only ’unfree’, and thus called into question in its role as a center, insofar as it can act beyond its material and mental possibilities to a very limited extent only, which, nevertheless, seems to correspond well to the human experience.

A simple subject is driven by material impulses; its ’freedom’ may be limited to following certain more or less randomly acquired rules somewhat sooner or later. However, a subject with a correspondingly complex material and mental structure will always be strongly underdetermined, especially when manipulating non-material bundles. We may, if her development has allowed it, imagine a person as so ’richly’ structured that she is free in the sense of a sovereign subject through the perception of the meanings of various complex mental constructs and the freedom to carry out further manipulations on the basis of these perceptions (provided that her current situation does not otherwise restrict this sovereignty). Kant’s three cognitive faculties of *Sinnlichkeit* (sensuality), *Verstand* (understanding) and *Vernunft* (reason) thus become explainable as the perception of qualia and abstracta, as well as the possibility of the free(!) manipulation of more complex structures. (Although the core subject only experiences its reasoning as free.)

It is not necessary to decide at this point, how comprehensive, or how uniquely structured real core subjects can be, because minimal and indistinguishable core subjects are sufficient for our purpose here. Even these ’faceless’ core subjects are tropes, since each of them can pick out a position *vis-a-vis* the positionless universals; each enables an individual consciousness. (If they were not, there would be only one subject.) Largely without direct consequences for the proposed model, this can be taken further: Each core subject or group of core subjects could differ in mode of perception or propensity to act, and the ’core’ of some subjects or groups could encompass more than just perceptual and action capabilities. Certain further properties would then be essential for these subjects, which may then be quite individualized in the everyday sense. Whether such an ’inflation’ of the subjects could later still be considered necessary must remain open here, as further arguments for this would first have to be found. In view of the central concern of this book, however, this question may well remain open. (And with regard to ethical considerations, the model is anyhow a warning against our irrationality, but also an invitation to use the rationality available to us.)

Accordingly, the proposed model should be of interest to both traditional idealism and deflationary panpsychism because, unlike materialist alternatives, it at least offers the possibility of developing the model further in one direction or the other with additional arguments, should the bridge between information and meaning be considered successful. In the case of traditional idealism, a possible further development via arguments, e.g. from the philosophy of religion, for the existence of God and manifold created (instead of only evolved) souls can be readily thought of. For panpsychism, there is an additional problem with the proposed model, besides having to provide a solution to the combination problem of how a bundle becomes a whole (so that, conversely, the core self can also be 'unbundled'): The connection between a material signal like for instance neuronal activity and a mental component like for instance colour appears to be neither materially causal nor completely random. As already mentioned in chapter 5, the above considerations are to be understood in such a way that a core subject makes the connection in the form of an act of mental causality; it perceives both neuronal activity and color and follows learned rules in manipulating each other. Due to the lack of material causality, at least in the mental world, more complex mental rules cannot have developed by any necessity from simpler mental contents, so that we also need a perceiving and freely acting core subject for our model at this point. In contrast, conceivable psychophysical causal relationships in panpsychism would always require further, purely mental causal relationships. As a result, a panpsychist solution would only be possible with a strong (e.g. emergent) subject, which conversely would rule out a fully deflationary panpsychist solution.

On the opposite side, the model experiences its greatest friction with traditional idealism not through its conception of the core subject, but through its understanding of a person as a Leibnizian ensemble of subjects. In the model presented, as with Leibniz, subjects are at work everywhere in the world: On the micro scale they ensure the maintenance of the causal processes in the material world, while on the meso scale they constitute the unchanging core of a human mind. A person is more than this core self in the sense that the causal processes in their material body are maintained by countless simple micro-subjects and that the unconscious processes at the interface between body and mind described above, as well as those processes that we summarize in the psychological concept of the unconscious, must be driven by further sub-subjects.

The core subject thus not only holds together bundles of universals, but also an ensemble of simple subjects. If I think this is absurd, but want to stick to a broad understanding of objective idealism, then I have to put all the work on God (or consider properties with dispositions). Also in view of the fact that we already know that our body includes a conglomeration of microbiological organisms, including the microbiome, and that the idea of understanding our ego as a kind of team has long been established in psychology, the ensemble variant appears to be the more minimal model, above all because of its proximity to the theory of evolution, as already described in chapter 5, and the associated potential for its reinterpretation, e.g. with regard to the gradual development of life and natural intelligence on the meso-scale.

In so far as the model does not aim to solve the problem of the core subject, but rather to bridge the perceived gap between matter and mind (here by means of a more helpful explanation of the relationship between information and meaning), and therefore, where it requires meaning and a consciousness capable of agency, simply assumes them as a given, it is of course also a disappointment. But at least in the natural sciences, paradigm shifts are of this kind: At the beginning of the scientific revolution is the positing of solid bodies as the sum of the properties that further theorizing had to presuppose at this point in history. And modern physics only became what it is today through the recognition of quantization and the constant speed of light as fundamental phenomena. More recent developments such as string theories then attempt to catch up with these assumptions with even more fundamental assertions.

Models such as the one proposed here therefore do not claim to be eternal wisdom, but at best milestones on the way to understanding our existence better and better. Meanings and a consciousness that is capable of agency, as at least currently necessary presumptions of (some parts of) philosophy, should therefore indeed be taken as the basis of model building, just as we do with those presumptions of our scientific theories that seem equally unavoidable; if we leave the former out, we do not get a simpler model, but one that is too simple. In case a model is then successful, this success is always at least partly an argument for the initial assertions. On the meagre basis of this chapter, this of course appears to be quite optimistic thinking or even hoping into the future; it is therefore first necessary to collect arguments for the proposed model, which I will attempt to do in the following three chapters by refuting counter-arguments from neurobiology, psychology and philosophy.

Chapter 8

Counterarguments from philosophy

Before attempting to refute counterarguments in the following three chapters, it seems appropriate to summarize the proposed 'Model A' (or *A-world*) once again at a somewhat higher level of abstraction. This also makes sense because, although we have arrived at this model in a certain way, namely through reflections on the nature of human thought, this way is not the only one through which such a model can be motivated. Alternatively, for example, open questions about the nature of quantum mechanical objects, fundamental cosmological properties, the nature of mathematical entities and also ethical or aesthetic values can lead researchers and philosophers to the assumption of objectively existing, non-material building blocks of qualitative and universal nature. A common feature of these paths is that they are not motivated by a 'narrow', but a 'broad' understanding of the mind/matter problem; that they do not 'only' want to explain the functioning of our brains, but reconcile two worlds. Given the millennia-long success of idealist positions in exploring the non-material world, it should then come as no surprise that Model A is an (objective-)idealistic one.

8.1 Model A at a glance

It is therefore the basic assumption of the objective existence of non-material building blocks, which we may have arrived at in various ways, that is at the core of Model A. The only way to reconcile this assumption with the successes of modern natural science without contradiction seems to be an objective idealism, if the problem of emanation – why and how the material world has arisen from the immaterial – can be solved. Alternatively, we would have to present solutions for the interaction or the (de)combination problem, as well as explain abstract entities differently, but at least with regard to the first two problems, no breakthrough can be anticipated at present. The uncontradictory unification with the natural sciences requires in particular – and in clear deviation from classical idealist models (which, however, neglect rather than deny the problem) – that the material world is already structured on the micro-scale and that a pronounced gap between the mental and material world of subjects must result from this on the meso-scale.

Since only mental causation processes are available as argumentative tools in a consistently thought-through idealism (which is only then immune to the interaction problem), the observed formation of a microscale material world can only have taken place via the evolution of a population of – in the sense of Ockham; as simple as possible – subjects and their interaction with the given building blocks. (The alternatives of a god or properties with inherent dispositions would both require additional, rather more problematic arguments.) The driving force of such a process can then only be the 'self-fulfilment' of a set of ideas, i.e. non-material building blocks, such as growth, identity, etc., which are initially acquired by the subjects by chance, but then lead to self-organizing and self-reinforcing evolutionary processes, the end point of which so far is the formation of a physical, then biological and finally also cultural world

Central to this is the idea of growth, which is 'self-fulfilling' to the extent that on large time scales only that which is preserved in the sense of this idea can be observed. In a world of universal building blocks, the idea of identity (of separates within a whole) then becomes equally important, as without it no stable growth of separate creations is possible. The formation and separation of the material world thus appears as an evolutionary implementation of these two ideas, as a result of which material objects receive identity through their positioning in space as the basis for their 'stable' growth. The functions of space – as well as all physical-causal processes – can then not be traced back to physical-causal properties of substances, but only to the mental causation of micro-subjects, which, however, follow evolutionarily fixed rules in the manipulation of (in themselves dispositionless) properties or bundles of properties (mental or physical objects): The 'extended' can thus develop from the 'non-extended' in the sense of 'Leibniz's problem', because having extension only means that micro-subjects will generally follow certain rules in the manipulation of 'spatial' properties. In addition to an idea of cooperation, a set of further ideas or values of fundamental importance for the growth of spatially-materially anchored subjects comes into view, above all the classical trio of truth, beauty and the good (more on this in chapter 11).

The requirement for the proposed model is that the biological section of such an 'extended' (not only biological, but also physical and cultural) evolution must be in line with what we know about biological evolution from modern biology to date. It should be noted, however, that the proposed alternative extends the model of biological evolution in the sense that, especially for higher organisms, the possibility of accessing universals provides 'superphysical' stable and conceptually broad attractors for feedback processes. (Among others, those abstractions we have been on the trail of since the beginning of the book). This will be discussed in more detail in the second half of chapter 10.

The extended evolution outlined above then results in a seamless transition from 'cellular automata' or 'physical microbes' on the microscale, via simple organisms and then animals, to humans, who are characterized by the fact that their 'core subject' manipulates a bundle of non-material universals that is only indirectly coupled to the causal network of the material world. (In contrast, if one would want to think of the cellular automata dualistically as purely material, the seamless transition remains unclear). The development of this 'world map' – which is not only to be conceived spatially – allows far more than the faithful representation of physical causal relationships and thus makes human cultural achievements possible. In particular, it enables both the selflocalization and subsequently the self-reflection of the person, as well as the intersubjective reality of social entities, i.e. our social world, which is superimposed on the material conditions.

Our understanding of individuals must then be expanded to the extent that they must always already be ensembles of subjects in which a 'core subject' must be supported by several 'sub-subjects', i.e. a rather complicatedly structured subconsciousness, if only to enable mental causation while adhering to the material consistency rules (more on this below). Unlike a core subject, however, an individual is thus always already integrated into a bodily-physical and, in the next step, a historical-social context, which allows the core subject to exercise its inherent freedom, but also restricts it by providing it with the abilities and possibilities associated with the context – and only these. [1]

[1] While the core subject in Model A is set as the basic building block and thus indestructible, a subsequent exchange of the core subject – which is practically impossible to realize in Model A – would result in the continued existence of both the now unconscious and identity-less former core subject and the person now marked as different by the new core subject, though identical in all characteristics. Since neither the (contentless) core subject nor the (intention-less) world map alone is a conscious mind, in Model A a person can only really be recognized when they come together, which means due to the anchoring in the material world necessary for their identity, only as a combination of body and mind.

Our understanding of physical objects must also be expanded; they are neither fully determined by the causal network on the microscale, nor by their bundle of qualities (shape, color, etc.) in the world map, that is neither by their primary nor secondary properties alone. As a finite section of reality, they can only be understood as part of the world map, since the causal network with its particles (as bundles of material properties) and their material-causal interactions do not ultimately allow a division into individual objects: On the atomic scale, meso-scale objects are more or less fluid. However, material objects are also not just finite sections of our world map, because it is precisely the feedback via the causal network on the microscale that makes them *physical* objects. Every object in the traditional sense (and including our body) thus bridges the gap between the material and mental world, and we do the same in every act of recognizing the world.

In this model, a person is the totality of body and mind. The mind is the totality of the core subject and the structured bundle of universals that the core subject can perceive and manipulate. The structured bundle is the representation of the person's world, their 'world map' (not only to be understood spatially), through which the core subject interacts with the world. World map and core subject can only be meaningfully understood in combination as a (self-)conscious mind. A 'mapless' subject is without any conscious content, a 'subjectless' world map is only an abstract object. Parts of the world map of the core subject are also part of the world maps of sub-subjects, which make up the person's subconsciousness and which allow interaction between the core subject and the body: Parts of the world maps of the sub-subjects are in turn part of bundles with physical properties that belong to the structure of the person's brain, whereby the respective sub-subjects can manipulate both the mental and physical properties of their bundles on an equal footing, albeit according to different, evolutionarily learned rule sets.

The brain as part of the person's body thus functions as an anchor for the non-material mind of the core subject. However, the person's body can only be understood as a hybrid entity; it receives its embedding in the causal world via the countless bundles of physical properties of which it consists of (organs, molecular structures, but ultimately elementary particles), but its unity as an entity only in the connection of these structures to structured bundles of universals in the mind of the core subject, which is only aware of its body in this form. Every perception or action thus builds a bridge between causal network and qualitative representation, which, as we shall see, is causal in nature, but whose evolutionary emergence must be understood as a creative

If we consider only the loss of the core subject instead of an exchange, what remains is not a (philosophical) zombie, but a coma patient.

act.

Human thinking then encompasses a whole spectrum from purely quantitative information processing in the form of physical processes to purely qualitative information processing in the form of the structured un/bundling of (bundles of) universals. Thought processes guided by formal criteria would correspond to the manipulation of quantities on the basis of rules 'copied' from the material world; however, the freer creativity of subjects with richly structured world maps would also allow much 'wilder' operations as we observe them in music, art and literature, for example. [2]

8.2 Is the model inconsistent or excessive?

After this brief overview, we can now ask whether the model is not misleading at one point or another. I consider it to be consistent, but (many) further investigations are certainly necessary here – especially in view of the physical considerations in chapter 6.

Criticism of idealistic theories that draw attention to logical inconsistencies due to the structuring of objects by ideas on the meso-scale can be averted by pointing to the structuring of the physical world already on the micro-scale and the hybrid nature of objects in the combination of causal network and world map. Thus, for example, the considerations developed by Plato in the Parmenides dialog; neither must there be separate ideas for all entities, since these can be composed of simpler ideas, nor is the logical consistency of the world map required in all parts; the idea of a round square, i.e. an irrational mental world, is readily possible as long as it cannot 'break through' to the physical world. The latter is logically-rationally structured by the material consistency conditions and not allowing fundamental irrationality break through is precisely the 'evolutionary purpose' of this world.

On the basis of the modern reappraisal of the arguments in the Parmenides, e.g. by Rickless [167] and Gill, [168] one could now try to show exactly which assumptions (such as purity, uniqueness, causal consistency of non-material building blocks, but also their separateness from and participation in objects) are fulfilled or rejected in Model A and where. In any case, in the paralogical non-material world of the model, there is nothing to be said against rejecting the assumption of purity (that no opposing properties are conceivable for structured mental entities), in accordance with Rickless' argumentation; the evolutionary purpose of the physical world, however, is to enforce it. A the central difference is the rejection of Plato's notion of

[2] A list of the key terms used in this summary can be found in the glossary at the end of the book.

a fundamental separation of ideas and matter, or subsequently ideas and objects as their 'shadows'.

Anyone who agrees with the above assertion at least for the time being, namely that the model seems to be consistent, can of course still ask whether the proposed model is not anyhow hopelessly overpopulated. The answer to this is that the model is indeed a kind of overkill for each of the individual problems addressed, but that it should be understood as a minimal model with regard to the sum of the problems: If we do not want to abandon qualities and subjects, we will have to introduce them (according to the current state of knowledge) as fundamental entities. For however science wants to explain the human being from the perspective of the observer, the problem of subjectivity, the unique view of the individual on the whole, is still in need of explanation: Neither subjects nor qualities are unfounded, but in a narrower sense actually the most consistent assumptions. On the other hand, qualities are not simply cognitive dispositions, but are integrated into a causal network, which is why unlike in classical idealism we must also integrate what we have learned from the natural sciences about the micro-structure of this causal network, i.e. the physical world. And as long as the interaction or de-/combination problems cannot be solved convincingly, we are left with only idealistic approaches for this integration.

In this respect, the idea of an extended evolution is still part of a minimal model, as there are no other argumentative tools available (beyond religious considerations or dispositional properties) than developmental processes caused mentally by minimal subjects. Here it certainly depends on the extent to which the reader is inclined to agree with the idea of the material world as a necessary 'anchor of identity'. In any case, this does not seem inconsistent to me and it additionally seems to be of argumentative value because a number of 'oddities' of our world appear practically necessary as a result of it. These oddities are:

1. The existence of the material world and the gap we observe between the mental and the material world.
2. The fundamental, 'identity-founding' properties of the world; conservation laws, finite speed of light, equivalence of heavy and inertial mass, etc. (see chapter 6 for details).
3. The quantum nature of the physical world on the micro-scale due to the structuring out of universals and the associated restriction of the occurrence of stable growth processes to the (sub-)meso-scale and beyond.
4. The possible tracing back of an essentially unclear physical causality to the mental causation experienced directly by us, but also the limitations of it.
5. The complicated coupling of world and individual; as a causal expla-

nation of (but not solution to) the problem of scepticism; the dual, or rather bridging, nature of objects, including our body and cognitive processes; our subconscious, as well as other psychological peculiarities of the human being (see chapter 9).

Even beyond extended evolution, Model A can be understood as an attempt to extrapolate down from biology, rather than up from physics, to describe our world: We 'know' on the one hand that we as subjects can freely manipulate abstract entities, and on the other hand of the existence of 'lower' forms of life; the thesis now is that this remains the same down to the (sub-)microscale. The assumption of universals as the only kind of basic building blocks avoids the interaction problem; the fact that disparate things are bundled in the subject, which thus also connects places and times, avoids the (de)combination problem.

On the other hand, Model A also avoids excessive, idealistic model building, in which the non-existence of a subject-independent world is (logically inadmissible) inferred from the fundamental 'pre-formatting' of our perception and thought processes. Unlike the 'classical-idealistic' systems, Model A does not recognize any 'grand unification' that goes beyond the minimal elements required for the solution of concrete problems. There may be good arguments for larger unifications, but these arguments would have to be added and critically examined, as they entail the risk of not too few, but too many answers: The above systems have mostly turned out to be less helpful for our individual and social growth than their authors assumed, essentially because their hermetic nature had less critical and innovative potential than the more cautious, deliberately kept open models that have consequently replaced them.

8.2.1 Are the alternatives not ontologically more economical after all?

As to the question of whether Model A could not be more ontologically economical (which will be an important, if not the most important question, at least for natural scientists), it should first be noted that already Ockham understood his 'razor' to mean that we should strive for simplicity when forming theories, but that we cannot dictate to the universe how simple or (infinitely?) complex it is structured. In the next step, we can then consider whether alternatives that are acceptable to us can really be set up more economically.

On the one hand, many scientists will not want to give up the 'phenomenological content' (i.e. the general elements) of our current best scientific theories – but will in principle be prepared to take steps such as from Newton to Einstein; matter in the traditional sense appears in quantum theory either way as a mental construct of a basically unclear reality, not as the concretely tangible origin of all being. On the basis of the theory of evolution, the conviction that there must have been a continuous transition from matter to simple life forms and then to humans will then additionally be found with many scientists. Perhaps even more natural scientists than philosophers seem to me to be willing to recognize a genuine problem complex of 'subjectivity' (qualia, intentionality, mental causation, free will, ...) and to advocate a Platonic realism at least for mathematical entities (albeit not vehemently), which seems to underlie the special cognitive possibilities of human thought.

However, once this set of beliefs has been formed, it becomes very difficult to get by with a more parsimonious ontology than Model A: Either qualia are fundamental building blocks, or there are fundamental laws underlying their generation, but then it is completely unclear how the categorial difference of qualia is supposed to result from fewer laws than building blocks. And either subjectivity is a basic building block, or there is an even more mysterious relation to the emergence of subjectivity from other building blocks. Analogously, we would have to assume a similarly mysterious relation for the generation of qualia (of whatever nature) by subjects. Every becoming of qualia or subject building blocks also makes them appear in a completely different light than mathematical entities whose great, at least intersubjective, stability seems difficult to reconcile with the idea of an evolutionary, continuous development of life, but also of the individual.

If we assume, e.g. with Frege and/or Popper, the existence of a 'third world' of logical-mathematical entities, it does not seem unlikely that more complex entities in this world, as in A-world, could be built up from simpler ones (including all misconstructions ever made!); but also those simple building blocks would have to be explained. Conversely, the idea that each person should have developed their own version of red or the number one seems itself to be an unnecessary duplication of entities. And the argument that universals are to be understood as concepts does not help here, because we are concerned with the realization of these concepts in the human brain and/or mind. (Ockham already considered two possibilities here, '*fictum*' vs '*subjectivum*'/'underlying', whereby the latter may well be understood as a mental entity in the sense of Model A.) The problems of realizing stable concepts in the usual connectionist models of the human brain were shown in the first chapters.

The probable necessity of at least mathematical universals in science, but also of an idea of what exactly the token/type distinction should constitute, should therefore rather lead us to find a more meaningful application of Occam's razor in considering which minimal ontology should be considered for an extension in the sense of Model A: In addition to mathematical entities, we would probably have to include in the list at least qualia for all our senses (incl. sensation of pain, etc.), but probably also for senses that are not accessible but are nevertheless known to us, such as echolocation. Finally, fundamental normative ideas, at least a regulative idea of truth, would probably be necessary. Ideas such as 'giraffe', on the other hand, should certainly be understood as composite. With regard to qualia, solutions such as the construction of specific colors from a few basic colors could be considered, as well as the construction of mathematical entities from a few basic building blocks, such as the number one and the idea of mathematical induction, or sets, etc. Such a model would at least be ontologically more economical than a truly nominalist model, because we would have to work with a large but manageable number of non-material building blocks instead of a practically infinite number of individual objects in the universe. However, picking out such a set of building blocks, or even categories of building blocks, seems just as arbitrary as the particle zoo of our standard model; an alternative would be to assume that there are, at least practically, an infinite number of simple qualities and thus possibilities of development, in our world.

Many natural scientists would probably be more comfortable with a dualistic model that would leave physics more or less untouched; but little would be gained and some options even lost in terms of ontological parsimony: Once I have gained the conviction that the problem of qualia exists, I can practically no longer avoid a richer ontology. In addition, an ontology of fundamental physical properties (charge, spin, etc.) is rather more economical than one based on a zoo of particles. The dualist could understand universal qualia analogous to physical 'particle fields' as universe-wide 'qualia fields', whose local excitation then corresponds to the impression of a sensation in my consciousness, but here too little more is gained than an apparent connection to the existing physical ideas; here too, one would not be able to arrive at an ontology with one (or a few) field(s) for all qualia without further ado.

What might motivate the natural sciences to consider a richer ontology in the sense of Model A is the odd parallel of the vanishing distinguishability of indiscernible objects in bundle theories based on unversals as well as in quantum theory (see Chapter 6), combined with the finding that a universal bundle theory cannot do without subjects and that an 'extended' evolutionary theory of such subjects is in turn only conceivable on the basis of shared, universal building blocks.

Model A, of course, explains neither qualia nor subjects, but only shows how they could be integrated into the natural sciences; possibly only as provisionally necessary placeholders like the solid bodies at the beginning of mechanics, whose nature might be further elucidated later on. Seen in this way, the natural sciences would not have to make a fundamental problem out of subjectivity and universals; but with an ontology as parsimonious as the standard model, this direction cannot be taken.

Many philosophers, on the other hand, will probably not want to approach the problem so pragmatically: In the footsteps of Quine, they will most likely not want to make assumptions beyond the standard model and mathematical sets. But here, too, the decision stands and falls with the intuitions regarding qualia, mathematical entities and possibly also values. Do I accept the 'hard problem'? Then I am already on the slippery slope. If I am prepared to add sets, why not also qualia, if they appear necessary for a more comprehensive theory? And the less prepared I am to engage with the possibility of a Platonic realism, the more effort I have to expend in order to ultimately arrive where I want to end up in any case, namely that the functionality of universals is available to science.

As an alternative, for instance following David Lewis, materialism can be consistently thought through to the end, as this book attempts to spin out idealism to the maximum: We could assume our reality to be structured in a purely local materialistic way, as a 'Humean mosaic' of spatiotemporally instantiated properties, from which an infinite number of objects result through 'Humean supervenience' of all possible particle combinations, of which our mind only filters out those that are relevant to us. (It should be noted that this alternative, like Model A, assumes a 'causal network' of physical entities that is in itself hardly structured.) Our ability to think counterfactual, then additionally requires the concrete existence of a practically infinite number of parallel worlds (or at least world states?) to which such thought processes can refer. Very successful as an argumentative tool, especially in modal logic and the philosophy of language, this construction appears from the perspective of Model A rather as a hermetically closed subsystem optimization of analytic philosophy. (And this even if one is prepared to recognize the concept of Humean supervenience as meaningful and possible).

What is irritating about this position is that it is taken in the name of natural science and 'common sense', but then rejects both concrete scientific results and everyday rationality when they would prevent the formation of a compact theory: The concept of a purely locally structured physical world seems so improbable from the point of view of modern science that it simply seems scientifically impermissible to build a theory on it. And the equally invoked common sense is ignored at the latest when the result of theorizing is the conclusion that there are not only an infinite number of objects in every world, but also an infinite number of worlds; only in order to be able to answer the question of what a possible world is as unassailably as possible, instead of – following common sense – as sensibly as possible. (A little more cautiously: The ontological status of possible worlds still seems quite unclear in this model.) The general thrust of wanting to be even more materialistic than the natural sciences is not even helpful for the natural sciences themselves, because the concrete observations and thus the actual problems do not inform the theory-building process, through which one normally hopes to gain further helpful insights. In this sense, Model A is both closer to the modern natural sciences and ontologically more economical: It is not purely materialistic, but naturalistic, whereas the above model is materialistic, but no longer naturalistic in the narrower sense.

Probably the most unclear element of Model A, and therefore the easiest to attack argumentatively, is the assumed ability of core subjects to reach out for new building blocks on the basis of existing ones (which is related to Plato's 'greatest difficulty'). With regard to the described alternative, however, it may be asked whether the idea of Humean supervenience does not ultimately require an equally unclear mechanism.

8.3 Does the model explain what it wants to explain?

Next, it should be questioned whether the model fulfills its originally intended (minimal) purpose, namely to bridge the gap between quantitative and qualitative information and thereby allow for an expanded understanding of human intelligence. (As already indicated above, the model offers nonmaterialistic explanatory options beyond this minimal purpose for fundamental questions in the philosophy of physics, mathematics and biology, as well as ethics and aesthetics; see Chapter 6, Chapter 10/II and 11).

By design, the model is able to capture the concept of semantic-qualitative information, but extends it beyond a purely linguistic understanding to a concept of qualitative information or meaning, which then also encompasses qualities such as colors, abstractions and values, i.e. 'ideas', 'forms' or universals in a broader sense.

The necessary bridge to quantitative information then takes place via the implementation of evolutionary-learned rules for the connection of (regular) changes of the bundles in the physical world with corresponding changes of bundles in the mental world by subjects that we would have to assign to the subconscious in the case of humans. The question of what it is like for such a sub-subject to perceive, for example, both electrical charges and colors, is as inaccessible to us as the question of what it is like to be a bat; important is that no interaction problem arises, since both the mental and the physical bundles are made up of non-material universals and the latter only acquire their material nature because sub-subjects have learned evolutionarily to adhere to strict consistency rules when manipulating them. There is also no decombination problem, because although the subjects share bundles, they each have a completely separate, undivided view of their 'world map'.

However, this bridge can only be built if the emanation problem can be solved, that is if the idea that also physical reality is made up of bundles of universals can be reconciled with the modern natural sciences. In addition to the fundamental question of whether this is compatible with the quantum-mechanical nature of particles on the micro-scale (or may even explain it, see Chapter 6), it is particularly important to critically examine whether a process such as the proposed extended evolution could really lead to such a complex construction as the claimed brain/mind combination. It is therefore not enough to show that mental causation is conceivable while adhering to physical consistency rules (this is attempted below) and not even that this can be reconciled with our psychological and neurological constitution (see Chapters 9 and 10); also the evolutionary and individual developmental processes must be made plausible (more on this in Chapter 10/II).

If the model does not succeed in explaining the mechanisms of the causal network of the physical world, then it could offer a (broad) understanding of qualitative information, but would not have a consistent concept of quantitative information. So here we have the interesting case of an initially philosophical theory being dependent on the natural sciences in the sense of an inductive metaphysics in order to be able to make further progress.

8.4 Isn't mental causation impossible in principle?

The problem of mental causation has been discussed for centuries, and the debate continues in the 21st century. [103, 104] In a nutshell, one could say that the current state of the discussion leaves sufficient room for a construct such as Model A. The two main lines of argumentation should nevertheless be examined in more detail in order to demonstrate the plausibility of the model to the uninitiated.

8.4.1 The causal nexus and the pairing problem

A number of important argument assume either that mental causation is not possible because a causal nexus (i.e. an interface) for the necessary interaction is missing, or that the 'pairing' (i.e. the coupling) of mental and physical processes is impossible because the non-physical mental processes have no identity through a positing in space that would allow an unambiguous coupling in the first place. [29] (The problem to be solved, formulated as a question, is then: Why is *this* mind connected to *this* body?)

In Model A, there is no need for a causal nexus, since the transition between the physical and mental worlds is continuous and only occurs for our perception across a principled and not just contingent gap. This avoidance of the interaction problem is, after all, one of the core arguments for idealist alternatives. Model A is a 'crypto-dualism' only in our perception of an apparent gap. As beings with body and mind, we are always bridging this gap, which is the fascination of a phenomenology in the sense of Merleau-Ponty, but which, unlike Model A, does not go beyond the perceptual limits of our own body.

Figure 8.1 illustrates this graphically: People are of hybrid nature, with a body, a subconscious that bridges the gap and a non-material (conscious) mind, whereby every exchange of information must be mediated via the material world (red arrows). Objects such as the bodies of other people are of a hybrid nature in the sense that they are only completed into a self-contained object through their representation in a mind. People as well as objects are not closed in the causal network of the material world; they continuously exchange materials with their environment. In our perception we then find a 'weak' dualism with regard to the nature of our person, because we are only aware of the bridging processes in the result; but a 'strong' dualism with regard to the nature of other people, because the assignment of bundles in my world map (my image of the other person) to a dynamic section of the causal network (the body of the other person) appears un-mediated (dotted red arrows). People and objects are then not only material/non-material hybrids, but also Janus-faced in the sense that they can be completed differently in different world maps (blue arrows).

The 'second world' of subjective content thus presents itself as the result of the interaction of the 'first', material world and the 'third', non-material world'. In Model A, however, the first and third worlds differ only in that micro-subjects act in them according to different rules. Thus, all three worlds collapse into an (idealistic) monism. The distinction between (core) subjects and qualities is also not a hidden classical dualism: The difference is not between mind and matter, but between active and passive non-material ele-

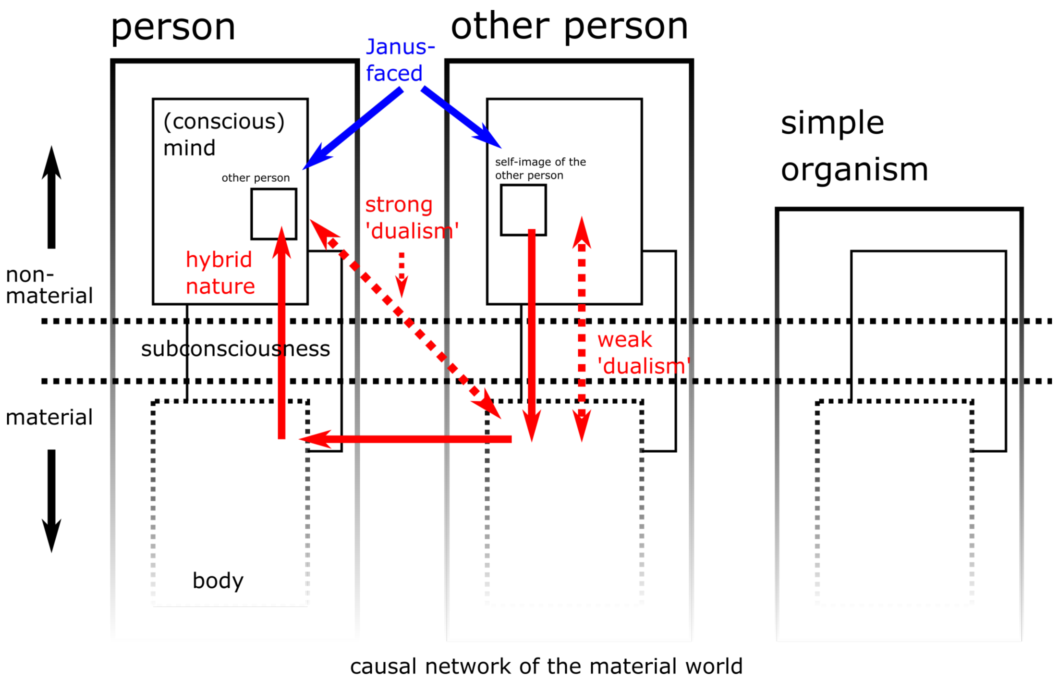

Figure 8.1: Model A is no crypto-Cartesianism: People and objects are of material/non-material hybrid nature and are not closed in the causal network of the material world (red arrows). In our perception, we therefore find a ’weak’ and a ’strong’ dualism (dotted red arrows). People and objects are also Janus-faced, i.e. they can be completed in different ways (blue arrows). The difference between material and non-material entities is not a matter of principle, but a contingent result of physical evolution. Details in the text.

ments.

The pairing problem, on the other hand, is turned on its head in Model A: The physical world has developed precisely because identity and thus causal coupling is possible here. However, the connection between subjects and their objects is a contingent one that can only be understood from the coupled development. The bundle ’held’ by the core subject in its perception and through its actions, is the result of this development and can therefore not be addressed by other subjects in the same way. (Chapter 10 discusses whether this should also be the case with very simple organisms.) Ultimately, then, coupling results from the ability of subjects to perceive qualities and to un/bundle them. This can be seen as the effect of an ’individualizing force’, insofar as each core-subject is in fact assumed to be its own.

8.4.2 Conservation laws and causal closure

A second important line of argument against the possibility of mental causation deals with the idea of a causal closure of the physical world, or at least the incompatibility of mental causation with the known conservation laws, especially for the quantities energy and entropy. Amongst others, Gibb [169] offers a first introduction to the English-language literature on conservation laws. In German-speaking countries, the topic of causal closure is discussed, for example, in the context of the so-called 'Bierri trilemma' [170] and on the basis of Brüntrup's [171] elaboration of the resulting decision alternatives.

The Bierri trilemma is a formulation of the classical mind-body problem in the form of three – seemingly irreconcilable – assumptions about our reality: A) The difference between mental and physical entities; B) the causal closure of the physical world; and C) the possibility of mental causation. Brüntrup has shown in detail the views to which the affirmation or denial of the respective parts of the trilemma lead us. Voigt argues, [3] that this classical trilemma should be thought of in a broader sense by not assuming a physical-causal closure, but rather a metaphysical-causal closure, which can do more justice to idealistic or panpsychistic models.

On the basis also of the discussion in the English-speaking world, it seems clear that the causal closure of the physical world in the sense of a clockwork (or more modern; in the sense of 'particle billards') is not an analytical result, but first and foremost an assertion and thus more a 'primeval dream' of materialism than an empirical finding of the natural sciences. Accordingly, in the history of the natural sciences, the conservation of energy has been a moving target for which we have found again and again new types of energy that had to be included in the accounting.

Even versions of this idea that are based on an overdetermination of processes by both physical and mental factors hardly seem tenable in view of the statistical and practically 'immaterial' nature of processes on the microscale. Here the world is entirely in flux and the most improbable is possible as long as the consistency rules behind the conserved quantities are observed. (We will see in chapter 10, however, that these rules still provide very tight constraints on how mental causality could be organized, even in a non-deterministic 'quantum world').

Models for explaining mental causation should make compliance with this seemingly necessary 'correct accounting' of physical processes a basic assumption already because our understanding of mental causation goes hand in hand with the expectation that our volitions reliably 'cascade' in the phys-

[3] U. Voigt, lecture 'Einführung in die Philosophie des Geistes', summer semester 2023, University of Augsburg.

ical world and thus lead to predictable results. (And ultimately, in Model A, it is the consistency conditions of the material world that give things their identity.) If one wanted to allow even small deviations, one would also have to ask why evolutionary processes would not have jumped at these possibilities in order to accomplish the 'supernatural'.

As early as 1925, Broad [172] argued that mental causation would require no more than the redistribution of energy in the brain; more recently, Gibb [104, 169], among others, has updated this argument. For mental causation on the meso-scale, however, it remains to be clarified what a corresponding coupling mechanism might look like and – even more unclear and therefore more urgent – how such a coupling mechanism should have evolved. Model A makes suggestions for both, which only becomes problematic beyond the microscale when, for example, subjects in our subconsciousness are supposed to influence electrical charges or the like on the basis of mental information. Subjects are initially only able to exert a minimal influence, which must not be lost in the noise of the constantly occurring physical processes. Here we must assume for Model A that brains have been evolutionarily optimized for this task and are then forced to make certain assertions about the concrete functioning of our brain, some of which could be empirically accessible and thus falsifiable. This should clearly be seen as a strength of the model and will be discussed in more detail in Chapter 10.

8.5 Doesn't Model A have a completely different concept of perception and knowledge than philosophy and the sciences?

With regard to the concepts of perception and cognition, it must be noted that the intuitive grasp of qualitative building blocks by core subjects is not the process that is normally discussed in philosophy or cognitive psychology under these terms, as in the former not an entire living being is involved, but only the core subject as its most basic, 'atomic' building block.

In Model A, a proper process of perception is a wide bridge between the material world (the causal network of bundles whose properties are manipulated according to strict consistency rules, e.g. elementary particles) and our mental world (our world map), whereby some subjects of our subconsciousness manipulate bundles with both material and purely mental properties. The change of a material constellation, e.g. a visual signal, thus leads to corresponding changes of some non-material properties in the mixed bundles, which are accessible to sub-subjects in our subconsciousness and effected by these sub-subjects on the basis of evolutionarily learned rules. This change is now 'passed on' until it can also be consciously experienced at the 'highest' level as a forced sensory impression in our world map. (It makes less sense to say at this stage that it can be experienced by 'us', because although our consciousness corresponds to the combination of core subject and world map, by 'us' we generally mean the whole person, i.e. going beyond this, not only the interlinked overall bundle of body and mind, but also the entire team of core- and sub-subjects).

On the way 'up', these basic qualities, generated purely on the basis of material signals, are additionally processed and contextualized in our subconsciousness, so that at the highest level qualities are experienced already in terms of form and intentional structure, i.e. as self-contained objects with intentional possibilities. In Kant's sense, a pre-structuring intellect appears alongside our basic sensuality, which (re-)presents the world to our reason, or in Model A more generally, to our conscious thinking. (Apparently not contingent, but only so for the highest level, which is not directly aware anymore of the evolutionary origin of the underlying processes.) Once a certain level has been reached, for instance that of concrete objects, higher abstractions in mathematics, or similar, the respective (sub-)subject can act freely on the qualitites, within the framework of the possibilities available to it.

Model A thus describes – seen from 'the top' – the effective dualism, which we indeed experience in everyday life, but with a strong empirical component: The complex representations in our world map are built up from simple universals as the 'atoms' of our thinking, but the assignment of these complex representations to material causal connections is evolutionarily learned and thus by no means unquestionable. Even the thought processes at the very 'highest' level are still indirectly shaped by evolution, amongst other things by what is set as rational in the world (see also the topic of logic and mathematics below). The whole object only arises through the combination of representation and the associated causal network, and is thus, in accordance with Kant, always already pre-structured by our perception. Due to the consistency rules of the material world and the evolutionary growth of us living beings in this world, we can trust our most basic perceptions to the extent that we are condemned to do so anyway. This allows us to draw indirect conclusions about the underlying causal network – except that these can never be completely immune to sceptical attacks; our goal can only be a 'best overall fit'.

In Modell A we find our representation of the world quite concretely as bundles of universals in our world map. These bundles include intentional possibilities ('can be thrown', 'is worth striving for'), but the core of our intentionality arises from the focus of the core subject on partial bundles of the whole world map, which then occupy our entire conscious perception. The presumably larger part of our beliefs, but probably not the larger part of our wishful thinking, would fall within the 'remit' of the core subject. With regard to our wishful thinking and general affects, at least some sub-subjects would most likely play important roles, especially since they would also perceive bundles, including 'beliefs', that would not be directly accessible to the core subject. A large part of our perception would therefore be completely subconscious. And even complex bundles that are perceived could be structured in such a way that essential properties would not be obvious without reflection, i.e. unbundling. (Sense impressions do not seem to mix like sets of particles, but rather follow a mechanism of substitution; more on this later.)

What is central to Model A is that with our world map we do not have an 'interface' to the material world, but already one side of the objects. Due to the evolutionary, ontogenetic, social, historical and (individual) life-historical development of our world maps, our perception and cognition are always already pre-structured in many ways (especially linguistically also), but this can be reflected to a large extent, and in some cases even corrected. The most fundamental, 'atomic' act of our perception is the creative combination of changes in qualities (e.g. material signals) with new qualities (e.g. colors). This allows for intersubjective understanding and objectification in connection with the universality of the qualitative building blocks and sufficient correspondences in our world maps constructed from them. Meaning is essentially universal, but (inter-)subjectively bundled and contextualized (more on this below under the topic of language).

The extent to which compelling connections, i.e. a kind of 'semantic logic', allow us to draw conclusions before any empiricism, at least between the non-material building blocks of our world-map maybe, will be examined further below under the topic of logic and mathematics and then in Chapter 11; it should, however, be clear that Model A is entirely dependent on empiricism with regard to the material world, but also neuroscience and psychology. Model A is a suggestion as to how an overall design could look like; it is nevertheless essential that previous observations are captured and that predictions to be made are empirically confirmed. The proposed objective idealism is not a rationalistic maximum one, but a pragmatic minimum one, without eternal truths that can be deduced purely by thinking.

With regard to the modern philosophical discussion of perception, the above model can be understood as a 'sense-data theory'. It can provide meaningful explanations for veridical, illusory and hallucinatory perceptions. The model can be defended against alternative theories (such as naive realism, disjunctivism, intentionalism, etc.) in analogy to Howard Robinson's [173] outline, without, however, having to advocate like him a theistic idealism.

Finally, there is obviously much more in Model A that needs to be found out about the processes of perception and cognition outlined above. In particular with regard to the subconscious pre-structuring and contextualization of the bundles that finally reach our world maps, philosophical phenomenology, and here in particular Husserl, could probably be of great help to cognitive psychology in Model A: From the perspective of the model, Husserl's phenomenology can be understood as an attempt to describe the regularities of our world maps, our 'horizons' and '*Lebenswelten*' (lifeworlds), as well as their subconscious generation. However, the extent to which central concepts such as *Ideation* or *Fundierung* could be transferred would still have to be examined in detail.

In any case, the development of a 'grammar for the formation of meaning' (*Grammatik der Sinnbildung*) would be a central goal for Model A as for Husserl. The central observation of a structuration of parts by a whole that results from its parts, for example, can also be found in Model A on another level in the creative connection of qualities with processes for which the qualities can then be used as regulative targets (see also Chapter 10). Further parallel observations would be, for example, that, as in Husserl, different phenomenological objects can supervene on the same causal connections, that the freedom of the transcendental ego or core-subject lies in the choice of its attentionalities, that intersubjectivity ultimately depends on empathy, and so on. Since in Model A our world maps should also be able to encode temporal dependencies, the model is conversely perhaps not entirely uninteresting for phenomenology, for example with regard to the 'holistic' perception of temporally distributed phenomena such as hearing a melody as a whole, which in Model A can be indeed present as such a whole (bundle) in my world map. Whatever benefits Model A, cognitive psychology and phenomenology might have for each other; from the point of view of Model A, this is a question of empiricism.

8.6 Does the model explain too much too simply?

If you have struggled this far, you may get the impression that the model is 'too good'. That it puts us in a very comfortable position, but one that is perhaps too comfortable to really contribute anything: In the natural sciences, everything fundamental is attributed to identity requirements, everything specific (particle masses etc.) to contingent evolutionary processes. (The 'particle zoo' must be structured in such a way that the microscale can be the basis of individuated objects on the mesoscale; however, there are practically infinite possibilities for this structuring, so that the structure actually observed can only be a contingent product of the extended evolutionary process.) The question here is whether there can be empirically verifiable conclusions that are not already known or otherwise explainable (see also the previous discussion in chapter 6). Even beyond the natural sciences, the model seems to collect the problems 'as is' and then ask the question 'What is the simplest model that explains most of the effects?', instead of genuinely solving any of those problems. If it really were a more accurate description of our reality, there would be a whole series of philosophical questions for which we could simply return to 'older' answers. Figure 8.2 attempts to provide an initial, graphical overview.

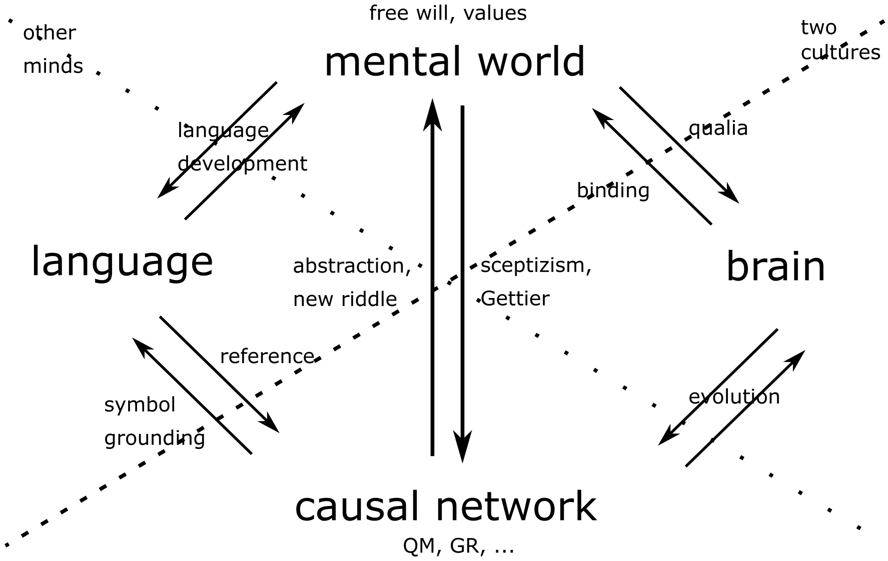

Figure 8.2: Overview of philosophical and scientific problems related to the mind/matter problem.

A large number of scientific and philosophical problems are framed by our ideas about the connection between the physical and mental worlds. The most important interfaces here are our body (in particular our brain) and our language. If I now consider the existence of an objectively existing mental world in order to explain the 'narrow' problem of how our brain functions, then I do also change the framework within which I can, or even must, discuss the other problems. In the area of the causal network of the physical world, for example, I have the possibility of rejecting a material understanding of the realism of micro-scale processes. Likewise, there are new (old) explanatory options with reference to the mental world.

In the case of problems relating to the interplay between these two worlds, the changed framework conditions can become particularly apparent. Here is one example:

Goodman's 'new ridle of induction' [163] should also be understood as posing the question of what distinguishes helpful from unhelpful hypotheses and why people are so good at making meaningful hypotheses. This question is very topical again in that the 'causal inference' approach, [58] for example, attempts to teach AI systems causality to escape the 'curse' of the uncertain conclusions from correlations on which DNNs are based. Unfortunately, machines still have a very hard time to automatically decide which possible hypotheses must be included in the 'causal graph' for a given problem. Model A provides an answer here (which in principle should even be realizable in the form of an 'organic' computing machine): Humans generate their hypotheses from the manipulation of (bundles of) universals, as outlined in Chapter 7.

Gettier's observation that knowledge cannot be understood as 'justified true belief' without further ado (since I could have acquired such defined knowledge on the basis of coincidences), may serve as an example of the fact that despite changing framework conditions sometimes no major change in the understanding of a problem occurs. Here, as with scepticism, the new approach does not solve the problem, but makes it at least well comprehensible, because the model makes it explicit, that objective knowledge is in need of a proper coordination between causal network and mental map. Model A thus describes a world in which the possibility and the practically necessary occurrence of scepticism and Gettier arguments are already inherent in the design.

The questions associated with the brain are examined further in Chapter 10, while those associated with the phenomena of language [174], logic and mathematics [175] can only be briefly touched on below due to their great complexity (though the problem of language development is briefly addressed again in Chapter 9).

8.6.1 Language in Model A

First of all, it should be noted that Model A is based on the assumption that thought also includes non-conceptual content, i.e. that thought comes before language; a successful refutation of this assumption, e.g. along the lines of McDowell, would render Model A invalid and considerably complicate the design of improved idealistic models of this kind. With regard to other discussions (internalism/externalism, etc.), Model A attempts to assume a mediating position; language functions here (necessarily) as an intersubjective bridge between a logically comprehensible physical and an initially at best paralogical mental world. Our initial symbol grounding problem is then closely linked to the problem of reference.

Model A suggests a multidimensional approach to defining the meaning and reference of words and sentences, i.e. more or less complex bundles. Meaning – Frege's '*Sinn*' – refers to elements of the individual's, developmentally and socially formed mental map of the world; reference – Frege's '*Bedeutung*' – optionally to other elements of the world map with or without reference to the causal network of the physical world. The connection of certain physical, e.g. audio-visual signals or signs with language elements is due to historically and socially developed conventions and – since it is essentially based on creative acts – is by no means pre-determined. At the next level, these language elements correspond to bundles in the world map, which can be identified as the meaning (with Frege the 'sense') of the language elements, but can also be referenced as language elements as such. Meanings can stand on their own or reference other elements of the world map. These may or may not be directly related to the causal network of the material world. If the latter is the case, the meaning references the nonmaterial part of physical objects, which is presented to us by our subconsciousness according to the underlying causal connections. Newly discovered, simple or complex entities in our world maps may then require new language elements.

For example, we use the auditory signal 'water' in accordance with historical-social convention for the complex bundle 'water' in our world map, which is part of other bundles of our world map and thus can reference its part in these bundles, of which for instance 'this water in the glass in front of me' is a concrete bundle in my world map that is connected to the causal network of the material world via my subconscious.

Unlike meaning, in Model A reference only arises from the context. The context of our language use is the sum of our world map and the assumed world maps of our interlocutors. Both substitution and indexicality can be represented in Model A in this way. Efficient language use is based on the (presumed) minimum required difference to the context shared across the world maps. The 'force of the argument' then results from the effort to achieve the – more or less vital – consistency of one's own world map, which is why no such force might be felt in the case of agonal differences.

The construction of our world maps is a combined biologically (evolutionärily, ontogenetically, ...) and socially (historically, psychologically, ...) embedded process: My map of the world is formed on the basis of my subjective sensory impressions and, from a certain age, also by linguistically conveyed ideas. Sensory impressions and language can only ever provide the subconsciousness and the core-subject with the material on the basis of which the next step is possible, namely the intuitive recognition and then use of a new non-material building block. There is a high evolutionary pressure on the alignment of the world map and the causal network, which results in a relatively fixed 'wiring' in the subconsciousness for the translation of signals into sense perceptions. On the basis of this shared alignment with the 'outside world', the further intersubjective alignment of world maps via language can then be halfway successful. (It takes a lot of effort, on the basis of such functioning world maps, to come up with the idea that Gagavai might not mean the hare).

Our world maps therefore not only provide our thinking with all the more or less shared prejudices of our '*Lebenswelt*', including the possibility that linguistically mediated thinking is channeled by linguistic conditions according to Sapir-Whorf. Conversely, they also contain a lot of potential for relativism, holism and 'language games', which in Model A, however, must have a (for human standards) reliable framework of certain physical and biological fundamentals. The possibility of the success of matching processes is ultimately based on the universality of the basic building blocks. With regard to our use of language, the above fits in with the fact that semantics and pragmatics are always already intertwined against the background of individual world maps and that linguistic context-setting ('framing') is as effective as it is, due to an almost automatic accommodation of such maneuvers, which are by no means always advantageous for the 'recipient'.

A special role is played by elements that are intersubjectively coordinated across our individual world maps (although probably never perfectly) and thus appear objectified to us, i.e. 'are a thing'. These can be, for example, complex social constructs such as social institutions, the idea of money, or more simply, also proper names, which are only conditionally 'ridig' in the model. Each such construction is subjectively combined with other elements in the individual world-map, which means they are both subjectively shaped in terms of content and subjectively contextualized, especially in terms of relevance. But social paradigms, blind spots, incentives and constraints act here like a 'social physics' and give these intrinsically immaterial objects a quasi-material stability and reliability. However, unlike in physics proper, these basic structures can easily undergo reformatory or even revolutionary developments on human time scales. Deeply embedded 'distortions' of our world maps may nevertheless stand in the way of overcoming for instance existing power structures (more on this in Chapter 11).

The most important evolutionary aspect is the resulting error tolerance of the system, which allows the processing and reflection of inconsistent information: *Was nicht (in die Weltkarte) passt, wird passend gemacht.* ('What does not fit (into the world map) is made to fit'). The observation of universal and innate aspects of language would then result from a coupled brain/mind development, which could also explain the existence of a critical period for language development and possible limitations due to brain lesions; more on this in chapters 9 and 10.

If one compares language in Model A with the ideas of the analytical philosophy of language from Frege via Russell, Wittgenstein, Quine, Davidson, Kripke to Kaplan, Montague, Stalnaker etc., the model appears to be largely compatible with current discussions, even if not positions. A central difference, at least to the thinkers after Frege, is certainly the special conception of meaning here as an 'idea' to be grasped intuitively and not as a lexical meaning, which, however, avoids the 'rule regress', holism and other problems. In contrast to complex linguistic bundles of meaning, primitive building blocks of meaning such as the core ideas of true or good cannot be further analyzed, but can only be paraphrased or 'shown' in the same way as colors. We can, however, analyze the complex bundle that is our everyday notion of true or good. And we should by no means fall into quietism with regard to fundamental ideas: We can certainly talk meaningfully 'around' what we cannot directly talk 'about'; make meaningful paraphrases, ultimately 'point' to ideas verbally, i.e. create the context in which the other person can take the implied step towards the intended building block.

In Model A, too, historical-causal developments determine the concrete contexts of world, thought and language. However, these developments are only possible through the objective existence of fundamental 'ideas', on the basis of which problems of radical translation or interpretation can then be traced back to insufficient commonalities in the world maps of the speakers. Especially with regard to the problem of pragmatics, the known analytical models of language do not appear to be better suited *per se* to provide solutions to all open problems, so that here too the question remains as to whether Model A as a whole (beyond language) is understood as an attractive alternative. Defining linguistic meaning via possible worlds does not seem hopeless, but an analogue for defining the 'meaning' of sensory qualities such as colors seems substantially less promising.

Counterfactual or modal considerations can be realized in Model A via the addition of (bundles of) universals to (parts of) the world map. This gives a very concrete and ontologically parsimonious answer to the question of what a 'possible world' is. Due to the paralogical nature of the mental, real contradictions are in principle also realizable here, since they would be without physical-causal consequences, so that correct (re-)thinking, i.e. thinking oriented towards the consistency rules of the material world, can become necessary. (Among other things, Leibniz's idea of compossibility does not apply here; the existence of objects cannot automatically negate the independent existence of other objects. With Hume the claim is that the opposite of every fact is conceivable without contradiction. And 'nothing' can also be a qualitative building block here.)

It could be discussed whether this construction could also provide a solution, or rather a psychological explanation, for the problem of material implication as 'deductive explosion': The logical conclusion meant by this is oriented towards the consistency rules derived from the material world, whereas in normal usage the possible world 'constructed' on the basis of false premises is directly rejected as a whole, because either way it is in conflict with the world map, which – in particular also for the evaluation of counterfactuals – must be kept as consistent as possible and in the best possible agreement with the individual's social and physical environment. Logical reasoning applies within the possible world, while everyday language looks at it from the outside. (The illogical behavior of people in Wason selection tasks [176] could perhaps be explained in a similar way). However, the possibility we are given of 'life' in fictional worlds, including the phenomenon of imaginative resistance to 'inappropriate' elements, shows that we can also immerse ourselves completely in such possible worlds.

An important 'prediction' of Model A would be that even possible worlds are initially always thought context-sensitively; every consideration, and thus also counterfactuals, are made against the background of the concrete world map. As already mentioned above, this is of great advantage at the semantics/pragmatics interface: Necessary pragmatic supplementations are automatically made against this background; but this only works if the interlocutors' world maps are sufficiently similar. Holistic aspects of language arise through the necessary embedding of language in common world views, aspects of ambiguity arise through the additionally necessary coordination with the physical world.

A whole series of answers would still have to be found before we could speak of a theory of linguistic meaning for A-world. But just as we could probably use a lot from Husserl for Model A with regard to the problem of the structure and organization of our world maps, we could probably learn a lot from Charles Sanders Pierce with regard to a theory of linguistic meaning for Model A. It is already favorable that Pierce, with his semiotics, strives for a general, not only linguistic theory for the connection between signs and concrete as well as abstract objects, since meaning in Model A is not only to be understood linguistically. However, while for Pierce the transitive identity of objects is central (also as a normative assumption of his pragmatism) and thus everything (including human thought) becomes a relation in context, Model A assumes, as explained above, that this only applies to signs, which at some point require substantial identities as reference points.

8.6.2 Logic and mathematics in Model A

The central characteristic of logic and mathematics (not necessarily as a sub-area of conceptual understanding; think of visual techniques in mathematics), would be in Model A their derivation from the consistency rules of the physical world – also and especially with the aim of being able to theoretically grasp the complexities of this world. (Which in turn could explain Wigner's observation of the unreasonable effectiveness of mathematics in the natural sciences). [177]

While the existence of logical-mathematical entities is readily accepted in Model A, there is no mechanism on the basis of which one could argue for compelling connections between these entities. (Accordingly, it would also have to be argued with Harman that rationality cannot be exhausted in deductive logic). For this there would have to be mental-causal necessities, which, like physical-causal necessities in the model, would all have to be traced back to the actions of subjects. Alternatively, one could assume a 'semantic logic' between mathematical entities, so that by recognizing the meaning of the entities, subjects are already given compelling connections with, and thus action guidelines for other entities. However, since in Model A, for the sake of consistency, the existence of such a semantic logic would have to be assumed not only for logical-mathematical entities, but for all qualitative building blocks, and since this does not seem to be the case in many areas, it seems more consistent and 'economical' to assume here that logical-mathematical regularities can be derived from their use to describe 'meaningful' relationships for us. Unusual logical-mathematical ideas such as 'round squares' can then simply be understood as a set of building blocks that are not necessarily mutually exclusive in the mind. The possibility of a semantic logic, or perhaps rather 'aesthetics', will be examined further in chapter 11. Somewhat unusually, in Model A we then find the mental world as initially non-rational and the physical world to be rationally structured due to extended evolution; which in Model A results from the harmony of metaphysical and logical principles in the evolution of the cosmos.

Between Plato and Mill, Model A is thus clearly on the side of Plato in the philosophy of mathematics, albeit with a Humean, empirical flavor, in the sense that mathematical relations do not arise directly from empirical experience, but still result from experience via the selection of 'permissible' relations, i.e. those that are somehow meaningful for the physical world. The assertion is thus that even where mathematics goes its own way, it still ultimately refers to 'physics' in its negation of it. But this is a different synthesis of rationalism and empiricism than that of Kant, in which the synthetic *a priori* in mathematics is derived from 'pure' intuition through the channeling of sensory perceptions:

In Model A, 'pure' intuition is not fixed in principle, but is conditioned by the world map, which can be extended to include new concepts, e.g. non-Euclidean geometries. Within our world maps, we can seemingly reason '*a priori*', since our idea of rationality derived from the material world must appear to us to be without alternative. But even this 'analytical' reasoning is ultimately empirically informed, which also means that it is by no means purely linguistic in nature. In essence, the derivation of logic and mathematics is thus a creative act, so that it is not entirely unexpected that previous attempts to systematically demonstrate their foundations run into the problems identified by Gödel. One could speculate somewhat more boldly as to whether the outlined connection with the consistency conditions of the material world provides the reason for the special position of first-order predicate logic, which, unlike higher order logic, quantifies over particular objects (here; individuated bundles) instead of universal properties, and thus offers itself as the 'logic of the material world'.

It should also be noted that the application of mathematics in Model A does not build a bridge from – objective, but subjectively contextualized – mathematical entities directly to the causal network of the material world, but from those mathematical entities to other entities in our world map, namely the mental, always already idealized side of the hybrid structures that make up physical objects. The bridge is thus built between two systems of rules; that of material qualities, e.g. for positioning in space, and a system of purely mental qualities selected and logically designed according to the intended use, e.g. the real numbers, for whose use certain consistency rules are prescribed and with which in this example a coordinate system can then be formulated.

Finally, it should be discussed whether the model can propose solutions to Benacerraf's problems [70] of epistemic accessibility and identifiability or whether it is not already based on an answer to these problems, which are often cited as central arguments against Platonic realism in mathematics: In Model A, Mathematical identities as such are epistemologically accessible (mentally-causally embedded) and open to flexible identification with referents. Each concrete mathematical identity is already a bundle of more fundamental entities.

According to Model A, some of the proposed solutions – not only with regard to logic and mathematics – appear too simple only because they do not have to deal with the paradoxes established in the current discourse, based on the assumption of a purely physical world. All in all, the question of whether the model explains too much too simply appears to be justified; it cannot be answered conclusively here. In a way, however, it does not seem surprising that a change in the fundamental world view must always lead to many new (old) answers. Both science and philosophy have spent several centuries filling in the gaps in the materialistic world view and mapping out the problems associated with it. The fact that a synopsis of these problems leads us to a new model that addresses precisely these problems at least in part should be seen as a strength, perhaps not of the previous model, but at least of scientific and philosophical work to date. Furthermore, the explanatory power of idealism in the field of philosophy of mind was certainly an important reason why humanity was never able to leave it behind completely over two and a half millennia.

Overall, Model A appears to be compatible with current philosophical discussions across a whole range of issues and on par with accepted positions in modern philosophy. The central open questions are then about the concrete structuring on the microscale (to which chapter 6 was dedicated), as well as the connection between the human psyche (chapter 9) and the human brain, including its evolutionary and individual development (chapter 10).

Chapter 9

Counterarguments from psychology

An alternative to materialism in the sense of Model A would most likely have its greatest practical impact on psychology and neuroscience, which is why this and the next chapter are dedicated to these subject areas. In the necessary 'reconstruction' of psychology in Model A, the most imminent danger seems to be a slide into esotericism, i.e. the setting of entities or relations motivated by something other than rationality, which should be avoided at all costs. There are two fantasies in particular that Model A will involuntarily evoke in many people: That of non-physically-anchored individuals like angels or ghosts, possibly linked to the idea that our physical life is something like a larval stage for sufficiently individuated spirit beings at some point; and that of the transfer of our mind to other physical anchors like artificial bodies or a 'matrix'. An argument against the first is that in Model A individuals without physical anchoring are extremely unstable; an argument against the second is that an extremely complex coordination of mind and matter would first have to be understood and then transferred largely error-free. At this point, let us assume that both seem at least practically impossible. (It must be admitted, however, that the local manipulation of the set of rules according to which micro-subjects maintain the physical world, which is conceivable in principle, could lead to technologies that would appear as magical to us as our current technology would to an inhabitant of the Middle Ages).

Thinking even further, we could speculate about complex mental structures on the basis of less integrated physical structures, for instance whether, in the sense of animism, seemingly inanimate things might not also have a proper mind. In Model A, however, complex mental structures only arise from the *interplay* of necessarily increasingly complex physical and mental processes, which is why this also seems very unlikely at the very least. But even apart from these imagined extremes, we must be careful not to use Model A to advocate a misguided lay psychology. In this sense, the following is to be understood as a first attempt to investigate what a psychology in Model A might look like, but which would then always require further, above all empirical investigations before it could be claimed in this way.

9.1 Is Model A compatible with the basic concepts of psychology? – Modules, information processing, psychological mechanisms of mental causation

We have already seen in Chapter 4 that two paradigms in particular can be regarded as central to modern psychology: [178] Firstly, the thesis of the modularity of the mind, which is to be understood as composed of different interacting units, [1] and the information processing paradigm, namely that these units and their interaction can be explained by information processing in the nervous system. Model A is compatible with these paradigms as long as they are not understood in a purely materialistic sense, namely that the modules only correspond to (possibly dynamic) brain structures and that the processed information is of a purely quantitative nature. In Model A, sub-subjects are added as modules that can also process information of qualitative nature. The first important implication is that, in addition to a purely physical-functional modularization (neuronal aspects of vision, motor skills, etc.), there is a psychological modularization of the subconscious: Model A requires that we do not have a monolithic subconsciousness, but a polyphonic one. If a subject alone could bridge the gap between mind and body, there would be no need for a subconsciousness; that this can hardly be the case will become clearer in chapter 10.

Against the background of our current scientific world view, the talk of several sub-subjects will appear unscientific to quite a few people. Ultimately, however, it comes quite close to the established idea of modules that exchange information. Nevertheless, it would perhaps be better to avoid talking about sub-subjects in this context, because a subject in psychology, as well as in everyday life, is much more than what is referred to here with the term sub-

[1] Ideas of modularity have accompanied psychology from the very beginning, from Freud's *Instanzenmodell* of the psyche to Piaget's stages of development, the 'core cognition hypothesis' in evolutionary psychology especially with regard to the cognitive performance of animals, to Chomsky's modular 'organology' as an explanation of the complex language ability of humans.

subject. However, we will continue to work with this term in the following, at least for the time being. In Model A, the human being would in any case be seen as a holobiont, i.e. a whole living being made up of individual organisms, not only with regard to its microbiome and other known biomes, but also to its 'psychobiome'. A second important difference to the established view is that some modules would only be partially organized under a uniform survival principle such as the 'predictive regulation of behaviour' or similar, but would instead bring an irreducible, unconscious irrationality, but also creativity, into the 'team'. Those parts of our subconsciousness would be able to an extent, however limited, to find and pursue their own ideas. Which would then lead, among other things, to the observed phenomenon of a possible weakness of will of the whole person.

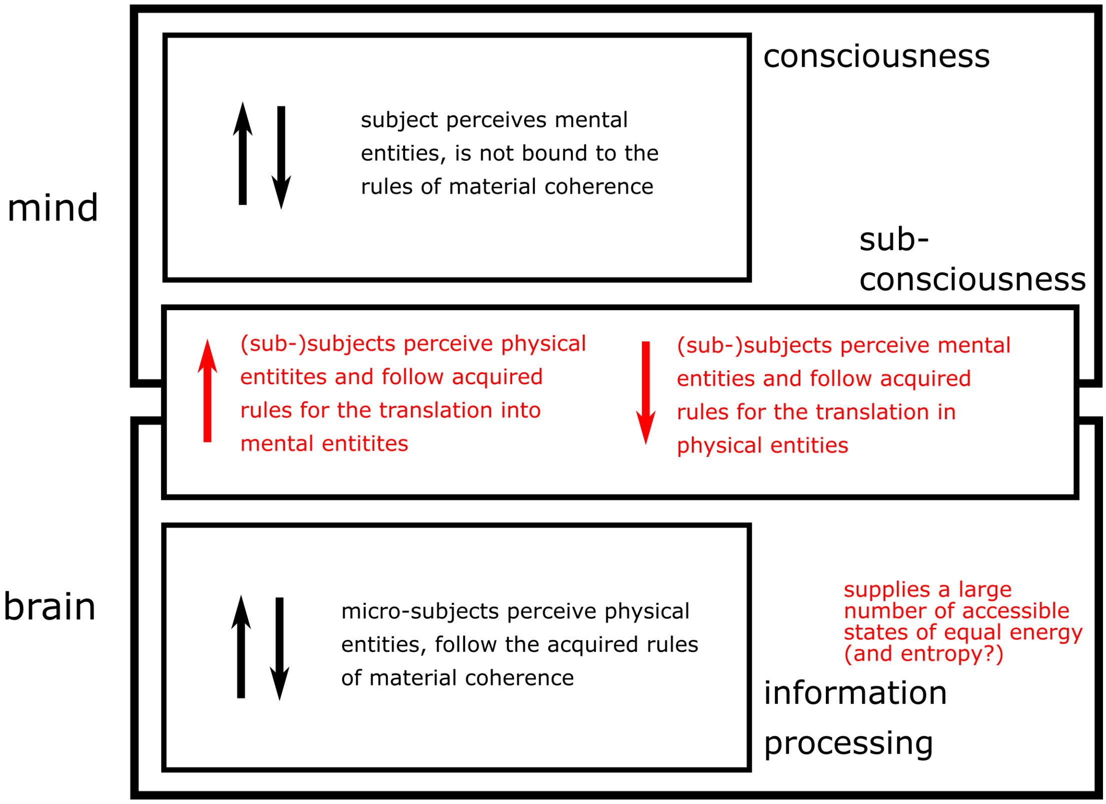
consciousness
subject perceives mental entities, is not bound to the rules of material coherence
mind
sub-consciousness
(sub-)subjects perceive physical entitites and follow acquired rules for the translation into mental entitites
(sub-)subjects perceive mental entities and follow acquired rules for the translation in physical entities
brain
micro-subjects perceive physical entities, follow the acquired rules of material coherence
supplies a large number of accessible states of equal energy (and entropy?)
information processing

Figure 9.1: Schematic representation of the interaction between brain and mind in Model A, including the (sub-/micro-)subjects involved.

Figure 9.1 shows a schematic representation of the interaction between brain and mind in Model A, including the (sub-/micro-)subjects involved. At the fuzzy interface, sub-subjects operate with bundles of physical properties as well as mental entities like colors or shapes. The latter can simultaneously be part of the world maps of other subjects, including the core subject. (There is an interesting parallel here with the medieval psychology of faculties of mind sharing forms. However, Ockham's critical distinction between mental action and mental content can be accounted for in Model A via the core subject and its world map). A critical aspect of this notion is that the brain would have to provide a multiplicity of accessible physical states of equal energy, so that sub-subjects could read out the core-subject's volitions from shared mental entities and then physically realize them, i.e. make mental causation possible. The next chapter examines whether this is conceivable in accordance with the findings of modern neuroscience, or whether it could perhaps even contribute to the explanation of certain phenomena. In any case, the result is a three-layered structure of brain, subconscious and conscious mind, as has been suggested by many others, including for instance Augustyn. [179]

The central importance of sub-subjects for the functioning of the brain/mind system puts a 'balance of souls' (i.e. sub-subjects) instead of a 'balance of chemicals' in the focus of possible considerations for the treatment of mental problems. Although one could argue that both ideas have the unfortunate antique-medieval flavor of the 'balance of humors' – which was actually already an important step beyond the alternative of divine causation –, chemicals can produce a proven track record in medicine. So, before we fall into the trap of trying to explain all psychiatric illnesses by faulty developments in the balance of the sub-subjects (and I think the potential is indeed huge), we must be careful not to rationalize beyond empiricism. (More considerations on mental illness and forms of therapy can be found below.)

Nevertheless, the talk of sub-subjects and their balance is not in itself less scientific than the established discourse: Both sub-subjects and non-physically bound, higher modules can currently only be identified via their function. And the fact that in Model A subjects bridge the physical and the mental world is just as little magical as the fact that neuroscience can only relate brain activities and experiences to each other on the basis of the testimony of subjects.

An interesting option for Model A is that communication between (sub-)subjects could explain the existence and functional mechanism of feelings as qualia: For the inquiry of complex, non-reflexive behavioral changes of the whole organism, the most abstract markers possible are favorable, which must then be made accessible to the core subject as mental entities. Sub-subjects cannot provide the core subject with purely quantitative information, which in many cases would also be too specific to demand complex behavioral modifications that go beyond the instinctive-reflexive. For example, we have hunger, but (normally) do not show any reflexive processing of this feeling, but (usually) have the opportunity to approach food intake with a good dose of rationality, which should be seen as an evolutionary advantage overall. In Model A, it is precisely the underdetermination that gives the subjects room for maneuver.

9.2 Does Model A provide a realistic picture of the basic functions of our mind? – Consciousness, private access, (sensory) Experiences, intentionality

Model A also succeeds in meaningfully reconstructing the basic functions of our mind: Subjects are – albeit by definition – conscious and have private access to their 'world map', but beyond this their access to the entire living being is limited, as interaction with the causal network of the physical world and other subjects can only take place indirectly via a complex subconsciousness. (This would then also correspond to the observation made since antiquity – by Augustine, Albertus Magnus and Thomas Aquinas, amongst others – that we are to us so close, yet so inexplicable; that the injunction 'know thyself' is not a paradox; that arguments such as Avicenna's flying man can be made.) In Model A, moreover, experiences are always qualitative and intentional in nature, because operating with the world map corresponds to the subject focusing on bundles of qualities.

9.3 Can a realistic picture of the performance of our mind be derived from Model A? – Attention, perception, memory, cognition, learning, psychomotor skills, language

In Model A, the core subject has a 'floating' attention even in a non-occupied state, since it cannot perceive itself as completely context-free, detached from the world map; the core subject itself is 'empty', devoid of any quality. If it has perceptions that are triggered by the arrival of 'individual images' of quantitative information, these are already integrated into a continuous 'movie' by the sub-subjects: After the initial creation of a mental entity, it exists continuously until further information causes (steady, i.e. 'judder-free') changes to the entity. In Model A, mental performance will thus also depend on the number and 'competence' of the sub-subjects. How many sub-subjects form a psychobiome is, like the established question of how many independent modules should be assumed, open to empirical investigation. Various observations can provide clues, including how many qualitative information units can be processed simultaneously (the 'Miller number' for this is 7+/-2, but today is assumed to be 3-4 on average), but sub-subjects would not only be involved in the cognitive processing of conscious information, so that one must probably assume many more sub-subjects in total.

Functioning memory mechanisms are central to the cognitive performance of humans, especially in comparison to other organisms. In terms of the idea of extended evolution, it can be assumed that the micro-subjects of physical evolution follow acquired rules without a memory of their actions, which makes them extremely inflexible, but therefore also extremely reliable. For the organisms developing in biological evolution, increasingly complex memory mechanisms can then be assumed, which not only make them more competent through the possibility of learning, but also increasingly flexible and ultimately freer in the further course of evolution. The thesis that humans, unlike simpler organisms, are characterized by the fact that they operate with a world map that is only indirectly coupled, would then also include the fact that humans can work with particularly efficient memory mechanisms: While the memory of simple organisms must be physically implemented in their neuronal system, higher organisms should then be able to make more and more use of the possibility of organizing memory functions in a non-material and semantically structured way. The human world map is therefore as already mentioned not to be thought of in purely spatial terms; it would be more like a semantic network that amongst others also encodes temporal relations.

For the sake of efficiency, however, this coding will be as abstract as possible, so that it should come as no surprise that human memory works in a highly intentional way and is unreliable in detail. (Though this is no different in the established models; the change/overwriting of information through the influx of further input can be understood both through storage in neural networks and through the use of universals). In particular, it seems likely that the details of sensory qualities are not stored as such – except maybe under special, e.g. strongly emotional circumstances –, but only as abstract markers, which when we remember them would then require a renewed activation of the sensory centers in the brain for the 'completion' of the remembered facts. Nevertheless, there may be individuals for which this mechanism works less efficiently or at least differently, which on the other hand could perhaps enable special skills; think of people with absolute hearing, for instance. (From a philosophical point of view, the conception of memory outlined above is interesting because it can answer the question of the nature of past and future, which also the initially presentist Model A raises, by arguing that past and future are structuring elements of human world maps).

In addition to this coding at the 'highest', purely mental level, the human memory must also include physically and mentally encoded elements (especially for sensory information), as well as purely physically encoded one (especially for psychomotor skills), i.e. it must continue to use also the evolutionary earlier memory mechanisms. The usual distinction between short-term and long-term memory is likely to be orthogonal to this stratification.

If we look not only at the unusual mental capacities of humans, but also at where they find a biological limit that is relevant for psychology, our focus shifts to mental exhaustion, the need for sleep and (e.g. intoxication- or drug-induced) 'blackouts'. These are all phenomena that are currently not well understood; here the model allows interesting hypotheses to be put forward (which still would have to prove themselves empirically!): Assuming that the core subject initially manipulates the world map as freely as possible, more and more conflicts between the map and the underlying causal network would accumulate over time. This could explain the feeling of mental exhaustion as 'emotional friction' or 'ego depletion', as well as the evolutionary benefit of a mechanism for the temporary isolation of the core subject during sleep as a period for 'repairs' to the world map. Which could then be partly accessible to us as dreams, but which we should not imagine as a simple correction of the world map, but in which – due to the inherent creativity of the sub-subjects – new things would also emerge. The isolation of the core subject would have to refer both to the sensory and to all memory mechanisms, which implies that the purely mental structures are not meaningfully accessible to the core subject without their sensory anchoring, i.e. without a 'painted' world map. Sleep should thus be understood as 'blacking out', not as 'letting go' of the perceived world map by the core subject.

Another fundamental characteristic of the human psyche in Model A would be a spectrum of thinking from fast, instinctive-emotional to slow, deliberate-logical conclusions, similar to that propagated by Kahneman (see also Chapter 4). The first mode ranges from purely reflexive, physical information processing to the subconscious actions of sub-subjects that appear instinctive or emotional to us. The second mode then covers the conscious manipulation of the world map by the core subject. The irrationalities of the subconscious that can be recognized by the core subject in reflection would then have to be attributed above all to the limited horizon of the sub-subjects. (With systems theory, much of psychology could presumably be understood as 'subsystem optimization'). In recognizing the importance of the sub-subjects for the core subject, Model A clearly goes beyond the usually strongly rationalistic model of the human mind in classical idealism and locates the mind in dependence on the subconscious in a body and a concrete situation. Model A is thus well suited to the '4E' ideas of an embedded, embodied, extended and enactive mind.

The special cognitive capacity of humans (quick comprehension, creative problem solving, etc.) would result from the possibility of manipulating (bundles of) universals as described in chapter 7, but always in the context of a body and its concrete situation. This would also gives rise to the particular cognitive susceptibility of humans to prejudice; everything is always conceived against the background of the individual world map. This includes other humans, who are initially only bundles in my world map and whom I can only experience as persons through empathy.

A central 'prediction' of Model A is that cognitive learning processes must ultimately originate from the individual: If organisms are fundamentally anchored in the physical world, we must always think of the exchange of information as being mediated via the physical world. The 'translation' of physical signals takes place in the subconsciousness, which, however, only has limited possibilities (and must not have more!) to evoke qualities in the world map of the core subject. If something new, possibly quite abstract. is to be learned, the qualities evoked must be sufficiently suggestive of this newness in the context established so far, so that the core subject itself can then incorporate it into its world map. In this process, universals can serve both as a goal and as a bridge. The world map, i.e. the already existing 'knowledge' of the individual, thus channels the further acquisition of knowledge, and also allows effective social learning through the common access to universals, as long as the backgrounds of the learners are sufficiently similar. From the perspective of the model, the symbol grounding problem therefore appears to be the 'materialistic fallacy' of falsely assumung that quantitative information could transport qualitative information instead of just being the basis for its generation.

In addition to the 'know that', structured probably like a semantic network in the mind, we would usually also find a physically and/or subconsciously organized 'know how'. Unlike the learning of semantic content, know how (think of playing a musical instrument, for example) requires the establishment of physical structures in the brain and beyond. Here, the idea of the activity in the world map can serve as an important target, by means of which stable feedback can be given for the initially awkward attempts, until corresponding structures have formed. If the complex behavior is learned, i.e. physically and/or subconsciously coded, it will be more of a hindrance to consciously intervene again.

It therefore seems likely that most psychomotor processes are physically anchored at some point in the course of life as more or less complete programs that run automatically in response to given environmental cues. (Think of grasping, walking, but also brushing one's teeth in the morning, etc.) The possibilities of the core subject would then primarily consist of cybernetic control through feedback, in the form of (micro-)inhibitions of such automatisms, while real behavior modification, such as learning new behavior, would have to be conscious and therefore laborious. Against this background, the Libet experiments that neuronal activities can precede conscious decisions could be understood: Our brain must always already offer the sub-subjects and the core subject behavioral patterns or action templates appropriate to the respective context so that they can make decisions.

A very impressive 'achievement' of humans is their language ability, the philosophical aspects of which were already briefly touched on in the previous chapter. With regard to the psychology of language and, in particular, language learning, it should be added that with Model A, in the sense of a continuous extended evolution, we would have to assume that the faculty of language in the broad sense (FLB) according to Hauser, Chomsky and Fitch [180] is not exclusive to humans, since in Model A this would initially 'only' involve the coordination of physical signals with mental entities, as well as the possible mental representation of entities. This would include the systems mentioned by Hauser *et al.*: Sensory-motor, conceptual-intentional, and recursive-computational. According to Hauser *et al.*, the faculty of language in the narrow sense (FLN), which is then apparently unique to humans, is based on the possibility of free recursion, which can easily be understood as the only indirectly or not at all coupled manipulation of universals in Model A, especially since this – as the authors also assume – is not bound to language and, according to them, cannot be understood as an evolutionary adaptation to communication requirements.

Jackendoff and Pinker [181] argue against this view to the extent that it is based on very specific ideas of cognitive processes and language, and point out that recursion itself is not yet a unique feature of humans; the thrust here is that the specifically human language ability can also be understood as an evolutionary adaptation. Model A offers a mediation between the different points of view by emphasizing the free recursion that is only possible for humans, but can explain this as evolutionarily developed. Model A also offers the possibility of a concretization of Chomsky's idea of a mental language organ: On the basis of and in interaction with material processes, certain non-material strutures (the 'mental organ') would be developed that could then channel human language development.

9.4 Can Model A contribute to our understanding of the intentional aspects of our psyche? – Volition, affects, motivation, needs and goals

Volitions are very complex processes in Model A. First of all, the core subject is to be seen as truly free in its actions. Nevertheless, the core subject can only think, feel and act in ways that its physical and mental structure allows. As a consequence, the less decisions are predetermined by the individual structure and the forces acting on it from within and without, the more freely the individual will be able to act. (There is an interesting parallel here to the medieval discussion between voluntarists and intellectualists as to whether will or knowledge comes first; in Model A, both must come together.) With humans, biological evolution now seems to have reached a level of mental complexity that makes people quite free, as long as sufficiently few internal and external forces act on them; think of the counter-examples of hunger or war. A complex mind will always find its decisions to be markedly underdetermined by its knowledge and its needs. In line with this consideration, we feel most free in our decisions when they play no role for us at all.

An important cause of forces that influence the core subject via modifications to the world map are the sub-subjects, which pursue their own goals that are nevertheless largely adapted to the survival of the individual as a whole in evolutionary terms. As already indicated above, emotions are then important messages from our subconsciousness, for which it would be neither sensible nor possible to be handed over to the core subject as purely physical signals or as explicitly formulated semantic information. Not possible, because access can only take place mentally and below the level of conceptual information; not meaningful, because the message is 'dense': It does not point to a completely defined pattern of behavior, but is to be understood as a call to activate the superior problem-solving capabilities of the core subject, for example in the search for food as a response to hunger. In Model A, feelings or a similar communication channel are therefore essential upwards from a certain level of development of organisms. In addition to physical and emotional needs, the core subject can develop further goals as motivation for actions, which ideally arise from a reflected overall view of the individual and its position in the world.

9.5 How are mental disorders and therapies to be understood in Model A?

In Model A, mental disorders and their therapeutic possibilities are to be understood as similarly complex as the formation of will. In the theoretical discussion of these illnesses, some interlocutors are more materialistic than the supposedly materialistic modern psychology itself (quite analogous to the discussion of the natural sciences in philosophy). An 'override' of the world by brain chemistry is then propagated: If only the chemistry is right, then the right feelings are produced, then thought content and the individual's lifeworld are relegated to their rightful places, because for our psyche these are only external factors that influence the balance of chemicals and thus our self-regulation. A psychology that is in touch with practice will view this more sceptically given our current level of knowledge about the processes in the brain. In Model A, we have the opportunity to take this scepticism into account: In addition to purely physical developments that can impair the function of physical structures (including the anchoring of mental entities!), there are mental processes, both subconscious and conscious, that can endanger mental health: Both affective and psychotic moments can be assigned an objectively real position in the world map of the core subject, but also in the 'hidden' world maps of the sub-subjects, which after all only partially overlap with the world map of the core subject. In Model A, a variety of possible clinical pictures appears to be practically necessary.

And such diversity is indeed observed: [182] Apart from physical trauma or age-related 'material' damage such as dementia and neurological developmental problems such as autism, we find affective and psychotic elements such as elevated, irritable or low mood ranging from mania to anxiety to depression, as well as distortions of subjective reality up to schizophrenia, and all this not infrequently in conjunction with sleep-wake disorders, eating disorders or somatoform disorders, substance abuse and self-harm up to suicide. Cognitive deficits such as reduced attention and memory, as well as language and social impairments are also common side effects. The boundaries between the various illnesses appear to be fluid, and co-morbidity, i.e. the shared occurrence of two or more problems, is widespread, which has led to the discussion as to whether it is not rather a spectrum that underlies all this. This in turn can be interpreted as an argument in favor of the above hypothesis that the neurobiology and the chemistry of the brain are the causes of mental issues; from the perspective of Model A, however, purely neurological mechanisms appear too homogeneous for the observed diversity.

In the philosophical discussion, Graham [183] accordingly feels compelled to define mental disorders as a partial impairment of fundamental psychological ability due to a mixture(!) of mental activity and neural mechanisms. It seems practically impossible to differentiate disorders precisely from one another and to define what should be regarded as an impairment regardless of context; nevertheless, there are also identifiable, objective elements.

In any case, most problems seem to result from a combination of susceptibility, stressful - or rather incisive - events and not infrequently 'lifestyle choices' such as continued substance abuse. Here, susceptibility is usually considered to be genetic/epigenetic and developmental/environmental, i.e. not necessarily understood to be only materially structured. A purely neuroscientific explanation of the problems, although it is certainly correct in some cases, often does not seem to correspond with the experience of those affected, who generally attach a great deal of life-historical significance to what they experience in the run-up to the actual illness – but also during the illness itself. [184, 185]

With Model A, we can consider both sides: Here, psychological problems can begin both in the material, as a result of developmental disorders, physical illness or trauma, as well as substance abuse, but also in the non-material, as significant problems that the individual must deal with in order to overcome obstacles to their growth. Due to their coupled nature, both material effects and non-material processes can lead to a spiral of mutual, negative influences. In practice, of course, psychology already operates as if this were the case anyway, because there is no other way to help people; Model A would offer the opportunity to better underpin this practice theoretically.

Fuchs' characterization of the development of schizophrenia [64], for example, would fit quite well into this picture: With Model A, a progressive 'decoupling' of the world map and the causal network of the physical world would appear to be the most likely central mechanism behind the clinical picture of schizophrenia. In line with this, Fuchs speaks of a progressive subjectivization of perception, with a reversal of intentionality, so that it is not I who turn to the objects, but the objects that turn to me; with objects that appear only as my perceptions and the associated solipsism; and finally the transition to delusion, which replaces my insecurity with a story that I can no longer critically distance myself from. In Model A, the resulting loss of the normally self-evident background of all our thoughts and actions would be attributable to neurobiological predispositions as well as a vicious circle of experienced influences of physical and mental nature.

To illustrate what kind of new ideas could be developed on the basis of Model A, two more (also highly hypothetical!) examples are given: Neurodiversity could be explained not only by purely neurological deviations, but also by natural variations in the interaction with and between sub-subjects;

e.g. a too 'loose' coordination of processes could complicate the inhibition of impulsive behavior and thus come to light as a disorder of executive functions, as it can be observed in ADHD for instance. And this could occur in particular for especially 'competent' sub-subjects.

Psychosomatic disorders or strong physical effects in post-traumatic stress disorders could be understood in Model A in such a way that information is shifted from the area of the conscious core subject into the areas of sub-subjects of our subconsciousness. In these cases, the body would actually and not only figuratively 'keep the score' in the sense of van der Kolk's book *The body keeps the score.* [186] It would remember the trauma and then show it, for example, in the form of 'affective bridges'. As the subjects would usually be unable to process the experience due to their limited world view, the aim of therapy would then have to be to make this task accessible to the core subject again. And since the sub-subjects cannot be addressed directly at a conceptual level, it would not be surprising that other approaches would have to be chosen here first, like bodily experiences or psychotropic substances. Somewhat unexpectedly, but demonstrably successful forms of therapy such as Eye Movement Desensitization and Reprocessing (EMDR) could then be interpreted in such a way that sub-subjects are given the opportunity not only to 'paint' the currently perceived scene, but are also given the opportunity to renegotiate the simultaneously remembered trauma with the core subject.

The existing, rather confusing therapy landscape would have to be expected accordingly. The high value of early intervention and cognitive behavioral therapy (CBT) could be explained by the risk of mutually reinforcing effects and the superior cognitive problem-solving abilities of the core subject: In talking about and with itself, the core subject opens up new possibilities for the sub-subjects. However, the ever-present link to physical processes would also explain why medication can help or is sometimes without alternative, especially in cases of advanced damage to physical/mental coordination. Placebo effects could be explained here as the setting of mental markers to control subconscious processes. The fact that strong physical stimuli such as the smell of ammonia can pull patients out of psychosis also seems fitting. We have already hypothesized the role of bodily experiences and psychotropic substances above. Analytical approaches or mixed forms of such approaches and CBT would in turn be justified in that they focus on the meaningful conflicts behind psychological problems.

If we turn our attention to the mental well-being of healthy individuals, it appears central in Model A to see the holobiont human as a team and not as the kingdom of the core subject. Corresponding considerations can, of course, already be found in psychological research, including Maslow, with his hierarchy of needs, etc. The observation that people rate their happiness on the one hand as a physically or subconsciously evoked emotion in the moment and on the other hand as consciously understood satisfaction with the course of their life would also not be unexpected for a psychology based on Model A.

9.6 Can we use the concept of personality in Model A in a meaningful way?

In Model A, personality is the product of circumstances, but also the sum of the individual's more or less free decisions. While free will can only be used quasi-randomly by the simplest subjects due to a lack of choice, it is normally underdetermined in humans and can therefore be used for essential life decisions. However, pronounced self-reinforcement effects will be observed; a restriction of my freedom of action makes it more difficult for me to avoid further restrictions; an extension makes this easier. In this sense, a relativism that attributes our concept of truth to power and thus ultimately to psychological factors is right: The unfree are driven by their psyche. People are thus partially removed from the complete determination of their circumstances; Stamer [187] then rightly wrote that a human life can only be told as a biography, as there can be no science of the individual human being and thus as a consequence also of our concrete world. Overall, the concept of personality is much less problematic in Model A than it is, for example, in purely panpsychistic models. [188]

For the psychological understanding of personality, it can be added here that in Model A, central personality aspects must be understood as important elements of the world map: My self-image as a person with a certain body for instance is part of my world map, which I can reflect on via the perception of this map. This way, my self-image can serve as an important target for the cybernetic/feedback-based coordination of processes in the human holobiont. This could also explain natural divergences between internal vs external views of the self (think of the issue of sexual identity), as well as the accumulating 'errors' in self-perception over a lifetime (think of one's mental vs physical age). More on current issues of identity can be found in Chapter 11.

Finally, one could discuss the extent to which basic personality traits such as the 'big five' (extraversion, openness, agreeableness, conscientiousness and neuroticism) could be derived from the basic options of the core subjects (exploration vs. exploitation of qualities, as well as self-initiative vs. collaboration with other subjects), possibly in connection with the further developmental possibilities of the structures that have been developed accordingly.

To summarize, Model A argues for a rich mental structure with strong subconscious, emotional forces because direct mind/body interaction only seems possible at the subconscious level. This suggests new mechanisms for the somatization of mental processes in addition to the processes of the body's influence on the mind. An open problem from the point of view of psychology is then how such a complex system could have developed or can develop in evolutionary and lifehistorical terms; here it must be explained above all how mental processes develop in step with the better known physical processes. A second open question is the extent to which Model A is consistent with the findings of neuroscience, which are of central importance also to psychology. Both questions will be addressed in the next chapter.

A number of explanations in Model A will be completely analogous to established psychology; the question here is what the added value of the model would then be. This cannot be decided with regard to individual problems, but only in the overall evaluation: Is the model as a whole a more helpful theoretical basis? Even if Model A were to prove itself in other areas, this question could still only be answered by future generations of psychologists. At least for the time being, Model A appears to be compatible with the central ideas of modern psychology.

Chapter 10

Counterarguments from neuroscience

If we now turn to neuroscience, it should be noted in advance that most of the knowledge acquired there to date and also much of the current research concerns the molecular and cellular basis of neuronal activity, which can provide us with arguments neither for nor against A-world, as these are not questioned in Model A any more than in the established scientific world-view. Equally unhelpful is much of the philosophical discourse in this area, as it deals with the problems that have to be overcome if one does not want to consider alternatives such as Model A. (An initial overview can be found on the pages of the Stanford Encyclopedia of Philosophy under the keywords 'Neuroscience' and 'The Neuroscience of Consciousness'.)

Those arguments that speak directly for or against Model A in this area, like the 'hard' problem of qualia, or the causal closure of the physical world, have already been addressed in the previous chapters. It should also be mentioned that there is a lively discussion in the 'Philosophy of Memory' about whether experiences can be thought of as stored in the brain or not, whereby the argumentative positions can be categorized roughly on the basis of the distinction between quantitative and qualitative information already outlined. [189] In this discussion, one could take a mediating position with Model A, since here information is stored both in the brain and in the mind, i.e. the world map; the decisive question would then always be what kind of information is being spoken of specifically. Nevertheless, it will be interesting to see further results of experimental neuroscience on 'engrams' as neuronal substrates of memories and on the manipulation of engram cells.

10.1 Is Model A compatible with the findings of neuroscience to date?

Neuroscience is certainly one of the most rapidly developing scientific fields in the 21st century, with a wealth of established knowledge, [157, 158] especially on the aforementioned molecular basis of neuronal processes, but also on memory and attention mechanisms, decision-making and executive functions. Continued rapid progress [190, 191] can also be observed in the modeling and simulation of cognitive processes, [192–194] as well as the realization of brain/computer interfaces. [195] There are of course still open questions, especially with regard to neuronal correlates [196] and theories of consciousness. [197, 198] More specifically, the neural code (how information is represented in the brain) and the binding problem (how higher symbols are constructed from elementary symbols; more on this below) are still not understood. [199, 200] Also in general, the 'gap' [201] between mind and brain can by no means be regarded as already bridged by neuroscience, [57] at least from the perspective of the philosophy of mind, be it with regard to the problem of qualia or that of mental causation. [103, 104]

Sterling and Laughlin [157] consider the central function of the brain to be the anticipatory regulation of the organism, which includes the control of its behavior. This view of the brain has important implications for our understanding of thoughts and feelings and fits seamlessly with the idea that the brain does not need to do anything other than process quantitative information. The structure and function of the brain can then be understood as the result of a long process of 'evolutionary efficiency optimization' in terms of our current understanding of the principles of biological evolution. The energetic costs of biological information processing increases disproportionately with increasing amounts of information and faster processing, which results in an important design principle for brains; information should always be transmitted as little and as slowly as possible. As a result, brains are organized in modules, with computations distributed over many small areas and then consolidated in centers such as the thalamus. Nothing should reach a higher processing level that could already be processed and returned in a lower layer. Accordingly, a large amount of sensory information and motor control signals never reach our consciousness.

Many of the mechanisms behind the computations on the lower levels have already been described in detail by neuroscientists: Basic operations are realized by protein folding processes on the nanometer scale, we find intracellular circuits on the micrometer scale, and finally neurons on the millimeter scale. At the lowest level, the computations are carried out using diffusion-controlled chemistry and are therefore very 'cheap', although limited to small scales for acceptable processing rates. Larger distances are then bridged with electrical impulses, except that the restoration of operational readiness is then very expensive in comparison. In all of this, several compromises must always be made at the same time, which has led to the development of customized cells for specific computation purposes, for instance with regard to their length and thickness, but also as the number and type of contacts to other cells.

A further problem is that the entire process is noisy already due to temperature-related movements on the atomic scale, so that the signals normally still have to be summed up, which makes further compromises necessary. Learning processes are then to be understood as an adaptation of static and/or dynamic neuronal structures, which has also been studied in great detail at the molecular and cellular level. However, recent results indicate that the plasticity of the structures is realized by multiple, collaborative mechanisms and does not simply result from an increased networking of frequently used neural connections.

Model A is well compatible with all these findings, but sees a new mechanism of converting quantitative into qualitative information at work on higher, hitherto not understood levels of processing, as outlined in the previous chapters. This is arguably unproblematic with regard to the readout of quantitative information in the activity patterns and the evocation of qualitative information in the non-material mind (as long as one can believe that this is conceivable at all). It nevertheless requires further discussion when it comes to the reverse, that is the generation of activity patterns as a result of the readout of non-material, qualitative information. In the previous chapter, we have already specified this to the extent that the brain would have to provide a large number of achievable physical states of equal energy and possibly entropy. This will be discussed further in the next section.

The most important 'prediction' of Model A for neuroscience would be that information processing at the higher as opposed to the lower levels cannot be decoded in terms of a 'neuronal code' in which specific patterns directly describe certain content, but only in the sense that these patterns refer to an intrinsically immaterial content. Some parts of the processing procedures at the highest level would no longer have a direct material equivalent. Furthermore, the assignment of pattern and content on the levels below would not be a necessary one, but an evolutionary and life-historically developed one, i.e. ultimately contingent. (We could never be able to establish a *logically necessary* connection between certain activity patterns and our subjective sensory impressions.) The connection between brain activity and content shown in the known experiments would also be expected in Model

A, but there would be a fundamental limit to such experiments.

The most important consequence of Model A would therefore be a methodological one: Talking about subjects and qualities can only be done on the basis of indirectly obtained 'measurement data'. Practically, however, this changes the situation less dramatically than it might initially appear: Even at present, the allocation of activity patterns and content is an indirect one that has to take a detour via the information provided by the people involved. The problem of the impossibility of materially grasping the non-material only arises in Model A exactly because that's how our reality seems to be like.

10.2 Can the mechanism of mental causation in Model A be reconciled with the findings of neuroscience to date?

So let us come to the core of the problem: Since with the turn to idealism we have elevated mental causation to the rank of a central principle, the individual – and here more precisely; also the core subject – must now be capable of meaningful mental causation. To do this, it must be able to make changes to its world map in accordance with the model outlined in the previous chapter, which is not in itself argumentatively problematic at this point; but which must also be perceived by sub-subjects and converted into brain activity in accordance with acquired rules, so that material effects can then 'cascade' from there into the physical world. Which brain activities need to be generated for this cannot be concluded at this point and is largely an empirical question, as there can only be a contingent connection between patterns and content rather than a necessary one, as described above. Nevertheless, some basic conditions can be derived, without which the outlined mechanism would not at all be possible.

In essence, the mechanism should above all not violate any physical laws, which, as explained in the previous chapter, requires in particular the correct balancing of the processes, i.e. the conservation of energy and possibly, but not necessarily, the conservation of entropy. (It is discussed to what extent the conservation of energy is violated within the framework of the general theory of relativity; however, this refers to processes on cosmic scales. In the case of entropy, only the overall decrease is 'forbidden'.) In Model A, in order for subjects to be able to translate non-material information into material activity patterns in a meaningful way, the brain would have to provide a large number of energetically equivalent states that could be selected according to non-material information in order to trigger a sequence specific to the respective state.

It should be noted that every macroscopic system practically automatically has a gigantic number of such states; imagine that every small, energetically relevant change in one particle can be balanced out by a similar, but opposing change in another. Nevertheless, the relevant states must also differ significantly in their realization in order to be able to trigger completely different material effects, and be 'accessible' in the sense that their invocation must be possible by influencing one or at least only a few types of material entities. Otherwise one would have a possibly enormous number of different interfaces, which would all have to have developed in parallel in evolutionary terms.

The influencing of meso-scale electromagnetic fields in the brain, or the concerted influencing of physically independent micro-scale particle processes, appear to be the most sensible initial hypotheses here, as these are the only ways for sub-subjects to achieve effects at the meso-scale while having to change not too many material entities. For the corresponding sub-subjects, these physical entities could be part of their world map, as well as some non-material entities that would function as markers for 'requested' changes. There is no interaction problem at this interface; the physical and mental entities are fundamentally the same in nature, except that additional consistency rules are followed for the former. So that the energy is properly preserved, the respective influence would have to start in a coordinated manner from a rest value, i.e. a value not equal to zero. The necessary non-local coordination can be achieved via joint access to non-material markers set from a higher level. The strength of the influence would be assumed to be as minimal as possible, but the signal must still stand out of the existing noise. Here one could further consider whether there would be an optimum with regard to the necessary effect strength and the number of influence points used for a control command, as a signal could also stand out from the noise under certain circumstances due to the coordinated combination of very many signal sources. An alternative idea would be the existence of a 'tipping' or rest point from which completely different states could be reached with minimal influence; at first glance, this seems less suitable for functioning despite noise. However, the situation could be different if, instead of a tipping point, there would be a kind of 'activity cliff', over which the system could be 'pushed' into the desired functional state by the concerted action of several physically, but not mentally independent parts.

In line with the explanations in the previous chapter, it should also be added that the control commands, which are themselves necessarily simple especially in the area of motor skills, would then mostly have to trigger physically learned programs, or would at least refrain from intervening in an inhibitory manner in the case of given environmental triggers. This would be in line with Moravec's paradox [159] that 'lower', often unconscious abilities require large, but 'higher', cognitive abilities require small physical computational resources (think of juggling and chess). As a further dimension, one could add that qualitative information processing in the mind would make it easy to act inconsistently, while following instructions motivated by the consistency rules of the physical world would be difficult.

Do these considerations fit with what we already know about the brain and could we derive any experimentally relevant consequences? The brain does indeed operate in a rest mode that is energetically very similar to its activity mode: The basic neural activity consumes over 95% of the energy. [202] Remaining differences are also conceivable in Model A, since here too, after the actual step of mental causation, further energy resources could certainly be mobilized for sensory or motor computations. In addition to a 'task-positive network' or 'dorsal attention network' (DAN), which is mainly related to the performance of new, attention-demanding tasks (in Model A: more strongly oriented towards the material, external world), there is a 'default mode network' (DMN), which is mainly related to emotional, self-referential and memory-related activity (in Model A: more strongly oriented towards the inner, especially also non-material world). [203, 204] The DAN can also be seen as a detector of new environmental conditions, the DMN as a self-centered prediction model for the world; with Model A the former could be understood as the actively writing part and the latter as the actively reading part of the interface to the world map.

The DMN always starts from a high baseline of activity and undergoes only minor changes through specific tasks, [203] as would be expected for the active reading side of the interface outlined above. Also, that the DMN is critical for planned and reflective behavior; that it appears to play a leading role for the whole brain; and that the activities of DAN and DMN are anti-correlated 80% of the time, [203] fits with the idea that with the DAN and DMN we observe the material processes of the brain-mind interface. According to this, the effects of mental causation should be found primarily as coordinated patterns in the DMN, from which mental commands would be read. Overall, the considerations here are admittedly still far too simplistic to derive experimentally relevant consequences in detail. Further below, an attempt is made to further elaborate a possible mechanism of mental causation in order to consider whether verifiable consequences could be derived; however, our considerations already suggest, for example, that an equivalent to the DAN/DMN system as an interface to the world map should be found for all organisms that sleep, since that would imply a world map that must be periodically 'repaired' (see chapter 9).

It can perhaps be concluded that the observed brain-wide organization of electromagnetic activity patterns, starting from a resting activity and with energetically balancing elements, fits well with what should be expected for Model A.

10.3 Does Model A contradict the basic ideas underlying current neuroscientific research?

With regard to current neuroscientific research, it should first be noted that the investigation of the molecular and cellular foundations is of great importance also for Model A. The only difference is that somewhere beyond the sensorimotor level, molecular and cellular activities no longer play a role for information processing, but are responsible for the provision of activity patterns that can serve as 'anchors' for non-material content.

Also the 'mapping' of the brain, the allocation of content and structures or activity patterns, is of central importance for Model A. It should be noted here that voxel-(spatial pixel-)based morphometry, which has become immensely important since the 2000s and is based on magnetic resonance imaging (MRI) data, has run into a replication crisis because, according to the established protocols, thought processes would still have to be assumed even for a dead salmon. [205] However, also from the perspective of Model A, this crisis should not be understood as a fundamental one, because in A-world, too, a physical implementation in assignable brain structures and processes is assumed for quantitative information processing below the non-material level.

Finally, the theory of complex or dynamic – i.e. spatially or temporally non-trivial – systems, which is central to theoretical neuroscience, is important also for Model A: It can be used to understand possible spatial or temporal superstructures in brain activity, which could represent certain internal and external entities as 'stable attractors' of neuronal activity. In Model A, however, this is limited to the level of processing quantitative information, which is why only concrete, fluid representations can be identified here, which find their stable abstraction in the non-material world map only. Concrete in the sense that they are only the sum of their examples and fluid in the sense that they are open to further, possibly 'catastrophic' modification without thresholds, because they never reach the point where they could be recognized as a closed whole and thus used for the discarding of contradictory information. (Similar to DNNs having no direct defense mechanism against data poisoning.)

In Model A, we have the additional possibility that in a kind of quantum leap, such semi-stable (since physically implemented) representations are connected to 'super-physically' stable (since mentally implemented) abstractions, which can then help to stabilize the physical structures as markers for feedback processes. It is then rather helpful that these real abstractions are 'broad', i.e. not based on a fixed set of examples, but entities in the world map could also assume the function of hypothetical 'Jennifer Aniston neurons', i.e. cells assigned to specific contents.

A major difference between Model A and established ideas in neuroscientific research is, of course, that consciousness, qualia, intentionality, and certain cognitive abilities according to the explanations in the previous chapters, could not be understood as quantitative higher order information processing; neither by means of 'higher order', 'global neural workspace', or 'recurrent process' theories, as evaluated within the framework of the Cogitate Consortium, nor by effective emergence models such as that of Integrated Information Theory (ITT). [197]

If we turn to the here relevant unresolved problems of neuroscientific research, [57] we find above all questions relating to the neural code (the modern equivalent, so to speak, of the movement of heavenly bodies in the late Middle Ages): How does it represent content? How is it stored in memory? How is it updated during learning? And subsequently, questions that are characterized as possibly unsolvable: How is flexible and generative human cognition possible? What role does consciousness play? And so on. Model A, as outlined in this and the previous chapters, offers possible new answers to all these questions, with neuroscience being assigned the central role of deciphering the physical neural code behind quantitative information processing, which should be largely in line with the scientists' self-image.

A central question of representation is the so-called 'binding problem', how uniform perceptions are constructed from a multitude of sensory impressions, or ultimately how higher symbols are constructed from more elementary ones. Model A could provide a solution here: After the initial creation of a corresponding 'empty' bundle in the world map, there is a stable reference point for the assignment of further information. This would not necessarily require additional meta-information on the physical level as to what piece of information should be assigned to which entity in the world map, if we do not think of the subconscious subject as sitting at the end of the physical 'information pipeline' and waiting for the result, but as being able to modify bundles on the basis of quantiative information from different sections of the pipeline. This task would be made even easier by the fact that integration always takes place against the background of the persistent world map even through states such as sleep, i.e. with a given framework for how certain information is to be classified.

This consideration is certainly not yet to be understood as a solution to the binding problem, but it at least offers a direction in which a solution is conceivable. More specifically, neuroscience would have to develop an approach in which a non-material storage location, perhaps conceivable as a semantic network, is placed alongside the established models of quantitative information processing. In this non-material storage location, the 'empty' bundle of a consciousness is already predetermined and fundamental structures are created at the very beginning of information processing. These structures are subsequently 'colored in' step by step with the help of physical information processing, on the basis of sensory perceptions, but also on an affective and finally cognitive level. Such models could then be seamlessly linked to psychology and, in the next step, to discourses in the social sciences and humanities.

Model A would thus also raise new questions for neuroscience (and evolutionary and developmental biology), in particular concerning the code for the brain-mind interface. Derived from this, knowledge regarding the concrete generation and bundling of entities in the mind could be expected. Finally, at a higher level, one would ask about the basic qualitative elements and their evolutionarily developed processing rules, perhaps in the sense of a 'characteristica universalis' of human beings.

10.4 How could we imagine the interface between mind and brain in concrete terms?

As promised above, we can now try to further concretize the brain/mind interface in Model A, although without experimental feedback this can only be done exemplarily and can thus only serve to generate ideas for future neuroscientific experiments.

At the (sub-)atomic level, micro-subjects are extremely limited in their actions already by the (non-)availability of more complex rules. The linking of further non-material building blocks with changes in physical properties must be seen as a creative act, the possible repetition of which depends on its individual and then evolutionary benefit. This benefit not only determines the fate of this individual rule for linking, but also the fate of the rules on the basis of which new rules were invented and are applied later on. Model A allows these 'rules', as structured bundles of nonmaterial building blocks, to take on an extremely complex nature up to the form of ethical and aesthetic considerations.

In addition to the fundamental perception of physical facts via the generation of qualia on the basis of changes in physical properties, the core subject should also be able to act in the physical world. To do this, only the manipulation of non-material building blocks in its world map is available, which must then be read out and implemented by its subconscious 'psychobiome', that is the team of its sub-subjects, in the case of physically relevant actions.

Due to the quantitative information processing involved, complex physical processes, as 'programs' or 'action templates' that must to have been learned physically, i.e. in the neuronal network of the brain, at some point, can only be triggered or vetoed by the core-subject. As mentioned before, Libet-like experiments, showing that neuronal activities can precede conscious decisions, could be understood against this background: Our brain must always already offer the sub-subjects and the core subject behavioral patterns or action templates appropriate to the respective context so that they can make decisions. In this sense, certain quantitative information processes would lead to the provision of action templates, with regard to the execution or at least prevention of which both sub-subjects and the core subject would normally still have intervention options. Purely mental thought processes, on the other hand, would not necessarily be dependent on such templates. Research into the quantitative information processes associated with action templates would remain almost entirely in the domain of neuroscience.

In any case, it would be possible for subjects to use certain brain states as input for the rest of our brain's neural network. These states would not have to be meaningfully different from each other, since any input, for instance the 'firing' of a certain nerve cell, could be assigned to any output via a sufficiently complex neural network and enough training. And since 'behind' the input there would be any number of complex, purely non-material processes, the minimum necessary information content would also be very low. (This could be significant because, for example, brain waves or the signals of individual nerve cells only have a very low information content.)

The question remains, however, as to how such an influence of the subject on neural structures can be conceived while adhering to the consistency rules of the material world. (Model A would in principle also be compatible with a minimal violation of conservation laws, since these would only correspond to learned rules. But this would be difficult to reconcile with the basic idea of a material world as an identity-preserving anchor of the mental world explained in chapter 6, for which conservation laws are a central aspect of identity preservation). The availability of a large number of energetically identical states, mentioned several times by now, is not yet sufficient here, because it remains unclear how the causal history of the interactions that would lead to the respective state can be determined both materially and mentally.

From the perspective of Model A, the most natural way out of this predicament is to fall back on the statistical nature of interactions on the (sub-)atomic scale. The description of the physical processes on this scale via quantum mechanics outlined in Chapter 6 means that these processes have a causal history only down to the level of their description as quantum systems. Below this, only compliance with certain consistency rules, but not the movement of the individual, no longer separate part, is guaranteed. A 'naive' use of this fact in the sense of a theory in which the brain uses meso-scale quantum effects either to generate consciousness, to perform quantum computer-like computing feats, or to allow a mind, however conceived, to influence global structures of the brain, are all most likely incompatible with the physics and neurobiology of our 'warm, wet and noisy' brains. Under such conditions, quantum effects are limited to the very small length and time scales of molecular systems. [206–208]

Nevertheless, a model in which a mind influences a large number of individual quantum systems in a coordinated manner in order to 'switch' meso-scale states would be entirely compatible with this. Statistical quantum fluctuations of a multitude of systems would then occur 'purely by chance' in such a coordinated manner that a larger structure would be pushed over a certain 'activity cliff', so that an immediate, chance-based return to the original state would be prevented. It should be noted that the individual quantum systems would then first need a 'recovery' or 'quarantine' period in order to maintain the statistically correct distribution of their fluctuations, which is directly linked to the conservation laws. Somewhat simplified: The 'random' events should only be triggered so rarely that the overall statistics could be maintained. A 'pattern out of nothing' would thus be able to lead to a targeted redistribution of matter and energy – without a traceable causal history. [1] Such a model would be characterized by a three-step process of a coordination of quantum fluctuations, an activity cliff and a necessary recovery time.

The alternative control via individual quantum systems appears to be unlikely due to the following considerations: The functional state must be sufficiently specific for its function, which would have to correspond to a very

[1] This could be followed by a lengthy discussion on the concept of energy conservation: Since in purely physical systems an energy redistribution is not possible without adding or subtracting energy, one might be inclined to see such a redistribution itself as proof of a violation of energy conservation. The point here is, however, that the redistribution does not occur via a purely physical mechanism, but via the mental coordination of quantum fluctuations, and that for this the system neither has to absorb nor release energy in the balance.

unlikely fluctuation so that it only occurs very rarely by pure chance. In order to maintain the correct quantum statistics despite such outliers, a very long 'recovery time' with purely random fluctuations would be necessary; not necessarily in the case of a single occurrence, but the biological functionality would have to be available repeatedly. It therefore seems to make more sense to switch a functional state via the coordination of several quantum systems, in which the individual fluctuation does not have to be particularly improbable, since it is only the improbability of the coordinated co-occurrence of all the necessary fluctuations that specifies the functional state. (The process could in any case include an initially slightly reduced probability of the 'utilized' fluctuation.) The greatly simplified picture would then not be that of a player who rolls a 6 30 times in succession, but of 30 players who all roll a 6 at exactly the same time. The individual quantum states would then also not be correlated with a specific information content; this would only be the case for the collective state.

In a purely physical model, the necessary coordination of quantum systems would not be possible, because this would then correspond to meso-scale quantum effects, which most people now reject as practically impossible in a brain. In dualistic and idealistic models, however, it is of course possible for these states to be 'entangled' not physically, but via mental entities, which would also give the entanglement a 'superphysical' stability compared to the conditions in the brain and quantitative information processing in general.

Ultimately, there would then still have to be a lowest level of our non-material subconsciousness that would interact with the material brain: There would have to be actions of sub-subjects that spill over from the non-material world of bundles of universals, for which no strict consistency rules apply, into the material world of bundles of universals, for which such rules are always followed. So what we are still looking for is the place (or the places) where our mind can influence the brain; the 'hooks' by which the two are connected.

It is tempting to think that this could best be done by coupling to electromagnetic fields, for example through changes in the oscillation modes of large brain networks such as the DMN. However, such a mechanism to connect *this* mind with *this* body seems rather unstable: Fields would probably have to be understood as properties of space(points) and would therefore not automatically follow our material brain in movements. And although we could easily imagine a mechanism for perceiving and 'tracking' changes in materially induced fields, overall this seems a rather failure-prone solution for maintaining the connection under all circumstances. (Admittedly, this would be different if we would not have to understand fields as properties of spatial points). Two further observations speak against fields as 'hooks':

Firstly, the integration of information in the brain does not seem to occur globally, but at the neuronal level. And secondly, the connection should also be understandable in terms of its evolutionary development, which suggests that it should be locatable already at the level of individual cells.

It therefore seems more likely overall that the mind and brain are not connected via fields, but via specific cells. Here, it is in turn unclear whether cells as a whole or only certain structures within the cell maintain the connection. In both cases, however, mental causation would most likely require that certain molecular structures could be effectively influenced. The list of conceivable candidate structures is long, but we are looking only for those that are suitable for a mechanism with an activity cliff. These are then rather not the much-discussed microtubules (whose spatial shape could allow long-range quantum effects, which on the other hand is now anyhow considered unlikely [209,210]), but structures like post-synaptic neurotransmitter receptors (and possibly post-synaptic ion channels). There is a whole 'proteome' of such systems [211] and a still poorly understood, extremely high variability in their activity is observed. [212] It would then seem most reasonable to assume that similar mechanisms have evolved for different classes of such proteins. The following is a first speculation on how a mechanism of mental causation based on post-synaptic neurotransmitter receptors might work within the framework of Model A.

Before we can start, we need to take a brief look at the energetics of proteins: The absolute energies of such systems are in the range of tens to tens of thousands of atomic energy units (called 'Hartree' (H)), because these energies are calculated quantum mechanically as the difference between the actual systems and all nuclei and electrons at an infinite distance from each other. In contrast, the relative energies between functional states of such proteins are only a few tens to hundreds of milli-Hartree (mH) small, i.e. many orders of magnitude smaller, because the different functional states differ only by changes in the spatial arrangement of the protein, and not in the actual binding sequence. This means that all 'covalent' bonds, that is 'actual' bonds with 'shared electron density', which make up the largest part of the absolute energy, remain unchanged. The differences between the functional states then only result from differences in the inter- and intramolecular 'non-covalent' interactions, that is from attraction and repulsion effects through space and not through actual bonds. Local interactions, for instance with medical drug molecules or neurotransmitters, also arise exclusively from these non-covalent bonding effects (again no actual bonds are formed or broken) and are in sum generally much lower than the energy of proper bonds, which correspond to energy differences of around 150 mH or 100 kcal/mol, depending on the strength of the bond, which in turn depends on the elements involved and their 'chemical environment'.

The fact that these interactions are so (locally) weak is of crucial importance for biological systems; if the interactions were stronger than the typical bond strengths, they could damage the structure of the organic molecules involved. Water is also very helpful here, because whenever a drug molecule or neurotransmitter is removed from the protein or has not yet 'docked', water molecules take its place, whereby most of the possible binding energy is recovered at this binding site, which in turn reduces the *differences* in energy differences. Further cancellation effects result from the compensation of energy and enthalpy effects; water is less well bound but can move more freely in the binding pocket, which leads to more favorable entropic effects on the (free) energy.

The non-covalent interactions between protein parts and with drug molecules or neurotransmitters can be approximately divided into contributions of 'Pauli repulsion' (electron clouds repel each other), polar (non-time-dependent charge) effects including the famous hydrogen bonds, and non-polar (time-dependent charge) effects. In bonding, we usually find a very delicate balance between these effects with comparable effect strengths for polar and non-polar contributions and both in opposition to the repulsion effects. Now we (finally) come to a very important observation: The nonpolar part, referred to in biology and some parts of physics as van der Waals interactions, but in chemistry named dispersion, arises solely through the coupled alignment of fluctuating dipoles in the electron densities involved, that is through purely quantum mechanical fluctuation effects without a causal history. This in turn means that a mind that could coordinate such fluctuations would have the 'hooks' through which it could influence molecular activity at an activity cliff (the synapse) without violating conservation laws. The energy for this is already available; the mind 'only' has to coordinate the otherwise randomly occurring activities spatially.

In Model A, there are now four more important conditions or actually possibilities: Firstly, in Model A, quantum systems would be 'managed' by a group of micro-subjects or 'cellular automata' that ensure that the material properties of the system are propagated correctly in the material world. Secondly, we can assume for Model A that this bundle of properties that is the quantum system additionally has certain universal properties as non-material building blocks which serve as control markers. (What exactly these properties look like is irrelevant for the time being.) Due to their universal, non-spatial nature, these markers can be added to several quantum systems, receptors of a synapse and/or protein class at the same time and can be 'activated' together. Thirdly, we can add rules, i.e. bundles of non-material building blocks, to quantum systems so that the group of micro-subjects knows what to do with the marker: If the marker is activated, the micro-subjects shift fluctuations to the binding pockets as far as possible, thereby increasing their binding affinity. And fourthly, we could add a 'bridge bundle' that includes markers from the 'lower' bundle that is the protein, as well as qualia or other non-material building blocks from a 'higher' bundle in our subconsciousness, handled by the corresponding sub-subjects.

Now we are ready to go through an example of a mind/brain interaction. First, if a core subject activates certain qualia, i.e. moves them in its world map, then sub-subjects in the subconsciousness activate certain other properties in their 'bridge' bundles, which they share as markers with the physical bundles of proteins, so that the activity of the binding of neurotransmitters can be strongly influenced in these systems. (A multilayer structure between core-subject and brain seems of course more likely.) Second, this non-materially caused physical impulse leads to a cascade of synaptic activity, the opening of ion channels, the formation of neuronal spikes, and finally oscillations in large brain networks. [2]

Very little information would have to be transmitted at the actual mind/brain interface, as all 'computationally intensive' sensory-motor processes would be materially encoded in the brain. Whereas, for example, complex memory content could presumably be stored more easily in the non-material mind.

With the post-synaptic neurotransmitter receptors we would therefore have systems that could be controlled via the coordination of quantum fluctuations and whose neurobiological functioning would correspond to an activity cliff. The quantum system would be influenced on the femto- to picosecond scale, which would leave plenty of 'recovery time' between mental events on the milli-second scale.

Finally, it should be noted that in the above model, *this* mind is clearly connected to *this* body, but the mind relies on the brain as an anchor and neurochemistry for mental causation: The individuation of universal properties in the mind would be realized via the (multi-layered?) anchoring in subconscious 'bridge' bundles, which in turn would be anchored in material protein bundles, which are individuated by means of the elementary particles that constitute them as in contemporary physics via their positioning in space.

[2] Analogous mechanisms could of course be developed in general for the realization of 'structuring ideas' in the material world, e.g. for gene expression or cell organization, but such proposals – like the proposal here – should be taken with extreme caution.

10.5 The role of quantum theory in Model A

We have encountered quantum theory three times in our previous considerations, which is why a brief review may help to avoid misunderstandings: First, Model A should be able to explain why modern physics needs to describe processes at the (sub-)atomic scale by means of such a theory; this is the case, according to Model A, because elementary particles are also 'just' bundles of universals, as was suggested in Chapter 6. Second, the observation that at least some cognitive operations of human brains seem to have similarities with quantum information-theoretic processes is not explained by the brain being a physical quantum computer in some form, but because physical as well as mental entities are built from universals, as explained in chapter 7. Finally, we have encountered quantum theory a third time in this chapter, as a justification that mental causation is possible without violating conservation laws, because the causal history of physical processes at the (sub-)atomic level is lost in the statistical noise of quantum theory.

Quantum theory is not used in Model A to equate physical entanglement phenomena with contents of consciousness such as qualia. It is not used to argue that brains are physical quantum computers and because of this capable of their special intellectual performance. And it is also not used to postulate that long-range physical quantum effects allow a mind, however conceived, to coordinate a brain. All of these ideas have been discussed in detail in the literature and appear, if not obviously wrong, at least very unlikely. [206–208] And although much more will certainly be discovered about the significance of quantum effects in biology, [213] it seems foreseeable that the ideas mentioned above will not have a chance of being rehabilitated.

10.6 What experimental evidence could support or refute the proposed model of a mind/brain interface?

Direct experimental evidence of mental causation processes would be extremely difficult to find also in A-world, precisely for the sake of energy conservation. Apart from specific patterns of synaptic activity, it would only be possible to identify extremely small energy redistributions, in the order of magnitude of 10^{-20} joules per neurotransmitter. Perhaps a few hundred of them would have to be involved per coordinated event, so that for a synapse it would be a matter of measuring temperature fluctuations of maybe 10^{-7} Kelvin in a living brain; a similar calculation can be found in Summhammer. [214])

Experiments that investigate states such as near-death, coma, anesthesia, sleepwalking and dreaming, including sleep deprivation and/or sensory deprivation, could provide interesting results for A-world. One central prediction of Model A would be the relative independence of the different subjects involved in the whole individual, which can possibly be corroborated by the investigation of such special states. With regard to the interface described, the targeted numbing or switching off of the elements involved would be particularly interesting. Another central prediction would be the possibility of purely mental activity without neural correlates and associated non-material learning and memory processes, which could possibly also be investigated with the observation of the special states described above.

Experiments with brain organoids, [215, 216] i.e. brain-like cell cultures, appear very interesting in this context. However, it should be noted here that in Model A micro-subjects cannot simply 'take over' such structures in order to become sub-subjects: For the meaningful control of material structures, also non-material structures must have been built in step with them. We would probably only be able to discover how this could be actively brought about once we have understood the material-neuronal code much better. And this is then a dilemma in the sense that understanding this code could in turn require at least a preliminary understanding of the non-material parts. In any case, with Model A, the non-development of higher functions rather than the opposite would be expected as an observation in experiments with brain organoids. But for the time being, any interesting observation of such systems will most likely be attributed to physical self-organization processes. At the moment, at least, it seems that not even a continued failure of purely material models could persuade the neurosciences to consider dualistic or idealistic models – and such a failure is of course by no means a foregone conclusion.

Arguments that rely on the possibility of reconstructing speech or video information from neural data are rather not suited to refute Model A and the proposed mind/brain interface. Although impressive experiments are possible in this area, [217, 218] they ultimately only show a correlation previously approved by humans, which would be the situation expected also in A-world: Even if certain content is stored abstractly and non-materially, it seems reasonable that its 'activation' in the case of visual and auditory references is accompanied by the renewed activation of certain material structures that clothe the abstract information in a sensually tangible 'garment', so that a remembered scene is actually 'colored in' anew. The considerations above also fit in with the observation, that electroencephalography (EEG)-based approaches are very sensitive, but not very specific, so that reconstruction or prediction only works well under laboratory conditions. A sufficiently long prior measurement of brain activity under controlled conditions would allow a limited form of mind reading also in A-world.

However, there would certainly be arguments that would refute Model A practically immediately: The definitive proof of a principled boundary between perception and cognition would be one such argument. [219] In Model A, perception is not as free as cognition only because it is not generated by us, but by our subconsciousness. Nevertheless, the building blocks of perception and cognition would be of the same nature; in the sense of a classical-philosophical understanding of ideaesthesia as the perception of ideas, as well as its modern-psychological understanding of a free linkability of concepts with perceptions. According to Model A, the relative freedom of perception would be evident, for example, in extreme cases such as schizophrenia (described in the previous chapter).

10.7 How can we imagine the evolutionary development of the human brain with Model A?

10.7.1 The established model

The core elements of our scientific explanation of life are genes, cells, individuals such as plants, animals, and humans, as well as eco-systems. The theory of evolution has established itself as an overarching body of thought which, like the physical theories discussed in the excursus in Chapter 6, must be regarded as a cornerstone of our modern scientific view of the world. This also means that any failure to adequately account for biological evolution should be seen as an argument against Model A. (The following summary roughly corresponds to the presentation in Herron/Freeman [220]).

The theory of evolution has been able to take its central place in biological thinking due to an overwhelming amount of consistent observations on natural populations as well as from experiments on laboratory populations. This applies both to microevolution (evolutionary processes within a few generations) and to macroevolutionary speciation (the splitting of lineages into different species). Arguments in favor of microevolution can be gained, for example, from experiments with shortlived organisms like fruit flies. This is often overlooked by evolution critics: Biology can observe evolution 'at work' and only has to assume that nothing prevents the same mechanisms from coming into play on longer time scales, too. Arguments for macroevolution, i.e. the emergence of new life forms, are often made on the basis of fossil finds and refer to structural and/or molecular homologies; we observe the same building blocks and patterns everywhere. The theory of evolution suggests that life has been evolving for around 3 billion years, which fits in well with the findings of the earth sciences (geology, geophysics, etc.).

The basic assumptions behind the theory of evolution, i.e. 'Darwin's postulates', are: 1. That individuals differ from each other, 2. that these differences are at least partially passed on to offspring, 3. that some individuals are more successful at reproducing than others and 4. that this is not just luck, but is at least partly due to inherited differences. Only later were these postulates combined with genetics, which provided a mechanism for variation and inheritance, as part of the so-called 'modern synthesis'. Subsequently, it was possible to subject each of the four basic assumptions to rigorous testing, with the result that all four do indeed appear to apply.

Further evolutionary mechanisms have been identified in recent decades. In addition to the mechanism of 'natural selection' described above, which can have both a positive reinforcing and a negative attenuating effect, the importance of 'genetic drift', i.e. the completely random 'selection' of traits (more precisely: alleles), has become a central element of the so-called 'neutral theory' of evolution. At its core, it claims that genetic drift is the most important mechanism; a position that is supported by the clockworklike evolution of certain genes, and which therefore now serves as a null hypothesis for the proof of (positive) natural selection.

Other relevant mechanisms are the migration of populations (and thus alleles) and non-random mating, i.e. influences that are to some extent related to the behavior of (groups of) individuals. Modern biology has developed predictive mathematical models for all these mechanisms, thus establishing evolution as the central theory of biology that it now is. As such, it is able to explain such far-reaching and complicated phenomena as the 'family tree of life', but also such specific ones as 'life-historical' characteristics of ageing, including the interactions between genes and the environment ('epigenetics'), human evolution and the development of social behavior. The extent to which all these phenomena are *fully* explained by the theory of evolution is the subject of ongoing debate.

A very interesting case is developmental biology, which deals with the growth of individuals from birth to death, and which was initially excluded from the modern synthesis partly due to the complexity of its subject matter, but is now at the forefront of research into the foundations of evolutionary theory as part of 'Evo-Devo'. An initial central finding of this research direction is that relatively small environmental influences can have an outsized impact, especially if they affect very early stages of development.

For the evolution of thinking, it is assumed, on the basis of a continuous evolution of the animal brain towards the human brain, that the specifically human thinking abilities are a further development of an intentionality already found in animals towards 'mentalization', a conception of the thinking of others that is helpful for the social human being. The extent to which this fits with the observation of a partially convergent development in, for example, octopuses, is an open question.

Without questioning its central importance in principle, the foundations of the theory of evolution continue to be critically examined in evolutionary biology, whereby some arguments have already led to the expansion of the model, while others are controversially discussed and still others are only considered interesting (or, depending on the point of view, daring) hypotheses. Arguments surrounding epigenetics, for example, have led to an expansion of the model: Some environmental conditions are not able to influence the genetic makeup, but its activity. The open question here is whether such changes are subsequently inherited.

It is discussed more controversially whether an 'extended evolutionary synthesis' (ESS) is necessary, which would integrate developmental biology more closely. [221, 222] Particularly interesting is, for example, that genetic variations may not arise entirely by chance but as a function of the environment and thus of development, and that some relevant influencing factors are 'inherited' socially or physically and thus extra-genetically; think of certain group behavior or certain physical living conditions.

The idea of plasticity, the essence of which goes back to Lamarck, but which is still being discussed, that properties could arise before their genetic fixation, is to be seen as one of the rather daring hypotheses.

Fundamentally unanswered questions exist in the area of the origin of the first living cells and the emergence of human consciousness, but for instance also with regard to the rudimentary language abilities of animals and the unusually long development of human children. In the philosophy of biology, Nagel's position, that the theory of evolution fails to explain not only human consciousness but also our unusual cognitive abilities and value systems in general, has been controversially discussed in recent years. [223]Within evolutionary biology research, however, the questions are of course much more technical; of central interest here is above all where the robustness of gene expression comes from and how traits are coded across a large number of genes (which seems to be the rule rather than an exception). [224]

10.7.2 Model A

The first thing to say about Model A is that it adopts large parts of the theory of evolution as formulated by evolutionary biology. It is additionally based on the assumption that evolutionary theory captures very deep insights about every world in which change and growth are possible. It therefore also uses 'physical evolution' as an argumentative tool, which we must understand as a very limited process in comparison to the details of 'biological evolution' presented above: Although micro-subjects differ already in principle (Darwin's first postulate) and although they can acquire further differences that contribute to some but not other micro-subjects remaining in the physical world (Darwin's third and fourth postulates), without a principle of reproduction (Darwin's second postulate), physical evolution runs into the situation we are observing of a universe that is static in terms of its basic principles. (However, the possibility of independent reproduction of biological structures – and just as important the death of successfully adapted ones – is already inherent in our physical world, among other things through the condition that entropy in general unlike in particular cannot decrease).

Biological evolution then marks the beginning of a completely new phase, and here we need to clarify the scales on which subjects act. We can be fairly certain in that in A-world subjects act on the micro-scale and on the scale of human life and we should, according to the neuroscientific explanations above, also assume that there are sub-subjects in our subconsciousness. But are organic molecules, cell organelles, cells, lower organisms, organs, plants, animals, ecosystems or even the earth organized as subjects? Since we have assumed a 'game of particles' between micro-subjects as the most likely scenario on the micro-scale (see chapters 5 and 6), an assignment of subjects to parts of cells above individual quantum systems for instance seems rather unmotivated.

However, we may find a radical change at the level of unicellular organisms (or possibly already in previously independent cell organelles, according to the endosymbiosis theory), from which multicellular organisms then emerged, for which we want to assume a structuring subject, at least in the case of humans. As far as is currently known, however, there is no physical mechanism other than the coordinated interplay of biochemical processes that controls a cell on the (sub-)meso-scale, quite unlike what we find higher-up with our nervous system. We would therefore have to assume that, although a micro-subject could have recognized the larger whole and expanded its minimal world map accordingly, this would still be a rather passive, very unfree state. Only minimal physical influence and the further development of mental structures would be possible, and the latter would still lack stimulation through organized sensory perceptions.

The most important aspect of this step would be that subjects could have become mobile by means of cells on the (sub-)meso-scale: Their world map would no longer be bound as a whole to concrete physical structures like elementary particles or molecules, but to an organism, that is a closely interwoven connection between an abstract representation in a world map and a causal network part of changing material properties. However, this also means that the evolution of simple organisms would have been practically purely material, i.e. to a large extent driven by chance events; and thus thoroughly in line with the findings of modern evolutionary biology. Whether cells are organized by micro-subjects should in any case be understood as an empirical question; an indication that this could be the case would be the observation of cell-wide control mechanisms whose robustness would be difficult or impossible to explain by purely physical processes. Such control mechanisms could then be realized via non-material markers, as we have already considered above.

Less daring seems the conclusion that, with the development of nervous systems and then brains, subjects were given the opportunity to intervene more actively in events. Initially only in a very rudimentary way, then in the sense of the instincts and intuitions of animals and our subconsciousness, and finally then in the sense of conditional human freedom. Where exactly subjects could still be found between cells and higher living beings can presumably also only be determined empirically and indirectly. For animals, this is practically undoubtedly the case in Model A, but not necessarily so already for plants. Larger organizational units than animals and humans in turn lack the possibility of direct interaction with the whole, which is why they appear unlikely or at least must be assumed to be purely passive. The driving force behind the integration of further non-material building blocks into the world maps of higher organisms – and thus ultimately the development of human thought – can be assumed to be the evolutionary benefit for pattern recognition, as explained in Chapter 7.

More specifically, Model A is therefore compatible with both the mechanism of natural selection and that of genetic drift. The role of migration and mating is emphasized even more, because this is where the freedom of higher organisms can come into play. This is generally the case with developmental and environmental factors, which is why in A-world we should vote for an extension of evolution in the sense of the EES. Macroevolutionary 'leaps' such as the emergence of a new species could be accompanied in Model A by the 'incorporation' of entirely new ideas or mental building blocks. Theories of the evolution of human thought based on a 'social brain' could certainly be incorporated as a further benefit of a world map, as could models that emphasize the importance of cooperation between organisms. (In Model A, organisms compete for material goods but share mental goods without loss). The model would furthermore certainly offer mechanisms for theories of plasticity, but whether such effects exist is an empirical question in Model A as well. The model has only a limited new answer to the question of the origin of the first cells (see above) and a trivial answer to the question of the origin of consciousness (with subjects as basic building block of reality). It argues for a continuous development from matter to life, from animals to humans, and from the nervous system to the mind, so that observations such as the rudimentary language abilities of animals and the long development of human children seem understandable.

An experiment that would argue strongly against Model A would be the successful, instantaneous construction of a living cell from molecular building blocks only, which in Model A would require a coordinated growth process of material and non-material regulatory processes. (Since we should assume, as explained above, that living cells as the basis of all further life should already be coordinated by subjects as a unit). In addition, there would be the general question of whether the evolution of complex biological rule sets for subjects would not also require non-material 'genetic input', or whether this could be realized entirely indirectly via material factors, which would then become instructive for non-material factors. In the first case, life could then really only arise from life. In the second case, evolution would have 'played it safe' and relied only on the more reliable mechanisms of material inheritance.

10.8 Does Model A extend the established ideas of biological evolution in a scientifically inadmissible way?

Here we must ask ourselves critically whether Model A, even if we concede that it is compatible with the theory of evolution, might not turn out to be more of a danger to the scientific discourse in evolutionary biology, in the sense that a meaningfully closed and very well-mathematized theory is softened and would thus rather lose potential for understanding and criticism.

In the sense of this objection, with Model A, we would close our eyes to what is actually most interesting about the almost optimal, evolutionary minimization of the error costs of a contingent genetic code (Crick's 'frozen accident'), or about ontogenesis (the development of the individual organism), which takes place without any central control and yet reliably and predictably, i.e. we would close our eyes to the information-theoretical complexity of these processes. This would then amount to the attempt to short-cut the challenging marathon of the biosciences with a cross-country run.

And here, indeed, the greatest caution is required, also and especially in the interest of the defenders of Model A: The possibility of replacing 'blind' evolutionary processes with 'informed' ones, i.e. assuming processes in which there can be feedback between physical and mental entities, should not be used lightly. Model A does not advocate grand-scale 'intelligent design', but rather tiny steps by individual subjects towards more growth. (The role of God as the creator of the 'set-up' remains conceivable either way.) Thinking broadly, our world has thus been structured by ideas, but necessarily via a material world that is itself physically-causally structured. Teleological elements are added here only in bits and pieces over very long periods of time, as regulating ideals of physical processes. It was only later that organisms were able to develop mechanisms that allowed them to exploit every little bit of developing freedom, e.g. first through mate choice.

Whether and where it is possible to work meaningfully with ideas such as Goethe's 'ideal leaf' in plant metamorphosis, a goal-oriented 'orthogenesis', or with extended teleosemantic concepts, can only be worked out by biology itself; and only in productive confrontation with purely materialistic explanatory approaches, such as those considered in Kauffman's *Essence of Life.* [225] This is, of course, conceivable for questions concerning the origin of the first cells, the existence of general developmental rules as 'styles' behind very different evolutionary processes, as well as mechanisms of gene expression, plasticity and ontogenesis. But only where an extension in this sense is still the most economical hypothesis, only there should it be dared; then, however, it is not scientifically inadmissible; then it allows to argue against religious, esoteric or even magical ideas for biological explanations.

10.9 How can we imagine the individual development of a human brain with Model A?

As written above, this model would certainly advocate closer cooperation between evolutionary and developmental biology because the increasing freedom of organisms is expressed in their development as a reaction to environmental experiences. However, developmental biology also plays a central role in Model A because the development of the mind cannot be directly materially caused and therefore cannot be directly genetically coded. With

Model A, we must assume that the development of a mind, especially a human mind, is only possible via a complex coupling of physical processes in the brain and mental processes in the mind. Physical processes can only be 'suggestive' for the development of certain mental structures, just as higher mental structures ultimately always have to be developed by the subject or initially the sub-subjects on the basis of the mental entities already available to them (as we have already considered for learning in Chapter 9). A large number of interesting research questions are conceivable here, especially with regard to the early stages of human development. [226]

If biology had understood the entirety of these coupled processes, it should be possible, according to Model A, to create 'real' artificial intelligence on the basis of alternative physical structures; however, such an evaluation of Model A is most likely still a long way off. The possibility of the inheritance of non-material structures, briefly mentioned above, could then become very relevant. We should perhaps think of it as being similar to the direct adoption of metabolic properties. But as with the assumption of new evolutionary mechanisms, great caution is also required with such proposals. For the time being, the controlled construction and reliable transmission of stable structures seems conceivable only via the material detour.

Overall, Model A appears to be at least compatible with the accepted viewpoints of modern neuroscience and evolutionary and developmental biology across a whole range of questions. The central prediction of Model A would be, amongst other things, that no complete material realization should be found for higher cognitive functions. Great caution should be exercised in prematurely extending the established theory of evolution to include further mechanisms.

Chapter 11

Value theories and current aspects

At this point, we need to pause. In the previous chapters we have seen that it seems possible to arrive at a model such as the one proposed both from a more scientific direction, starting from minimal subjects and objective qualities as the most fundamental, ineluctable units of knowledge; and from a more philosophical direction, starting from the epistemological problems of bridging the gap between subject and object. For such a model, the natural sciences open up to the non-material world, remain naturalistic, but no longer only materialistically oriented, and philosophy opens up to a closer integration with this world view in the sense of an inductive metaphysics. In contrast to the discussion of the 'narrow' mind/matter problem of the integration of mind and brain, with the 'hard' minimal problem of qualia, the above approach aims to understand the 'broad' mind/matter problem in the sense of an integration of the material and non-material world, also with regard to universal mental content. It can now rightly be asked whether the proposed model has succeeded in doing this, or even whether such a model can succeed at all. The problem is taken to the extreme with the question of whether ethical or aesthetic value judgments can be meaningfully conceived and implemented in it.

The following considerations are not primarily concerned with the specific 'solutions' proposed, but rather with illustrating that Model A allows a meaningful connection to ethical and aesthetic discourses, however these are designed in detail. Ideally, Model A would allow us to build a bridge between the discourse projects of the natural sciences, the social sciences and the humanities, in which all parts retain their individual validity, i.e. are not drained of their meaning. Accordingly, Model A does not aim at an objective idealism in a rationalistic maximum form, but 'only' in a pragmatic minimum form; without eternal truths and dependent on the continued philosophical-scientific and philosophical-social dialog. Model A thus does not stand for a return to Hegel, but precisely for the avoidance of a universal philosophy of nature, mind and, subsequently, of history, which by always already having an answer to all questions forfeits both the critical and the innovative potential of more ’cautious’ approaches.

11.1 The possibility of ethical thought and action

By design, Model A provides values as non-material building blocks and thus objectively. But these values are to be understood as qualities such as colors or feelings such as pain, i.e. initially completely ’naked’ and without a direct, intellectually defined extension. Such values must therefore always be imagined in practice with a historically grown ’corona’ of further qualities, which defines the extension from the very general non-material to the very concrete material for a specific context of life or groups of such contexts. However, this is not to be understood as a ’weakness’ of such an ethical or aesthetical value; although it can never be fully grasped in words, it offers conversly a very broad understanding of itself, with itself as the ultimately unreachable goal.

The necessary historical ’decoration’ of values then means that individual values cannot stand on their own; without further values as a corrective, each individual value can be overloaded to the point of absurdity. This could be found, for example, in a world that is only good, but not true or beautiful, an illusion of the good; or in a world that is good and true, but not beautiful, and so on. More specifically, an autocratic regime can claim a historically corrupted good for itself as the good that serves the regime, but this will hardly stand up to a correction that also demands truth and beauty in the sense of aesthetic diversity (more on this below). But what gives our ’coordinate system’ of truth, good, and beauty its special significance compared to the infinite number of other conceivable values? And even if we can find a justification for this, what gives this coordinate system its normative power? With objectively existing values, we have only postponed the problem of the naturalistic fallacy that an ’ought’ cannot be derived from an ’is’; in addition to their pure existence, values, or a coordinate system of values, would also have to excert some kind of binding force.

This leads us to the even more fundamental question of what overarching meaning and thus what normative forces A-world could have at all. On the one hand, we have minimal but fundamentally free subjects, separated from each other by the fact that they can only interact meaningfully with each other via an objective world. On the other hand, we have this objective world, which can only gain complexity and diversity through the work of the subjects.

What can be the point of all this? In A-world, one would have to say: That something is instead of nothing. That this something is not uniform, but diverse. And that this diversity is not static, but develops dynamically. And this seems to be inherent in exactly what we observe: The existence of subjects and objective qualities. That objective qualities can enter into changing constellations. That such changes can be freely brought about by subjects. In short, the growth and flourishing of the whole appears to be the only conceivable meaning, if no particular events or goals are to be arbitrarily singled out instead. (In this sense, the question of the ought from the is is not one that is independent of our metaphysical ideas.) It seems difficult, for example, to postulate any normatively distinguished state in the past or future of A-world, regardless of whether this is understood as a creation to be preserved, a state of lasting redemption or the perfect unfolding of abstract principles. Model A certainly allows the connection to such discourses; however, further arguments are needed here for the meaningfulness of the assumption of such ideal states.

Nevertheless, even the minimal sense of 'flourishing' outlined above would be sufficient to bring our coordinate system of values into its special position: Such growth requires the development of a multiplicity in unity, of individual parts in a whole, and this in turn requires the recognition of the individual, the understanding of the identity of the particular, i.e. an idea of truth (in this sense also as a 'basic duty' of ethics). And it also provides a direction for further development, thus enabling the pursuit of a good. Finally, in view of the minuteness of the part in comparison to the whole, it gives beauty its particularly important role: To see the contribution of the individual to the whole. (This can then be the overflowing variety of leaves, rippling, glittering in the beauty of nature, as well as the opening of new intellectual paths in the masterpieces of modern art; and in this way, even an aesthetic of ugliness can as well be thought of as enrichment.) The overarching meaning of the florishing of the whole thus lends our coordinate system its normative power: To turn away from the coordinate system means to turn away from this florishing; it means to choose nothingness instead of something. Not wanting to be part of it, not wanting to florish, is possible, but only at the price of withdrawing from this very florishing. The normative power of values is therefore not principled, but not following it is in practice ultimately self-contradictory. This normative power is not absolute, but it applies to us. It does not apply to all possible worlds, but to this unique one.

The concept of growth outlined above is of course not unproblematic and it is precisely here that the bridging envisaged in the model relies on the independent validity of ethical discourses: Is the perfect killer an enrichment? How should a competition of growth processes be evaluated? How do we evaluate the growth of different organisms relative to each other? How do we evaluate the contribution of individual events in relation to future possibilities? How are ultra-rapid fluctuations in growth and destruction to be evaluated? Is 'growing into each other' or 'growing side by side' of particular value? In other words, how should we generally evaluate the individual contribution, as a co-constituting part of the whole, in relation to the whole, as the sum of contributions? Obviously, our concept of growth needs to be further developed. Complex growth may well be seen as more enriching, e.g. the development of an inhabited planet instead of even more uninhabited ones, but such growth will usually only be possible on the basis of stable fundamental elements, as is already the case with elementary particles. How much freedom and how much stability and thus necessity does productive diversity require? How much reproduction and how much variation of the same? Without further arguments, one would probably have to assume that in A-world the concept of growth itself grows with it.

In the physical world, florishing then primarily means more of the same, but in the biological and cultural world, larger leaps in growth are possible, precisely because they are anchored in the physical world on the one hand, but can increasingly change their own elements 'disruptively' on the other. The existence of ethically and aesthetically capable organisms such as humans would then be a 'miracle'; made possible, but by no means guaranteed, by the basic outline of the world. Accordingly, this miracle would have to be protected and preserved at (almost) all costs especially with a view to its future possibilities. This, however, would simultaneously require the protection and preservation of large parts of the 'whole' as the lifeworld of these organisms. If, on the other hand, one would not follow this and would rather want to allow further arguments, these could of course also be of a theological or idealistic-ethical nature; in the latter case, for example, oriented towards Green's perfectionism. [227]

11.2 How is ethical action conceivable for the individual?

What then is our place in the whole? The most obvious answer is: To contribute to this growth our very own part. However, this should be understood in a much less elitist way than one might initially fear: What we do is always already a joint achievement; we owe almost everything to others, just think of parents, teachers, employees, healers, suppliers and disposers, etc. Additionally, we always contribute not only to our growth, but also to the whole. We cannot know whether the direct part of the whole, our own growth, or the indirect part, our contribution to the growth of others, will be of greater importance, especially as practically all growth that could still result from our actions lies in the hands of future generations. Our legacy (a child, an institution we helped to build, a theory or a story, but also a stranger I helped) can seem great at first and then suddenly fade away, or remain dormant for centuries before it becomes relevant for everyone. In any case, our own growth is already contained in the meaning of the whole, and can therefore be pursued also for its own sake – except that even this tender little plant, if it is to bear any fruit, will be aiming to interact with the whole.

We can therefore only accept ourselves in our world as quite unique on the one hand and strive to increase our freedom on the other and then use it in pursuit of our further growth as part of the whole. Here, truth, good and beauty are beacons for us in the chaos of possibilities and entanglements – less consistently sanctioned, but also less helpful in individual cases than for the world as a whole; and yet the best sea marks we have. What exactly we contribute is predetermined in many ways, but is always also the consequence of our free actions. This freedom is to be thought of as participation in a common freedom and thus includes the freedom of others. We can be proud, first of all of the joint performance, if we have contributed our own part to the best of our knowledge and belief (roughly in the sense of act-consequentialism). We should feel shame when we use our freedom to destroy what is good, true and beautiful in all its diversity. The suffering of others as well as our own pain are clear signals to us here.

Already our basic philosophical impulses (according to Jaspers with reference to Plato) are signs of this free striving for growth; to wonder, to search out of doubt, to help ourselves and others out of devotion. The fundamental possibilities of the exploration and exploitation of object possibilities and the competition and cooperation with subjects are open to us, which leads to the central role of creativity and love for our florishing. In Model A, the subject's need to be seen as an individual is not only a psychological one, but already a metaphysical one. The successful growth of the individual does not necessarily have to be expressed in the form of peak performance, but can also take the form of networks (think of care work, for example) and all intermediate forms.

The physical and social conditions of our growth restrict us, but also liberate us; for example, we age with our body, but our body also makes the complexity of our thoughts and actions possible in the first place. It transforms our absolute but negative 'freedom from' as a core subject into a conditional but positive 'freedom to' as a whole person in interaction with our other life circumstances. In Model A, the conditions of this freedom also include normative ideas, as objectively existing, socially shaped building blocks of our world maps and thus of our person and society.

We can act more freely and thus more conducive to our growth when external and internal conditions give us the space to do so; when we are not starving, driven by hatred or mentally ill; when we have a healthy body, conducive living conditions and a rich spirit. This leads to an asymmetry: We are most responsible for our true, good, beautiful works that are the result of our own free decision. But we should show understanding and empathy towards our bad deeds and those of others, as they will generally be signs of a lack of freedom in the sense of a lack of intellectual, emotional or physical possibilities: The person acting in such ways cuts others off from opportunities for growth, but always also themselves. (This does not call into question the necessary prioritization of victims, nor the question of the necessity of condemnation and sanctions).

In many cases, however, our potential deeds can not be so clearly classified as good or bad. Here one would have to argue with Model A that there are, for example, genuinely unsolvable ethical problems; that the question of whether I would rather let one or five people die, etc., cannot in principle be answered 'correctly'. The ultimate goal of free individuals would then be to recognize the inescapability of these conflicts and to avoid such situations through prevention if at all possible (think of the difficult issue of abortion). Although it will then usually be possible to weigh up various evils, this can only help in choosing the lesser evil and can ultimately only be decided by the person(s) concerned. (If we meet a person in such a distressed situation, the first thing to do is to show them understanding and empathy and to offer them our support for the difficult path they now have to take.)

Again, the above should only be understood as one conceivable minimal model of many possible ones, with which, starting from Model A, a meaningful connection to ethical discourses could be found. If I am prepared to admit further, for instance religiously or idealistically motivated arguments, then the spectrum of possible discourses widens accordingly.

11.3 Can we speak meaningfully of virtues with Model A?

With Hippocrates and Leibniz, '*sympnoia panta*' generally applies to Model A – all things must come together so that our more or less noble intentions can find their realization in A-world: All things are, after all, connected to each other as a whole via the causal network of the physical world. (The basically not incorrect observation that we need certain 'constellations' in the world in order to achieve our goals has accompanied mankind for thousands of years in the form of astrology; only that here the practically necessary oversimplification leads the practice *ad absurdum*).

In Model A, too, many growth opportunities are to be understood as 'transformative experiences', [228], i.e. as experiences whose meaning only really becomes clear to us once we have had them (think, for example, of having children). Our growth is therefore necessarily based to a large extent on 'trial and error', which is why the classical virtues aimed at our coordinate system of values take on their significance. A certain amount of courage to try and perseverance after making mistakes, but also prudence and moderation, are essential for creating and successfully seizing opportunities in such a complex world. Justice and kindness are then necessary for the success of joint growth. This results in moral duties at all levels: To preserve the world, as well as one's own body; to cooperate benevolently in dealing with fellow human beings, etc.

A virtue ethic of growth (not just physically conceived) would finally have to emphasize that we are all weaving on the same web, that we should celebrate the successes of others as successes of our team, that we can strive to play well, but that we should not take our failures too much to heart. (Wisdom then simply means having acquired a rough idea of when it is worth taking a small hop in the ever-evolving constellations of the world). Here, Model A catches up with the important transition of ideas from the Ancient to the Middle Ages; the 'hero ethics' of the successful individual is expanded to include the originally religiously mediated idea of a meaningful significance of the whole (as faith) and of each individual life, including the failed (as hope). Failure here means above all the failure of life plans, of possibilities already explored in our world maps, which must be mourned and then abandoned. Hope, on the other hand, exists until the end, because there are always undiscovered possibilities left. It is to note here again that the above must be understood as a possible minimal outline of how virtue ethics could be discussed following Model A.

11.4 Can we talk about happiness in a meaningful way with Model A?

In contrast to emotional happiness, which is also important for our growth, the satisfaction of a successful life is emphasized with Model A as a goal worth striving for. But what success then means is much more colorful than what is classically associated with the concept of *eudaimonia* ('welfare'). In Model A, the essential quality of people is not their capacity for reason, but the attainable positive freedom to love and create – which, however, will generally depend heavily on their capacity for reason. Physical, social and mental goods thus stand side by side for the time being without ranking; philosophical contemplation does not occupy a special position in the abundance of possible true, good and beautiful life plans. The diverse, the original, even the strange ('all things counter, original, spare, strange' writes Gerard Manley Hopkins) are an essential part of the growth of the whole, just as, conversely, the beneficial embedding of the diverse, original and strange in the whole is an essential part of the growth of the former. This does not have to call into question our practice of socially recognizing individuals who, through their top performance, give a face to important growth processes with their very own contributions; it should only always be considered with a corresponding footnote.

The success of a life plan requires the setting of goals, which – since the goals are then always already partially realized in the world map – can cause demands, frustration and anger if physical and/or social conditions require a correction of my world map. The central goal of self-empowerment of individuals starting with school education would therefore be to show them how to choose conducive goals. The reference to the big picture and our individual contribution can only be a starting point here, because the individual must of course shape their life with very specific goals and under very specific conditions. This seems especially important also because it is well known that people are often not good at choosing goals that, in retrospect, were actually conducive to their personal growth. In a way, this can be seen as a natural continuation of the shift from a focus on knowledge to a focus on competencies and now goals.

First of all, the individual will need time for this – in addition to all the known pedagogically beneficial conditions. A number of subgoals will always be associated with each goal found in this way: The goal of conducive social circumstances or that of maintaining one's own health for instance, as all of these will generally be conducive to achieving my individual goals. Here it should be considered whether preventive psychoeducation (as many young people are already acquiring privately today) could not help in the choice of individual goals; to be sensitized to the processes of one's own psyche and to receive hints on how these psychological processes can be better understood in order to enable acceptance and perspective, i.e. to increase or restore individual freedom of action in the sense of Model A. [229]

11.5 First conclusions

From a meta-ethical point of view, Model A thus leads to an interesting construction: Deontological in its basic construction (the good ultimately receives its normative force by the fact that it is already available as an objective building block), (act-)consequentialist in relation to the big picture (the good is the growth of the whole), for the individual, however, oriented towards the asymptotic approach to universal values and virtues derived from them as the best available sea-marks for the good.

There seem to be interesting parallels here to Iris Murdoch's ideas in *The Sovereignty of Good* in particular, although their conclusiveness still needs to be investigated further: This concerns not only the idealistic foundation, but also the central importance of the human psyche for ethical considerations, which gives the 'naked' but broad values their concrete form against the background of a lifeworld, as well as the observation that people's freedom increases with their knowledge.

To summarize, it can be said that Model A could make meaningful ethical thought and action comprehensible: Ethical (as well as epistemological and aesthetic) scepticism appears as a practically necessary but self-contradictory position. A 'coordinate system' of universal values is emphasized, but the contingent shaping of these values against the background of concrete lifeworlds and thus power structures is always already given, so that ethical theorizing must nevertheless be understood as a project to be continued.

With this result, Model A can also address the fear of modern societies of the 'undemocratic' tendencies of value Platonism postulated by Hösle: [230] Values are given to us asymptotically, but further convergence requires democratic agreement on the progressive discovery and development of the objective standard that acts as a reference point.

Model A should therefore be used in international discourse to argue against relativist positions and in favour of universal values, except that these should not be understood as a form of immutable (power) knowledge about all of humanity owned by the West, but rather as a continous project of approximation to the core meanings of these universal values, for which diverse voices, especially also from the Global South, must be understood and recognized. Particularly with regard to the practical implementation, however, there is an urgent need to drive this project forward, especially also within Western societies themselves.

11.6 Towards more concrete questions

Up to this point, the discussion has been very abstract, which is why it will now be concretized using three current examples, if only to pre-empt the arguments made against Rawls that working with universal concepts always either collides with at least some concrete traditions or is useless in practice. The first attempt will be to use the above proposals to mediate in the postulated conflict between identity politics and universal values. Secondly, an attempt will be made to use Model A to better understand the observed polarization of Western societies. Thirdly, the climate justice crisis of capitalism is addressed. At this point, somewhat unexpectedly, we will find that aesthetics will have to play a central role in our considerations. And then finally, we will return to our initial topic of information and intelligence, pondering the open question of the whether artificial intelligence can contribute to overcoming the three mentioned crises.

11.7 Identity and universals

The postulated conflict between identity politics and universal values is based on two quite justified concerns: On the one hand, the concern that to take identity-related considerations into account would lead to a relativism that is detrimental to our search for truth as such, as it would mean to self-contradictorily reject the high good of universal truth. And, on the other hand, the concern that accepting universal truth claims would subject the search for truth to a detrimental dogmatism propagated by those who benefit from not making their socially dominant identity explicit.

The above-mentioned minimal value model for A-world allows us to critize the first position for the fact that universal truth itself can only be a reference point for its concrete formulation. The consideration of the different world maps, the identities, is therefore not only not detrimental to the concrete understanding of universal values, but in fact indispensable; reason has bodies, fellow human beings, its time. The aim must be to allow the diversity of identities to contribute to further improving our universal coordinate system.

The second position is to be criticized for the fact that whoever claims an identity already claims a truth with universal aspiration; without it, talk of identities is self-contradictory. The aim must be, to develop our universal coordinate system accordingly by incorporating diversity.

As a result, those who think universally should also try to make their identities explicit and critically question the social functions of their arguments. This would mean to first think about power; what is considered a valid argument is often already a question of subject cultures, for example. And second to diversify the discourse as much as possible in practical and concrete terms; universalists will often be the ones who manage the opportunities to do so. In line with the above, taking diversity into account will not harm the concept of truth.

Conversely, those who think primarily in terms of identity should try to strive for an intellectual emancipation of diversity and not a tyranny of vested interests, i.e. to discuss questions of power where this is fruitful; it will not be the case for every, though admittedly for many universally conceived arguments. So how does diversity contribute to gaining knowledge? In all contexts through active participation, but not in all contexts through explicit thematization. Recognizing universal values does not erase diversity.

The end result, namely that national and international democratic discourse on values must be continued at all costs, still seems very abstract, of course. We would now have to delve into the individual discourses. As an example, we can take a look at the extreme case (in terms of complexity) of the debate about biological sex, social gender and the introspective sexual self. In Model A, it seems almost necessary that we find the evolutionarily essential variation of characteristics not only in relation to physical, but also mental ones and, in particular also to the coupling of these characteristics. A 'gender spectrum' would thus be expected, though in Model A it would be more fitting to speak of sex-, gender- and self-image-spectra, the latter in the sense of a rather special hypothesis of sexual essence. And also the in terms of numbers rather rare trans-coupling of physical biological sex and mental sexual self would not come unexpected. (Even a double or completely different coupling or no coupling at all would be readily conceivable.) This clearly distinguishes the model from a materialist view of the problem, in which such a coupling can only be interpreted as a developmental-biological 'error'. The intensity of the debate arises from the fact that those affected and those who doubt this operate with different world maps, each of which is objectively real for those involved. (The issue of sexual orientation should be assumed to be orthogonal to the above, but could most likely be treated analogously).

In Model A, the world map of those affected is the result of an evolutionarily normal variation of the possibilities of our human existence; an essentially important process that ensures the survival of life when environmental conditions change drastically. The person is simply 'born this way'; thrown into a life without a choice of starting conditions, and, depending on the exact combination of traits and coupling(s), will either be able to accept their socially unusual nature as completely individual, or will want to change the physical parts of their person to bring themselves into balance, because to the core-subject the mental part must always seem more essential. It seems more likely, however, that the individual will live their life somewhere between these extreme positions, fundamentally dissatisfied and punished for their fate by large sections of society.

In constrast, the world maps of the doubters are completely cut off from this knowledge. Since the structure of our world maps is always physically mediated and the information about the body of the other person appears unambiguous against the background of our pre-existing world map, the objectively real image of the other person's body in our world map appears unambiguous to the doubters, too. The body of the other person, as well as mine for him, is not only a hybrid object but also a Janus-faced one; based on the same causal network parts it can experience a substantially different completion into a whole object in different world maps. Furthermore, our pre-existing world maps are not only physically informed, but also socially and historically structured. The core problem then is that the correction required here must intuitively appear as objectively wrong to the doubters and that they will in all probability put their entire rationality at the service of this intuition, because in Model A the consistent world map is the most powerful tool of man. As with all diversity debates, it is now necessary for the doubters to engage in reflection; to recognize the conflict and their own limitations as well as those of others.

In any case, the greater suffering is experienced by those affected; even if it is possible to overcome the confusion and doubt about one's sexual self, and even if one's own body, whether modified or not, can be accepted, the social conflict remains: Ultimately, it is not so much the body that is the problem, because the core subject perceives it via its world map and thus first of all 'correctly'. Rather more problematic is the – at least in our current social situation – *inevitable* connection between the signals that the functioning of the body in the causal network of the physical world sends out and the accompanying – also own! – 'sorting' of the person according to our socially constructed gender concepts. This results, among other things, in the important role of clothing for self-perception. The socialization that would be associated with the 'right' body and that could make the person 'whole' according to current standards is most often denied to those affected today. (This has thus a substantially different effect than the 'wrong' coupling of skin color, which is also discussed in the literature and is indeed also conceivable in Model A; these two cases should therefore not be lumped together.)

What has been said so far concerns the case that biological sex and self-image are in fact not congruent; quite apart from this, of course, it must be taken into account that the process of establishing a human identity is not a straightforward one without errors and doubts. Irreversible interventions in this development should therefore only be made on the basis of careful consideration and consultation. The fact that the debate is being conducted must nevertheless be seen above all as an important step towards a better society; at its core, this is as always about individual human dignity. In my opinion the discussion can additionally be seen as an argument in favor of idealistic models: With such models, the debate gains selectivity and its emotionality becomes more understandable.

Closely related is the problem of our physical beauty and body ideals, which plays an important role in many feminist considerations. Irrespective of the fact that there is certainly a separate dialectic of making the female body available for male pleasure, which repeatedly escalates catastrophically for women, physical beauty should be seen with Model A as much more closely linked to social and intellectual beauty: In our map of the world, people are never entirely reducible to their physical attributes; there is always an 'aura' of their social behavior, achievements, and so on attached to them. The biggest problems with our ideals would then be that, on the one hand, we keep fuelling the dream that they could all collapse into one person and that, on the other hand, we contribute to the objectification of (mainly women's) bodies.

Also closely related to both topics is the problem of cultural appropriation, particularly in the field of fashion, but also music or dance: Not infrequently, it is those who are excluded from the central power structures, and therefore have few other resources at their disposal, who 'invent' fashion in order to materialize parts of their world map, i.e. their very individual identity. However, those who are better integrated into the central power structures have the prerequisites to market this fashion. The adoption of fashionable inventions is nevertheless also a recognition of such contributions and thus of the identity of the (now somewhat less) excluded. As with beauty ideals, the problem of cultural appropriation must therefore be about avoiding objectification and enabling mutual participation. Which also means that some things might not be open to commercialization.

Fundamentally important for all three problems addressed so far is a mechanism of 'aesthetic packing' of qualities in the design (not only of values) in our world map: For reasons of efficiency alone, but above all to enable 'narratives' conducive to finding (growth) goals, we weave comprehensive bundles of qualities that then appear to us as one and inseparable:

Just as our historical-social and thus contingent idea of good appears to us as completely clear before in-depth reflection, also established 'packages' of identity characteristics like traditional images of women and men, body images, or images of 'the others' can appear to us as an inseparable whole and as such as an objectively recognizable building block of reality. The world map of the subject is, after all, a structure that is only subjectively accessible, but objectively existing. (The purpose and problems of this 'aesthetic packing' will have to be discussed again below.)

With Model A, Western societies would have to be conceded that in the past they have, although not without major setbacks, gradually aligned themselves better with the goal of shared growth. But so far, they have failed to do so in the international context, where the 'empty' diversity discourse must then be misunderstood, especially by the Global South, as an imperialism of the mind. Furthermore, even within its own societies, the West no longer seems to be able to take the next steps; instead of gratitude and joy over immense prosperity, freedoms and aesthetic diversity, we observe a polarization borne of anger and hatred; which brings us to the second example.

11.8 Social polarization

The interesting thing about the polarization in question is that it seems to emerge according to a psychological understanding of progressiveness and conservatism; [231] practically as 'new soul' vs 'old soul'. Thinking in terms of Model A, it is not arguments that clash here, but life concepts that are not based on the selection of clearly separable parts, but must be understood against the background of entire world maps that are to be kept as consistent as possible, i.e. physically and socially mediated, but completely individual and yet objectively real worlds. Even more so than our bodies, social institutions are Janus-faced objects that may exhibit substantial differences between political camps. Scientific evidence that these institutions indeed rely on a 'social imaginary', i.e. a set of mental entities on the basis of which individuals experience their society, and in which agonal value orientations play a role, can be found, for example, in Hofstede's theory of cultural dimensions, Inglehart/Welzel's world map of cultures (traditional vs secular-rational orientation) or more generally Schwartz's theory of basic human values, as well as the Rokeach Value Survey of universal values, etc.

This is not to say that (especially economic) inequality and/or insecurity do not play an important role. However, these factors would be at work indirectly by increasing or decreasing the individual and collective freedom to grow – as envisaged in the respective world map. On the one hand, this would explain why alternative paths to growth as part of a traditional family, nation or religious community play an equally important role in polarization, and on the other hand, why polarization is less pronounced in continental Europe, where the financing of important basic conditions for freely chosen life plans is much more socialized, considering for instance health care or education. (The difference between two-party and multi-party political systems is likely also playing a role.) One would have to conclude that redistribution should primarily be aimed at enabling freely chosen life plans, which in turn can be achieved primarily through rich common goods such as freely accessible medical care and education.

According to the above remarks on Model A, both sides, as well as society as a whole, which is less resilient due to polarization, would be well advised to leave behind a materialistic view of the conflict for the time being: In defending the traditional values that are important for their growth, conservatives need not fear the complete annihilation of these values, which are, after all, objective and universal, though in their historical-social understanding also subject of the democratic discourse. The defense of progressive values, on the other hand, requires such an understanding of values, to immunize itself against the accusation of relativism, i.e. ultimately the materialistically conceived, instinct-driven abuse of power.

The necessary combination of universal and historical perspectives would thus safeguard the debate against populist ideas such as the 'integralism' of a clearly definable and historically enduring religious national culture of the West. However, it would also sensitize people to conservative arguments, including those against the formation of elites based on a concept of meritocracy that is too narrowly conceived. The latter becomes immensely dangerous when elites have progressed to the point where they do no longer see any danger to themselves in internal or external takeovers of society.

With Model A, one could therefore argue that the excesses of populism and elitist abuse of power are also a consequence of our current materialistic world view, which denies our effectively balancing and binding coordinate system of basic values, sees instead only an agonal plurality of them, and thus torpedoes the democratic discourse on values as such. Once a more idealistic discussion has been initiated, the problem of the correct intertwining of equality and diversity plays a central role. The sub-problem of economically conceived equality is particularly topical and important due to the climate crisis, which is why it will serve as our third example in the next section.

11.9 The climate (justice) crisis: daring less capitalism

Economic equality can only work in the long term on the basis of sustainable economic activity, which is why the current overuse of terrestrial resources demands that fundamentally important ecological aspects be taken into account in the discussion of fair economic activity and vice versa; climate-neutral and fair economic activity become two sides of the same coin.

In view of the excesses of the elite formation mentioned above, which is based on a concept of meritocracy that is narrowly conceived in material terms, the question arises of whether democracies and market economies are capable of effectively and efficiently implementing the solutions that do indeed exist. Whereas in the past political elites held all the power in their hands, now it seems to be economic elites who have steered the development of the system in their favor in such a way that the problems seem practically unsolvable despite best intentions. In purely mathematical terms, this is undoubtedly wrong; the money, the technologies and the manpower for it are available – but they can hardly be mobilized for the greater good. This observation rightly calls private property, meritocracy and the market economy into question; it is just that the alternatives are still less well suited for solving problems of such complexity.

Like sedentarism, writing and science, and in fact all tradition, private property is also a practice of retaining mistakes that allows these mistakes to be subsequently recognized and dealt with. Thus, at least a certain amount of private property, even if only as a guaranteed right of use, seems ultimately unavoidable for responsible economic activity.

And while talent and time invested will normally be Gaussian distributed, success will normally, that is even without malicious intent, be distributed according to power laws (according to 'hockey stick' graphs), due to self-reinforcing network effects. (Whereby success, however, only loosely correlates with performance, which is why it is so important to give one's own luck sufficient opportunities and to also look out for performance outside of the top lists). The fact that meritocracies are in principle always in danger of running out of control should be a reason for a fair redistribution of the opportunities that are ultimately always generated by society as a whole; but to turn away from the merit principle altogether would be to throw the baby out with the bathwater.

Finally, market economy models are certainly flexible enough to be part of the solution, especially since, unlike strictly socialist models, they were and are never completely materialistically conceived in reality, but were and are usually politically framed by noble, religious and/or enlightened, ideals. (The fact that the economy itself is increasingly concerned with non-material goods is also very conducive to solving our problems; for 2016, 'peak stuff' has even been proclaimed.) Material resources, labour and attention are never completely outside the control of the individual and thus of society; the interlocking of market economy and socializing democracy is therefore not fundamentally in need of reform, though the specifics clearly are – and not a little.

The author would now have to contribute his own opinion to the democratic discussion in more concrete terms: A temporary wealth tax to finance the necessary restructuring seems to him to be the simplest solution, while corresponding consumption taxes with compensation for low incomes would probably make more economic sense. Whether and when such solutions are possible certainly depends first of all on the level of suffering, which unfortunately is usually lowest among the main culprits. The most important insight is probably to recognize that many problematic processes are ultimately not self-driven, but are driven by us and can therefore be shaped differently through collective decisions. We are for instance only beginning to fully understand the extent to which legal frameworks influence the (unjust) allocation of resources. [232–234]

Up to this point, we have not even considered that with Model A we could also argue for a more comprehensive 'renovation' of our economic value system and then go, for example, with Hösle [235] not only for ecological and social tax reforms, but also for a new legal category of the organic, for which people, unlike for things, could only have ownership in the sense of a right of use, but not a right of possession, where possession implies amongst others the right to destroy.

But what does all this have to do with Model A? First of all, the model can make it clear why in such debates not arguments but life plans collide. The situation itself does not force us to choose between (turbo) capitalism and socialism; we can 'dare less capitalism', that is unravel the overall bundle of 'capitalism' owed to our impulse for aesthetic packing and see what works and what does not. However, our answer will not be context-free; it will have to be different in the Global South than in the West. Nevertheless, at least so far, democratic market economies have proven to be more adaptable, that is less self-limiting, than autocracies of any kind; and this primarily by tapping into the potential of their diversity. Only that they are not selfregulating with regard to the framework conditions of economic activity, but require active political participation.

And the above has to do with Model A in yet another respect: It shows where the meaning of physical growth, especially as having more and more, has a limit and it emphasizes the possibilities of alternative, social and intellectual growth. A successful life then revolves around the growth of a whole person, not only of their body and possessions, but also of other, equally objectively existing aspects of their world map, such as the development of virtues. Model A thus offers a much richer background for future, fairer and more sustainable life plans than materialistic models can do. Admittedly, life is based on a certain amount of surplus, including material possibilities; the pursuit of this is not without substance or useless – but what a certain part of the Western world lives and experiences is simply too much of a good thing: The benefits of more goods for further growth as a person are becoming increasingly smaller, so that even more resources are then only accumulated as confirmation of one's own life achievement, where immaterial goods have lost their value. [236–238]

In particular, for example, as the saying goes, youth is wasted on the young, but wealth is often wasted on the old. The same applies to the Global South versus the West. (Here, every golden toilet is a lesson in impotent growth anyway.) Seen in this light, the Anthropocene merely marks the beginning of the hard part of the enlightenment; enlightenment for adults, so to speak, whose central task is now to take responsibility for all life. This will also require humility, accompanied by modesty in terms of economic expectations. Model A shows that this can be a gain as well as a loss: Diversity makes my individual world map objectively more beautiful; if I learn of the suffering of others, this makes it truer; and if this suffering can be reduced, it makes it better. The associated problems of sustainability and justice are a problem in the material world where growth processes compete; in the non-material they are not necessarily doing so.

Even for ethical considerations following Model A, the question of how political power can be organized to intervene in the crisis in a regulating way remains central; individual consumer behaviour will not be able to achieve sufficiently large effects. The ideologies in the wake of classical idealism, the strong leader or the communist party, hardly recommend themselves as solutions according to Model A. After all, the crisis is not of such a nature that it could not be solved by the liberal constitutional state, which in addition seems to be the only option with conditions that allow for the provision of the necessary competences. But if we can simply choose the solution via our politicians, what is stopping us? More and more, unfortunately, the identity-based polarization of our societies, which brings us full circle to the first two problems mentioned above.

According to Model A, the basic conflicts would be of fundamental nature: The anchoring of our existence in the material world already implies that at some point limits will be reached and further use will have to be negotiated. Our various options of exploration and exploitation for achieving growth already imply the competition between progressive and conservative strategies. Finally, the pursuit and maintenance of identity is what makes growth possible in the first place. The core of the identity problem is the absolutization of my world map; the core of the polarization problem is the ignorance of the world map of the other; the core of the resource problem is the supposedly possible decoupling of the world maps from one another.

The normative work that is supposed to bring us to action can then not take place in such a way that I can develop my solution for everyone out of an all-encompassing rationality or worse, a decisionist 'calling'. On the contrary, I must rely on a pre-existing shared humanity that cannot be 'rationalized' or commanded, but can only be marveled at and must therefore be protected and preserved wherever possible. This assumption of the pre-rational existence of shared values corresponds to the observation that in Model A it is not reason but the freedom to be creative and to love that is the central characteristic of human beings; new ways of life are possible without theory. The political power that is really able to intervene in the crisis in a sustainable way, both ecologically and socially, can therefore only be organized in democratic discussion.

If it were possible to ward off the great danger of autocratic tendencies in this way, the problem of deeply interwoven power structures, codified in our world maps, which must be overcome in the effort to achieve justice especially also in Western countries, would remain. This then leads us to Marx, Foucault, du Bois, Fanon, Beauvoir, Butler, Anna Julia Cooper etc. The question of coming 'open societies' will increasingly have to be how fair(er) institutions can be built and/or strengthened. According to Model A, this will also require the at least pragmatically conceived assumption of the objective existence of a coordinate system of fundamental values as regulative ideas. [1]

11.10 The central role of aesthetics

In Model A, aesthetics plays a central role: In it, the triad of metaphysics – ethics – aesthetics is inverted; with the positing of qualities – free subjects

[1] I would accordingly expect for the not too distant future to see, for instance in sociological discourses, an increasing move away from relativistic positions, which in the dual crisis of overuse and underdistribution run the risk of providing authoritarian positions with arguments for the populist polarization of our societies. (On the basis of the denial of a common coordinate system in favour of the assumption that individual contexts are not simply framework conditions for our being-in-the-world, but rather construct the realities of this being-in-the-world in the first place).

– the world, aesthetics appears as the 'first philosophy'. Beauty is then the perception of multiplicity in unity; when the individual is realized as an individual in the whole. (A somewhat related idea of beauty as diversity can be found in Alain Locke, for example). Whether this definition can be argued to hold up in reality requires further investigation. However, if we are willing to grant it to be true for our considerations here, then in the outlined model it is not truth, but also not really the good, but rather beauty that allows for the diversity of life plans. (And it then makes sense to assume with Plato and Plotinus that we involuntarily and most deeply strive for the beautiful and not the true or good). The 'aesthetic packing' plays an important role in the 'invention' of these plans, but does not lead to static results, as they can always be re-bundled. We need narratives of the true, beautiful, and good life before we can tackle this. Art and culture therefore also have an important ethical and subsequently political and economic function. According to Model A, people should be moved already by the breathtaking beauty of the overall construction to strive not only for individual, but also social, and above all not only material growth in their life plans. With Meadow's theory of twelve 'leverage points', [239] it can be assumed that such an aesthetic reorientation of the life plans of future generations would have the greatest potential to solve our current problems.

In this sense, all life is art; however, it makes sense to only call something art in an academic context if it is in some way ahead of 'normality'. Rarely will it then be seen as avant-garde in retrospect; most of the time it will simply fail and be forgotten. It also makes sense to limit the concept of art in distinction to science and social activism to that avant-garde which intensifies expression, as Mutschler describes it as a central characteristic of art. [240] The aesthetic experience then reveals the contribution of the individual as such to the whole, which in the case of professional art is often precisely the development of the new of the avant-garde. What is interesting about Model A is that works of art can really be seen in the moment not only as physical givens, but including their objectively real 'aura' of intersubjective meanings for the informed viewer. This could explain the special role of specific artifacts; they are assigned mental realities that are absent from copies. However, it is the ongoing reception that keeps this aura alive as intersubjective meanings distributed across subjects: We receive what we use and become part of what we see, hear, read, or perform. This could then explain at least part of the great identification we experience with 'our' stars and their works.

But there is another sense in which aesthetics plays a fundamental role in Model A: In the operation of linking quantitative with qualitative information, which is so central to our perception and thinking, e.g. the assignment of a color to a physical signal, but also in the linking of qualitative information with one another, e.g. in 'aesthetic packing', certain (sets of) building blocks seem to be particularly suitable for this use due to their quality(ies). The aesthetic relations between different colors, for example, seem to be well suited for the representation of light signals, but different qualities of pain do not. For physical properties we accept such relationships without further ado, but we also count 1, 2, 3 and it is hard to imagine that a different evolution could have led us to count 1, red, round instead. Some building blocks may be interchangeable, such as visual and auditory qualities; think of echolocation. Some may not be perfectly interchangeable, but could still be used evolutionarily in the same functional contexts. In addition, there may be more qualities that are completely alien to us, such as those involved in the perception of electric fields, which is possible for some animal species. In all these cases, the building blocks stand in as broad, stable markers for noisy, complex signals. Some signals, however, do not need markers; we do not taste fat, although we have corresponding 'sensors'; and sometimes the marker can be set by deceiving the physical system; think of sweeteners, for example.

But is there not, perhaps at least beyond these cases, maybe at least at a higher level, a 'logic' of assigning or packing qualities? On the one hand, the 'grammar of meaning formation' for the structuring of our world maps, outlined in chapter 8 and motivated by Husserl's phenomenology, could be considered here. On the other hand the 'semantic logic' of relations between mathematical entities, also considered in chapter 8, could be a candidate theory, if we assume it to be extendable to all underlying building blocks. Further above, the ideas of growth and identity also developed such a 'necessity' of their realization that can be observed in extended evolution. Does not the assumed ability of the core subjects, to intuitively grasp qualities, already require a certain logic behind them?

At least for mathematical entities, quite a few will be inclined to assume logically necessary relations, *against* which, however, I argued in chapter 8. In contrast, if we were prepared to take this step, it would seem inconsistent to assume that similar relations do not exist between other qualities, too. Then Hegel's system suddenly appears as an attractive alternative again: His dialectic can be understood as such a semantic logic, which then unfolds very far-reaching consequences in his work. Hegel's central element of the synthesis of thesis and anti-thesis appears argumentatively difficult to maintain, and is in any case not clearly formalizable. (If only because, depending on the context, many anti-theses can always be thought of; thus the opposite of kicking can be non-kicking, as well as hitting.)

But apart from logical-mathematical relations, precisely the lack of such strictly valid relationships seems to be a characteristic of qualities. And logical-mathematical relations can be developed, as in chapter 8, as initially not necessary by tracing them back to the above 'grammar of meaning formation' of our world maps as only 'practically necessary' in our world. (Which can even take the horror out of logical-mathematical paradoxes.)

In my opinion, with Model A we would therefore have to argue for a 'semantic aesthetics', that is indeed the existence of connections between meanings; but not for a 'semantic logic', which would additionally imply these connections to be necessary. This would also once again emphasize the creative freedom of the subjects and subsequently the commitment to democratic discourse in A-world. Thus, the assumption of the possibility of an intuitive 'shimmy' from idea to idea, when ideas can be used as a transition from a given idea to a new one, or when a given context can suggest achievable new ideas, would imply an aesthetic rather than a logic between these ideas, as does, ultimately, the notion of the possible 'breadth' of ideas in terms of content. Then, however, it could also be assumed that this breadth need not be uniform, so that perhaps for some, for instance logical-mathematical ideas, the fuzzy aesthetics of the connections between them could merge into a sharp logic of relations.

As with the ethics of growth outlined above, one would again have to assume that this 'semantic aesthetic' develops in parts with the whole. Logic would be a high art, but art, with Nelson Goodman, would be a form of search.

11.11 Artificial intelligence as part of the solution

To return to the beginning of the book, it should not go unmentioned that despite the weaknesses of our current concepts of human and artificial intelligence pointed out in the previous chapters, the latter naturally also has great potential to contribute to solving our problems. [34] Many everyday contexts do not require any specific human intelligence and with AI systems, we will have a practically infinite number of employees with a normal level of competence for many contexts at our disposal in the future. Last but not least, we will adapt where there is no other way to achieve further automation gains; think of the 'education' in correct use that modern deposit machines provide us with. It is therefore undoubtedly the case that humanity has a new, quite formidable tool at its disposal in the form of modern AI systems. What it does with it will hopefully not be decided on the basis of purely materialistic considerations alone.

Chapter 12

Outlook

As explained in chapter 5, Tse distinguishes between epistemological, for instance conceptual, and ontological idealism models and further divides the latter into subjective and objective approaches. For our question of how to bridge the gap between quantitative information and meaning as qualitative information, and then formulate an alternative understanding of human thought, only ontological, objective idealisms appear to be helpful, since in the other cases the concept of quantitative information does not appear to be easily recoverable.

In analogy to Meixner's classification of panpsychistic models, also mentioned in chapter 5, one could then further differentiate ontological, objective idealism models into holistic and 'atomistic' ones, the latter assuming reality to be fundamentally composed of distinct parts. The well-known classical idealism theories would have to be assigned to the holistic camp. Here, on the one hand, the decombination problem, of how the individual subject can be understood as only part of the totality, must be solved; and on the other hand, the emanation problem, of how at least the observation of a material world can be understood from the totality, is exacerbated by the fact that the perceived existence of individual objects must always first be argued for. Perhaps more importantly, such theories cannot simply be derived directly from our basic intuitions, as subjects perceiving individual objects, and are difficult to make practically fruitful for discussions about the perceived interfaces between the material and non-material world; we do not get easily from Hegel to modern neurobiology.

In this book, an attempt has been made to sketch a first model for an *atomistic*, objective, ontological idealism. Starting from our most fundamental intuitions, namely the existence of subjects and qualities, it was shown that the structuring of the material world is conceivable in such a model already on the micro-scale, i.e. in line with the findings of natural science. If no God-like world spirits or non-material building blocks with dispositional properties are to be used, then at least the assumption of the evolutionary development of a population of initially and later still mostly very simple subjects is necessary. The advantage of such an atomistic idealism is that now 'only' the emanation problem remains to be solved, where we can additionally profit from a large amount of existing research on the re-interpretation of scientific theories, e.g. in the context of string theory research. The suggested model is a Platonic, but in a second step also a scientific realism and a naturalism in the sense that it is informed by the natural sciences in the form of an inductive metaphysics.

The model thus obtained is an objective idealism, not in a rationalistic maximum form, but in a pragmatic minimum form; without eternal truths, but dependent on the continued philosophical-scientific and also philosophical-social dialog. This not only enables a meaningful, non-devaluating connection to value theories and discourses in the humanities and social sciences; Model A is not a hostile takeover of the humanities by the natural sciences, nor vice versa; but circumvents a number of problems of alternative, idealistic, panpsychistic and dualistic models.

The following is avoided: 1) The assumption that the framing of the world by our mind corresponds to the complete creation of this world, or even proves its non-existence. 2) The assumption that the mental comprehensibility of the world allows the derivation of unquestionable knowledge ('eternal truths') about this world. 3) The assumption that mesoscale objects are structured directly via (Platonic) ideas or (Aristotelian) forms. 4) The abstraction problem of materialism 'by definition', with the positing of non-material building blocks. 5) The combination problem of panpsychism in an analogous way, with the positing of subjects as the focus of separate mental contents. 6) The interaction problem of dualism, with the assumption that whatever plays the role of the material world is that world. At the interface, no laws of nature but evolutionarily acquired rules executed by micro-subjects apply to physical properties differently than to nonphysical ones. Here we are not extrapolating from physics to human biology, but vice versa; everywhere (micro-)subjects now manipulate abstract entities – although the existence of these (micro-)subjects will admittedly remain as unprovable as the existence of natural laws as such.

If one were to allow physical properties with dispositional forces in the model, one could derive a dualistic variant without micro-subjects. If one were to allow non-physical properties with dispositional forces in the model, one could derive a panpsychistic variant in the etablished sense. In the latter case in particular, however, the necessary causal links seem too inflexible to me; in any case, one returns to the interaction or combination problem.

The proposed model could offer interesting solutions to a number of problems at and near the mind/matter boundary: Proposals for the interpretation of quantum mechanics, the problem of molecular symmetry, the neuronal code and the binding problem in neuroscience, mental causation, a more holistic understanding of mental processes, etc. were considered. The extent to which the model threatens to promise far too much was discussed critically in Chapter 8. Here it must be questioned again and again whether the positing of subjects and qualities can really be regarded as absolutely necessary, or whether this is already one step too many, which would then deny the resulting model critical and innovative potential rather than opening it up, e.g. in the area of mind/brain interactions.

The most important aspect of the proposed idealism is the direct connection to the pragmatically conceived natural sciences, for which the stable individuation of objects from universals was postulated as a unifying core principle. Space was interpreted functionally; a connection to an existing interpretation of quantum theory was attempted. With regard to neuroscience, a possible connection has also been shown, which must then postulate a coupled mind/brain development and our subconscious as a 'psychobiome'. As a first attempt, a possible mind/brain interface was proposed, with the conclusion that certain mental activities should manage without neuronal correlates, and that some neuronal correlates should have less information content than necessary for their functioning.

As explained in chapter 8, a person is then a totality of body and mind. The mind is the totality of the core subject and the structured bundle of universals that the core subject can perceive and manipulate. The structured bundle is the representation of the person's world, their (not only spatially understood) 'world map', through which the core subject interacts with the world. Parts of the world map of the core subject are also part of the world map of sub-subjects that make up the person's subconsciousness and allow for the interaction between the core subject and the body. Parts of the sub-subject's world maps are in turn also part of bundles with physical properties that belong to the structure of the person's brain, whereby the respective sub-subjects can manipulate both the mental and physical properties of their bundles on an equal footing, albeit according to different, evolutionarily learned rules. The brain as part of the person's body thus functions as an anchor of the non-material mind of the core subject. However, the person's body is only to be understood as a hybrid entity: On the one hand, it receives its embedding in the causal world via the countless bundles of physical properties of which it consists; organs, molecular structures, and ultimately elementary particles. On the other hand, it receives its unity as an entity only in the connection of these structures to structured bundles of universals in the mind of the core subject, which is only aware of its body in this form.

With regard to our core question of how an alternative model of human thinking could be formulated, some conclusions can then be drawn: In opposition to the current consensus of understanding human thought as 'purely' quantitative information processing, the idealism outlined above allows both to establish a meaningful connection between quantitative information and meaning as qualitative information, and to formulate a more complex model of human thought. The positing of independently existing non-material building blocks or abstract entities makes it possible to understand quantitative information as a change in such qualities, but qualitative information as these qualities themselves.

Human thinking is then characterized not only by the formation of dynamic attractors of neuronal activity in the brain, but also by the evolutionarily learned but essentially creative combination of quantitative and qualitative information, where for instance the neuronal signal for red is 'radically' translated by a subject in our subconsciousness into the quality red for us. The initial evolutionary benefit was presumably the availability of 'superphysically' stable and broad attractors like colors or shapes, as the target of physical feedback processes, and in a second step the availability of stable and broad abstractions. It must then follow that some mental processes should be possible 'super-physically' quickly, e.g. pattern recognition via universals, but possibly also language comprehension. Human thinking would then have to be understood on a spectrum between purely quantitative information processing in the form of physical processes and qualitative information processing in the form of the structured un/bundling of (bundles of) universals. Thought processes guided by formal criteria would correspond to the manipulation of qualities on the basis of regularities that correspond to those in the material-physical world. The freer creativity of subjects with richly structured world maps would additionally allow for much 'wilder' operations as we observe them in music, art and literature, amongst other things. Like for calculating machines, the purpose of brains would be to provide reliably available, appropriately logically structured causal relationships between qualities.

With such a view of human thinking, one can argue against any 'goal post shifting' of a (re)definition of human intelligence in the terms of our current AI models. Central to natural intelligence would not be machine-like cognitive performance, but the intentional perception of qualitative information as abstract entities, as well as the free, ultimately creative linking of patterns of quantitative information with such qualities. The practical, if not already theoretical failure of our current neuroscientific and technical models, which can certainly be observed, would then already be inherent in the fundamental unavailability of stable and broad abstractions. A (bio)technological development of artificial general intelligence would nevertheless be conceivable in Model A, if the coupled mind/brain development could be understood in all detail. Here, however, the difference between machine and organism would be completely blurred.

Model A relies on empiricism and falsification in the sense of Locke and Popper, but claims that this requires an at least pragmatic commitment to the objective existence of our basic building blocks of perception, as well as a commitment to the objective existence of a coordinate system of basic values as regulative ideas. It can be used to argue in favor of open societies and strong institutions in Popper's sense, which, however, would in turn require a commitment to the coordinate system of our fundamental values.

In the end, the circle closes, but the question remains; what is the human being and here more precisely; human intelligence, whereby one might add; in times of AI. This book has attempted to answer this question by developing an updated (panpsychist, objective) 'scientific idealism' for the modern world. Whether this has been successful will, also according to Model A, be decided by current and, above all, perhaps future readers. If we take Kuhn's theory of paradigm shifts [241] seriously, worldviews do not change due to logically compelling evidence alone, but because some worldviews prove to be more helpful for our progressive growth than others. The possible adoption of new world views is ultimately in the hands of future generations. It should hopefully have become clear that this book is not intended to present a solution, but to formulate a research project still to be carried out.

In the end, all that remains is the hope that, if not the model, then at least some considerations on the construction of models of this kind might have a certain value: That we probably cannot avoid qualities and subjects as basic building blocks. That a broadly understood concept of evolution is our best chance of explaining the growth that has taken place. That the material world is not simply given, but performs very basic functions for subjects that manipulate universals. That the mental world offers us 'superphysically' stable, conceptually broad attractors for physical information processes, as well as stable and broad abstractions. And that idealistic models can be used to describe a continuous development of physical, biological and cultural processes. Some possibly experimentally accessible 'predictions' of the model have been made throughout the text; again, it remains to be seen what the future holds.

Acknowledgments

My thanks go to the supervisors of this work, Christian Illies and Uwe Voigt, the participants of the Bamberg seminars ’Nature and Mind’ and ’Philosophy of History and Ecological Crisis’, the Bamberg and Augsburg advanced seminars, the working group ’Artificial Intelligence and Philosophy’ of the InterKI teaching program at the University of Münster – here in particular Benedikt Lennartz, the participants of the Seventh Annual Workshop on Biological Mentality – here in particular Kenneth Augustyn and Johann Summhammer, as well as Olimpia Lombardi and Christian Thiede.

Glossary

Abstract entities – here: Non-material (bundles of) building blocks of qualitative and universal nature. Without further qualification, initially equivalent in meaning to **qualitative properties**, **qualities**, **universals**, **ideas** or even **forms**. This includes colors, numbers, terms, concepts, affective qualities, values, etc.

Agents – here: All subjects, be they micro-subjects, sub-subjects or core subjects.

Emanation problem – here: Problem of ontological, idealistic approaches, which must explain in detail how exactly the observations and conclusions of the modern natural sciences could result from the idealistic basic assumptions.

Brain – here: Material 'anchor' for a subconsciousness and then also the world map of the core subject. Consists of bundles of physical properties, ultimately elementary particles.

Mind – here: The totality of 'world map' and core subject.

Idealism – here: Approach that assumes that the material world is ultimately completely traceable to the non-material world. This need not contradict physical assumptions if it can be shown that the physical world can be identified with that which performs its function, but is essentially composed of non-material elements.

Body – here: Hybrid entity, which is integrated into the causal network of the material world via the countless bundles of physical properties of which it consists (organs, molecular structures, but ultimately elementary particles), but which only receives its unity in the connection of these structures to structured bundles of universals in the mind of its core subject.

Materialism – here: Approach that assumes that the non-material world is completely traceable to the material world. To be understood as an epistemic instrument; hardly any real physicalism will see itself as materialism according to this interpretation.

Material/physical causality – here: Effective outcome of the agency of micro-subjects, which follow consistency rules for the manipulation of material/physical properties.

Human thinking – here: Spectrum from purely quantitative information processing in the form of physical processes to purely qualitative information processing in the form of the structured (un)bundling of (bundles of) universals.

Naturalism – here: Approach that is (not necessarily only) informed by the natural sciences and can be formulated consistently with them. The term is thus understood epistemically here.

Person – here: The totality of body and mind.

Physical world – here: Sum of the bundles of properties that are manipulated by micro-subjects according to evolutionarily acquired consistency rules. In contrast, there is the non-physical world, whose properties or property bundles can be manipulated without strict consistency requirements.

Subject – here: A consciousness that is able to perceive universals and manipulate them through (un)bundling.

Subject, core- – here: A subject that has central control over the world map of an organism. In the case of humans: the 'I'.

Subject, micro- – here: A subject that, as a 'cellular automaton' or 'physical microbe', manipulates physical properties on the (sub-)micro level according to evolutionarily learned consistency rules. The consistent development of the physical world follows from the actions of these micro-subjekts, as a result of which we observe the operation of natural laws.

Subject, meso- – here: A subject that is perceived as an individual on the meso scale of our everyday experience, e.g. a human being. From the perspective of Model A, meso-subjects must be understood as holobionts of micro-, sub- and core-subjects.

Subject, sub- – here: A subject that, together with other sub-subjects, acts as a subconsciousness that allows a higher being the interaction between non-material mind and material brain.

Subconsciousness – here: **'Psychobiome'** of sub-subjects that share parts of the world map of the core subject and (possibly via several intermediate layers) can also perceive and manipulate physical properties of the brain.

World map – here: Structured bundle of universals that the core subject can perceive and manipulate and that functions as a representation of the world for a core subject.

List of Figures

Bibliography

[1] Y. Bar-Hillel and R. Carnap, "Semantic information," *British Journal for the Philosophy of Science*, vol. 4, no. 14, pp. 147–157, 1953.

[2] L. Floridi, "What is the philosophy of information?," *Metaphilosophy*, vol. 33, no. 1-2, pp. 123–145, 2002.

[3] C. E. Shannon, "A mathematical theory of communication," *The Bell System Technical Journal*, vol. 27, pp. 379–423, 1948.

[4] A. Turing, "On computable numbers, with an application to the entscheidungsproblem," *Proceedings of the London Mathematical Society*, vol. 42, no. 1, pp. 230–265, 1936.

[5] G. Frege, *Begriffsschrift: Eine der arithmetischen nachgebildete Formelsprache des reinen Denkens.* 1879.

[6] S. Harnad, "The symbol grounding problem," *Physica D*, vol. 42, pp. 335–346, 1990.

[7] J. Searle, "Minds, brains, and programs," in *Philosophy of Mind: A Guide and Anthology* (J. Heil, ed.), Oxford: Oxford University Press, 1980.

[8] G. W. Leibniz, *Monadology.* 1714.

[9] M. Taddeo and L. Floridi, "Solving the symbol grounding problem: a critical review of fifteen years of research," *Journal of Experimental & Theoretical Artificial Intelligence*, vol. 17, no. 4, pp. 419–445, 2005.

[10] L. Floridi, *The Philosophy of Information.* New York: Oxford University Press US, 2011.

[11] L. Floridi, *The Ethics of Information.* Oxford: Oxford University Press, 2013.

[12] L. Floridi, *The Logic of Information: A Theory of Philosophy as Conceptual Design.* Oxford: Oxford University Press, 2019.

[13] C. Penco and M. Benzi, "Review of: Luciano floridi, the logic of information: A theory of philosophy as conceptual design," *Notre Dame Philosophical Reviews*, 2019.

[14] S. Lapuschkin, S. Wäldchen, A. Binder, and et al., "Unmasking clever hans predictors and assessing what machines really learn," *Nat Commun*, p. 1096, 2019.

[15] A. N. Whitehead and B. Russell, *Principia Mathematica to *56.* New York: Cambridge University Press US, 1962.

[16] H. L. Dreyfus, *What Computers Still Can't Do: A Critique of Artificial Reason.* Boston: MIT Press, 1992.

[17] M. Polanyi and A. Sen, *The Tacit Dimension.* Chicago: University of Chicago, 1966.

[18] W. S. McCulloch and W. Pitts, "A logical calculus of the ideas immanent in nervous activity," *The Bulletin of Mathematical Biophysics*, vol. 5, no. 4, pp. 115–133, 1943.

[19] A. Newell, J. C. Shaw, and H. A. Simon, "Elements of a theory of human problem solving," *Psychological Review*, vol. 65, no. 3, pp. 151–166, 1958.

[20] J. A. Fodor, *The Language of Thought.* Boston: Harvard University Press, 1975.

[21] H. Putnam, "Psychological predicates," in *Art, Mind, and Religion* (W. H. Capitan and D. D. Merrill, eds.), pp. 37–48, Pittsburgh: University of Pittsburgh Press, 1967.

[22] F. Jackson, "Epiphenomenal qualia," *Philosophical Quarterly*, vol. 32, no. April, pp. 127–136, 1982.

[23] F. Jackson, "What mary didn't know," *Journal of Philosophy*, vol. 83, no. 5, pp. 291–295, 1986.

[24] P. Ludlow, Y. Nagasawa, and D. Stoljar, eds., *There's Something About Mary: Essays on Phenomenal Consciousness and Frank Jackson's Knowledge Argument.* Boston: MIT Press, 2004.

[25] T. Nagel, "What is it like to be a bat?," *Philosophical Review*, vol. 83, no. October, pp. 435–50, 1974.

[26] D. J. Chalmers, *The Conscious Mind: In Search of a Fundamental Theory (2nd edition)*. Oxford: Oxford University Press, 1996.

[27] D. J. Chalmers, *The Character of Consciousness*. New York: Oxford University Press US, 2010.

[28] J. Shear, ed., *Explaining Consciousness: The Hard Problem*. New York: MIT press, 1997.

[29] T. Horgan, M. Sabates, and D. Sosa, eds., *Qualia and Mental Causation in a Physical World: Themes From the Philosophy of Jaegwon Kim*. Cambridge, UK: Cambridge University Press, 2015.

[30] A. Pautz and D. Stoljar, eds., *Blockheads! Essays on Ned Block's Philosophy of Mind and Consciousness*. New York: MIT Press, 2019.

[31] D. Rumelhart, G. Hinton, and R. Williams, "Learning representations by back-propagating errors," *Nature*, p. 533–536, 1986.

[32] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, pp. 1735–1780, 1997.

[33] K. Hao, "AI pioneer Geoff Hinton: 'Deep learning is going to be able to do everything'," *MIT Technology Review*, 2020.

[34] R. Vinuesa, H. Azizpour, I. Leite, and et al., "The role of artificial intelligence in achieving the sustainable development goals," *Nat Commun*, 2020.

[35] C. B. Frey and M. A. Osborne, "The future of employment: How susceptible are jobs to computerisation?," *Technological Forecasting and Social Change*, vol. 114, pp. 254–280, 2017.

[36] J. Zeng, "Artificial intelligence and China's authoritarian governance," *International Affairs*, vol. 96, pp. 1441–1459, 10 2020.

[37] K. Crawford, *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. New Haven: Yale University Press, 2021.

[38] M. H. Solon Barocas and A. Narayanan, *Fairness and Machine Learning, Limitations and Opportunities*. New York: The MIT Press, 2023.

[39] P. N. Stuart Russell, *Artificial Intelligence, A Modern Approach.* London: Pearson, 2021.

[40] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, p. 436–444, 2015.

[41] J. Sevilla, L. Heim, A. Ho, T. Besiroglu, M. Hobbhahn, and P. Villalobos, "Compute trends across three eras of machine learning," in *2022 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–8, 2022.

[42] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," 2023. arxiv:1706.03762.

[43] P. Christiano, J. Leike, T. B. Brown, M. Martic, S. Legg, and D. Amodei, "Deep reinforcement learning from human preferences," 2023. arxiv:1706.03741.

[44] S. Bubeck, V. Chandrasekaran, R. Eldan, J. Gehrke, E. Horvitz, E. Kamar, P. Lee, Y. T. Lee, Y. Li, S. Lundberg, H. Nori, H. Palangi, M. T. Ribeiro, and Y. Zhang, "Sparks of artificial general intelligence: Early experiments with gpt-4," 2023. arxiv:2303.12712.

[45] A. Zou, L. Phan, S. Chen, J. Campbell, P. Guo, R. Ren, A. Pan, X. Yin, M. Mazeika, A.-K. Dombrowski, S. Goel, N. Li, M. J. Byun, Z. Wang, A. Mallen, S. Basart, S. Koyejo, D. Song, M. Fredrikson, J. Z. Kolter, and D. Hendrycks, "Representation engineering: A top-down approach to ai transparency," 2023. arxiv:2310.01405.

[46] E. M. Bender, T. Gebru, A. McMillan-Major, and S. Shmitchell, "On the dangers of stochastic parrots: Can language models be too big?," FAccT '21, (New York, NY, USA), p. 610–623, Association for Computing Machinery, 2021.

[47] W. Saba, "Machine learning won't solve natural language understanding," *The Gradient*, 2021.

[48] D. T. Langendoen and P. M. Postal, "The vastness of natural languages," *Linguistics and Philosophy*, vol. 9, no. 2, pp. 225–243, 1986.

[49] A. Branco, "Computational complexity of natural languages: A reasoned overview," in *Proceedings of the Workshop on Linguistic Complexity and Natural Language Processing* (L. Becerra-Bonache, M. D.

Jiménez-López, C. Martín-Vide, and A. Torrens-Urrutia, eds.), (Santa Fe, New-Mexico), pp. 10–19, Association for Computational Linguistics, Aug. 2018.

[50] G. Marcus and E. Davis, *Rebooting Ai: Building Artificial Intelligence We Can Trust*. New York: Vintage, 2019.

[51] E. M. Bender and A. Koller, "Climbing towards NLU: On meaning, form, and understanding in the age of data," in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (D. Jurafsky, J. Chai, N. Schluter, and J. Tetreault, eds.), (Online), pp. 5185–5198, Association for Computational Linguistics, July 2020.

[52] K. Mahowald, A. A. Ivanova, I. A. Blank, N. Kanwisher, J. B. Tenenbaum, and E. Fedorenko, "Dissociating language and thought in large language models: a cognitive perspective," 2023. arXiv:2301.06627.

[53] A. Saxe, S. Nelli, and C. Summerfield, "If deep learning is the answer, what is the question?," *Nat Rev Neurosci*, p. 55–67, 2021.

[54] L. Zaadnoordijk, T. Besold, and R. Cusack, "Lessons from infant learning for unsupervised machine learning," *Nat Mach Intell*, p. 510–520, 2022.

[55] H. Maurer, *Cognitive Science: Integrative Synchronization Mechanisms in Cognitive Neuroarchitectures of Modern Connectionism (1st ed.)*. Boca Raton: CRC Press, 2021.

[56] J. L. M. David E. Rumelhart, *Parallel Distributed Processing, Volume 1: Explorations in the Microstructure of Cognition: Foundations*. New York: The MIT Press, 1986.

[57] R. Adolphs, "The unsolved problems of neuroscience," *Trends Cogn Sci*, pp. 173–5, 2015.

[58] J. Pearl, *The Book of Why: The New Science of Cause and Effect*. New York: Basic Books, 2018.

[59] "The new NeuroAI," *Nat Mach Intell*, p. 245, 2024.

[60] P. Hitzler, A. Eberhart, M. Ebrahimi, M. K. Sarker, and L. Zhou, "Neuro-symbolic approaches in artificial intelligence," *National Science Review*, vol. 9, p. nwac035, 03 2022.

[61] A. Garcez and L. Lamb, “Neurosymbolic ai: the 3rd wave,” *Artif Intell Rev*, p. 12387–12406, 2023.

[62] G. Hinton, “The forward-forward algorithm: Some preliminary investigations,” 2022. Preprint at ArXiv: https://doi.org/10.48550/arXiv.2212.13345.

[63] D. Kahneman, *Thinking, Fast and Slow.* New York: Farrar, Straus and Giroux, 2011.

[64] T. Fuchs, “Delusion, reality and intersubjectivity: A phenomenological and enactive analysis,” *Phenomenology and Mind*, pp. 120–143, 2020.

[65] D. J. Chalmers, “A computational foundation for the study of cognition,” *Journal of Cognitive Science*, vol. 12, no. 4, pp. 323–357, 2011.

[66] D. J. Chalmers, “The computational and the representational language-of-thought hypotheses,” *Behavioral and Brain Sciences*, vol. 46, p. e269, 2023.

[67] N. Goodman and W. van Orman Quine, “Steps toward a constructive nominalism,” *Journal of Symbolic Logic*, vol. 12, no. 4, pp. 105–122, 1947.

[68] W. V. O. Quine, “On what there is,” *Review of Metaphysics*, vol. 2, no. 5, pp. 21–38, 1948.

[69] H. H. Field, *Science Without Numbers: A Defence of Nominalism.* Princeton: Princeton University Press, 1980.

[70] P. Benacerraf, “Mathematical truth,” *Journal of Philosophy*, vol. 70, no. 19, pp. 661–679, 1973.

[71] G. Frege, “Der Gedanke: Eine Logische Untersuchung,” *Beiträge Zur Philosophie Des Deutschen Idealismus*, p. 58–77, 1918.

[72] K. Popper, “Epistemology without a knowing subject,” in *Logic, Methodology, and Philosophy of Science III*, p. 333–373, Amsterdam: North Holland, 1968.

[73] S. Cowling, *Abstract Entities.* New York: Routledge US, 2017.

[74] W. Künne, *Abstrakte Gegenstände, Semantik und Ontologie.* Frankfurt am Main: Klostermann, 2007.

[75] K. Gödel, "What is cantor's continuum problem?," in *Philosophy of Mathematics: Selected Readings, 2nd edition*, p. 254–270, Cambridge: Cambridge University Press, 1964.

[76] F. Bradley, *Appearance and Reality.* London: Swan Sonnenschein, 1897.

[77] J. R. Searle, *The Rediscovery of the Mind.* New York: MIT Press, 1992.

[78] R. Penrose, *The Emperor's New Mind.* Oxford: Oxford University Press, 1989.

[79] P. K. Unger, *Ignorance: A Case for Scepticism.* Oxford: Oxford University Press, 1975.

[80] H. Chang, *Realism for Realistic People: A New Pragmatist Philosophy of Science.* Cambridge, UK: Cambridge University Press, 2022.

[81] N. Cartwright, J. Hardie, E. Montuschi, M. Soleiman, and A. C. Thresher, *The Tangle of Science: Reliability Beyond Method, Rigour, and Objectivity.* Oxford: Oxford University Press, 2022.

[82] U. Meixner, "Idealism and panpsychism," in *Panpsychism: Contemporary Perspectives* (G. Bruntrup and L. Jaskolla, eds.), Philosophy of Mind Series, New York: Oxford Academic, 2016.

[83] D. J. Chalmers, "Idealism and the mind-body problem," in *The Routledge Handbook of Panpsychism* (W. Seager, ed.), pp. 353–373, London: Routledge, 2019.

[84] M. Korth, "Two comments on Chalmer's classification of idealism," 2022. https://philpapers.org/rec/KORTCO-18.

[85] D. Skrbina, *Panpsychism in the West.* Boston: MIT Press, 2005.

[86] W. Seager, "Idealism, panpsychism, and emergentism: The radical wing of consciousness studies," in *The Routledge Handbook of Consciousness*, London: Routledge, 2018.

[87] G. Brüntrup and L. Jaskolla, eds., *Panpsychism: Contemporary Perspectives.* Philosophy of Mind Series, New York: Oxford Academic, 2016.

[88] U. Meixner, *The Two Sides of Being: A Reassessment of Psycho-Physical Dualism.* Paderborn: Mentis, 2004.

[89] D. J. Chalmers and K. J. McQueen, “Consciousness and the collapse of the wave function,” in *Consciousness and Quantum Mechanics* (S. Gao, ed.), Oxford: Oxford University Press, forthcoming.

[90] P. Tse, “Metaphysical idealism revisited,” *Philosophy Compass*, vol. 17, no. 7, p. e12856, 2022.

[91] D. Kodaj, “Humean idealism,” *Australasian Journal of Philosophy*, vol. 101, no. 1, pp. 34–50, 2021.

[92] F. S. M. und Vittorio Hösle, ed., *Idealismus heute: Aktuelle Perspektiven und neue Impulse.* Darmstadt: wbg Academic, 2015.

[93] B. P. G. Joshua Farris, ed., *The Routledge Handbook of Idealism and Immaterialism.* London: Routledge, 2022.

[94] T. Goldschmidt and K. L. Pearce, eds., *Idealism: New Essays in Metaphysics.* Oxford: Oxford University Press, 2017.

[95] P. MacEwen, ed., *Idealist Alternatives to Materialist Philosophies of Science.* Leiden: Brill, 2020.

[96] P. Goff, *Consciousness and Fundamental Reality.* New York: Oxford University Press US, 2017.

[97] J. Foster, *The Case for Idealism.* Boston: Routledge US, 1982.

[98] T. Sprigge, “The vindication of absolute idealism,” *Philosophy*, vol. 60, no. 234, pp. 546–548, 1983.

[99] J. Foster, *A World for Us: The Case for Phenomenalistic Idealism.* Oxford: Oxford University Press, 2008.

[100] B. Kastrup, *The Idea of the World: A Multi-Disciplinary Argument for the Mental Nature of Reality.* Winchester: Iff Books, 2019.

[101] P. Lewtas, “The impossibility of emergent conscious causal powers,” *Australasian Journal of Philosophy*, vol. 95, no. 3, pp. 475–487, 2017.

[102] L.-C. Chan and A. J. Latham, “The possibility of emergent conscious causal powers,” *Australasian Journal of Philosophy*, vol. 100, no. 1, pp. 195–201, 2022.

[103] E. J. Lowe, *Personal Agency: The Metaphysics of Mind and Action.* New York: Oxford University Press US, 2008.

[104] S. C. Gibb, “Mental causation,” *Analysis*, vol. 74, p. 327–338, 2014.

[105] L. Smolin, *Time Reborn: From the Crisis in Physics to the Future of the Universe.* Boston: Houghton Mifflin Harcourt, 2013.

[106] N. C. Martens, “The metaphysics of emergent spacetime theories,” *Philosophy Compass*, vol. 14, no. 7, p. e12596, 2019. e12596 10.1111/phc3.12596.

[107] D. J. Chalmers, ed., *Constructing the World.* Oxford: Oxford University Press, 2012.

[108] D. J. Chalmers, “Finding space in a nonspatial world,” in *Philosophy Beyond Spacetime* (C. Wüthrich, B. L. Bihan, and N. Huggett, eds.), Oxford: Oxford University Press, 2021.

[109] P. Goff, “Did the universe design itself?,” *International Journal for Philosophy of Religion*, vol. 85, no. 1, pp. 99–122, 2019.

[110] P. F. Strawson, *Individuals.* London: Routledge, 1959.

[111] J. Hawthorne and T. Sider, “Locations,” *Philosophical Topics*, vol. 30, no. 1, pp. 53–76, 2002.

[112] D. M. Armstrong, “Universals and scientific realism volume 1: Nominalism and realism; volume 2: A theory of universals,” *Noûs*, vol. 16, no. 1, pp. 133–142, 1982.

[113] D. K. Lewis, “New work for a theory of universals,” *Australasian Journal of Philosophy*, vol. 61, no. 4, pp. 343–377, 1983.

[114] F. MacBride, *On the Genealogy of Universals: The Metaphysical Origins of Analytic Philosophy.* Oxford: Oxford University Press, 2018.

[115] A. R. J. Fisher, “Structural universals,” *Philosophy Compass*, vol. 13, no. 10, p. e12518, 2018.

[116] S. Wolfram, *A New Kind of Science.* Wolfram Media, Inc., 2002.

[117] N. J. T. James Read, ed., *The Philosophy and Physics of Noether’s Theorems: A Centenary Volume.* Cambridge, UK: Cambridge University Press, 2022.

[118] K. Olive, “Review of particle physics,” *Chinese Physics C*, vol. 38, p. 090001, 2014.

[119] R. D. Klauber, *Student Friendly Quantum Field Theory, Basic Principles and Quantum Electrodynamics.* Fairfield: Sandtrove Press, 2013.

[120] B. Zwiebach, *First Course in String Theory.* Cambridge, UK: Cambridge Unversity Press, 2009.

[121] J. M. Butterworth, "The standard model: how far can it go and how can we tell?," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 374, no. 2075, p. 20150260, 2016.

[122] M. Schlosshauer, J. Kofler, and A. Zeilinger, "A snapshot of foundational attitudes toward quantum mechanics," *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, vol. 44, no. 3, pp. 222–230, 2013.

[123] M. Esfeld, *Naturphilosophie Als Metaphysik der Natur.* Frankfurt am Main: Suhrkamp, 2008.

[124] C. J. Isham, *Lectures on Quantum Theory.* London: Imperial College Press, 2001.

[125] J. S. Bell, "On the problem of hidden variables in quantum mechanics," *Rev. Mod. Phys.*, p. 447–452, 1966.

[126] C. Budroni, A. Cabello, O. Gühne, M. Kleinmann, and J.-A. Larsson, "Kochen-specker contextuality," *Rev. Mod. Phys.*, vol. 94, p. 045007, Dec 2022.

[127] A. Cabello, "Bell non-locality and kochen–specker contextuality: How are they connected?," *Found Phys*, p. 61, 2021.

[128] S. Kochen and E. P. Specker, "The problem of hidden variables in quantum mechanics," *Journal of Mathematics and Mechanics*, vol. 17, pp. 59–87, 1967.

[129] M. Morganti, "Inherent properties and statistics with individual particles in quantum mechanics," *Studies in History and Philosophy of Modern Physics*, vol. 40, no. 3, pp. 223–231, 2009.

[130] B. Spinoza, *Ethica, ordine geometrico demonstrata.* 1677.

[131] T. Thiemann, "Lectures on loop quantum gravity," in *Quantum Gravity. Lecture Notes in Physics*, vol. 631, Berlin: Springer, 2003).

[132] H. P. Lovecraft, *The Colour Out of Space.* 1927.

[133] I. Johansson, “Container space and relational space,” in *Ontological Investigations: An Inquiry Into the Categories of Nature, Man and Soceity*, pp. 1–21, Berlin: De Gruyter, 2004.

[134] P. Busch, *The Time–Energy Uncertainty Relation*, pp. 73–105. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008.

[135] L. Hardy, “Quantum theory from five reasonable axioms,” 2001. arxiv:quant-ph/0101012.

[136] L. Masanes, T. Galley, and M. Müller, “The measurement postulates of quantum mechanics are operationally redundant,” *Nat Commun*, p. 1361, 2019.

[137] A. Cabello, “Quantum correlations from simple assumptions,” *Phys. Rev. A*, vol. 100, p. 032120, 2019.

[138] M. t. Vrugt, “How to distinguish between indistinguishable particles,” *The British Journal for the Philosophy of Science*, 2024. Just accepted.

[139] O. Lombardi and M. Castagnino, “A modal-hamiltonian interpretation of quantum mechanics,” *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, vol. 39, no. 2, pp. 380–443, 2008.

[140] J. S. Ardenghi, O. Lombardi, and M. Narvaja, “Modal interpretations and consecutive measurements,” in *EPSA11 Perspectives and Foundational Problems in Philosophy of Science* (V. Karakostas and D. Dieks, eds.), (Cham), pp. 207–217, Springer International Publishing, 2013.

[141] F. Holik, J. P. Jorge, D. Krause, and O. Lombardi, “Quasi-set theory for a quantum ontology of properties,” 2021.

[142] B. Falkenburg, *Particle Metaphysics, A Critical Account of Subatomic Reality.* Berlin: Springer, 2007.

[143] L. Lang, H. M. Cezar, L. Adamowicz, and T. B. Pedersen, “Quantum definition of molecular structure,” *J. Am. Chem. Soc.*, pp. 0002–7863, 2024.

[144] J. González, S. Fortin, and O. Lombardi, “Why molecular structure cannot be strictly reduced to quantum mechanics,” *Found Chem*, p. 31–45, 2019.

[145] S. Fortin and O. Lombardi, “Is the problem of molecular structure just the quantum measurement problem?,” *Found Chem*, p. 379–395, 2021.

[146] S. Fortin, O. Lombardi, and J. C. Martínez González, “A new application of the modal-hamiltonian interpretation of quantum mechanics: The problem of optical isomerism,” *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, vol. 62, pp. 123–135, 2018.

[147] B. Drossel, *On the Relation Between the Second Law of Thermodynamics and Classical and Quantum Mechanics.* Berlin, Heidelberg: Springer, 2015.

[148] J. B. Hartle, *Gravity: An Introduction to Einstein’s General Relativity.* London: Pearson, 2002.

[149] M. Lopez-Corredoira, “Tests and problems of the standard model in cosmology,” *Found Phys*, p. 711–768, 2017.

[150] S. Fortin, O. Lombardi, and M. Pasqualini, “Relational event-time in quantum mechanics,” *Found Phys*, p. 10, 2022.

[151] H. Reichenbach, *Relativitätstheorie und Erkenntnis Apriori.* Berlin: Springer, 1920.

[152] R. Samaroo, “The principle of equivalence as a criterion of identity,” *Synthese*, vol. 197, no. 8, pp. 3481–3505, 2020.

[153] H. Fritzsch, “Fundamental constants and their time variation,” *Progress in Particle and Nuclear Physics*, vol. 66, no. 2, pp. 193–196, 2011. Particle and Nuclear Astrophysics.

[154] M. Milgrom, “MOND theory,” *Canadian Journal of Physics*, vol. 93, pp. 107–118, Feb. 2015.

[155] J. D. Bekenstein, “Relativistic gravitation theory for the modified newtonian dynamics paradigm,” *Phys. Rev. D*, vol. 70, p. 083509, Oct 2004.

[156] M. Morganti, “Fundamentality in metaphysics and the philosophy of physics. part ii: The philosophy of physics,” *Philosophy Compass*, vol. 15, no. 10, p. e12703, 2020.

[157] P. Sterling and S. Laughlin, *Principles of Neural Design.* Boston: The MIT Press, 2017.

[158] L. F. Abbott and P. Dayan, *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems.* Boston: The MIT Press, 2001.

[159] H. Moravec, *Mind Children, The Future of Robot and Human Intelligence.* Boston: Harvard University Press, 1988.

[160] M. Howard, J. Wallman, V. Veitch, and et al., "Contextuality supplies the 'magic' for quantum computation," *Nature*, p. 351–355, 2014.

[161] D. Aerts and S. Aerts, "Applications of quantum statistics in psychological studies of decision processes.," *Found Sci*, p. 85–97, 1995.

[162] D. Aerts, "Quantum structure in cognition," *Journal of Mathematical Psychology*, vol. 53, no. 5, pp. 314–348, 2009. Special Issue: Quantum Cognition.

[163] N. Goodman, *Languages of Art: An Approach to a Theory of Symbols.* Indianapolis: Bobbs-Merrill, 1968.

[164] E. Cassirer, *Philosophie der Symbolischen Formen.* Berlin: B. Cassirer, 1931.

[165] A. Plantinga, *Warranted Christian Belief.* New York: Oxford University Press US, 2000.

[166] R. Rorty, *Philosophy and the Mirror of Nature.* Princeton: Princeton University Press, 1979.

[167] S. C. Rickless, *Plato's Forms in Transition: A Reading of the Parmenides.* New York: Cambridge University Press US, 2006.

[168] M. L. Gill, "Design of the Exercise in Plato's *Parmenides*," *Dialogue*, vol. 53, no. 3, pp. 495–520, 2014.

[169] S. Gibb, "Closure principles and the laws of conservation of energy and momentum," *Dialectica*, vol. 64, no. 3, pp. 363–384, 2010.

[170] P. Bieri, "Generelle Einleitung," in *Analytische Philosophie des Geistes* (P. Bieri, ed.), p. 9, Königstein: Hain, 1981.

[171] G. Brüntrup, *Philosophie des Geistes: Eine Einführung in das Leib-Seele-Problem.* Stuttgart: Kohlhammer, 2018.

[172] C. D. Broad, *The Mind and its Place in Nature.* London: Routledge & Kegan Paul, 1925.

[173] H. Robinson, *Perception and Idealism: An Essay on How the World Manifests Itself to Us, and How It (Probably) is in Itself.* Oxford, GB: Oxford University Press, 2022.

[174] S. Soames, *Philosophy of Language.* Princeton: Princeton University Press, 2010.

[175] S. Shapiro, *Thinking About Mathematics: The Philosophy of Mathematics.* New York: Oxford University Press US, 2000.

[176] P. C. Wason, "Reasoning about a rule," *Quarterly Journal of Experimental Psychology*, vol. 20, no. 3, pp. 273–281, 1968.

[177] E. Wigner, "The unreasonable effectiveness of mathematics in the natural sciences," *Communications in Pure and Applied Mathematics*, vol. 13, pp. 1–14, 1960.

[178] R. J. Gerrig, *Psychologie.* München: Pearson Deutschland, 2018.

[179] K. A. Augustyn, "Life transcends computing," *Journal of Cognitive Science*, vol. 22, pp. 1–40, 2021.

[180] M. D. Hauser, N. Chomsky, and W. T. Fitch, "The faculty of language: What is it, who has it, and how did it evolve?," *Science*, vol. 298, pp. 1569–1579, 2002.

[181] R. Jackendoff and S. Pinker, "The nature of the language faculty and its implications for evolution of language (reply to fitch, hauser, and chomsky)," *Cognition*, vol. 97, no. 2, pp. 211–225, 2005.

[182] H. Thomä and H. Kächele, *Psychoanalytische Therapie.* Berlin: Springer, 2006.

[183] G. Graham, *The Disordered Mind: An Introduction to Philosophy of Mind and Mental Illness.* New York: Routledge US, 2010.

[184] J. Read, J. van Os, A. P. Morrison, and C. A. Ross, "Childhood trauma, psychosis and schizophrenia: a literature review with theoretical and clinical implications," *Acta Psychiatrica Scandinavica*, vol. 112, no. 5, pp. 330–350, 2005.

[185] J. Addington, M. Farris, D. Devoe, and et al., "Progression from being at-risk to psychosis: next steps," *npj Schizophrenia*, p. 27, 2020.

[186] B. van der Kolk, *The Body Keeps the Score: Bain, Mind, and Body in the Healing of Trauma.* New York: Viking, 2014.

[187] G. Stamer, *Kritik des lebendigen Verstandes, Erkenntnistheoretischer Entwurf zu einer Theorie der Einheit von Geist und Leben.* Würzburg: Königshausen & Neumann, 2021.

[188] U. Voigt, "Menschliche Personalität als Problem des Panpsychismus," *Salzburger Jahrbuch für Philosophie*, vol. 65, pp. 145 – 158, 2020.

[189] S. Robins, "The role of memory science in the philosophy of memory," *Philosophy Compass*, vol. 17, no. 10, p. e12880, 2022.

[190] D. Bassett, P. Zurn, and J. Gold, "On the nature and use of models in network neuroscience," *Nat Rev Neurosci*, p. 566–578, 2018.

[191] D. Bassett, K. Cullen, S. Eickhoff, and et al., "Reflections on the past two decades of neuroscience," *Nat Rev Neurosci*, p. 524–534, 2020.

[192] W. Gerstner, H. Sprekeler, and G. Deco, "Theory and simulation in neuroscience," *Science*, vol. 338, no. 6103, pp. 60–65, 2012.

[193] X.-J. Wang, H. Hu, C. Huang, H. Kennedy, C. T. Li, N. Logothetis, Z.-L. Lu, Q. Luo, M.-m. Poo, D. Tsao, S. Wu, Z. Wu, X. Zhang, and D. Zhou, "Computational neuroscience: a frontier of the 21st century," *National Science Review*, vol. 7, pp. 1418–1422, 06 2020.

[194] C. Presigny and F. De Vico Fallani, "Colloquium: Multiscale modeling of brain network organization," *Rev. Mod. Phys.*, vol. 94, p. 031002, Aug 2022.

[195] B. Blankertz, L. Acqualagna, S. Dähne, S. Haufe, M. Schultze-Kraft, I. Sturm, M. Ušćumlic, M. A. Wenzel, G. Curio, and K.-R. Müller, "The berlin brain-computer interface: Progress beyond communication and control," *Frontiers in Neuroscience*, vol. 10, 2016.

[196] C. Koch, M. Massimini, M. Boly, and et al., "Neural correlates of consciousness: progress and problems," *Nat Rev Neurosci*, p. 307–321, 2016.

[197] A. Seth and T. Bayne, "Theories of consciousness," *Nat Rev Neurosci*, p. 439–452, 2022.

[198] F. Crick and C. Koch, "A framework for consciousness," *Nat Neurosci*, p. 119–126, 2003.

[199] W. Gerstner, A. K. Kreiter, H. Markram, and A. V. M. Herz, “Neural codes, firing rates and beyond,” *Proceedings of the National Academy of Sciences*, vol. 94, no. 24, pp. 12740–12741, 1997.

[200] C. von der Malsburg, “The what and why of binding: The modeler’s perspective,” *Neuron*, vol. 24, no. 1, pp. 95–104, 1999.

[201] J. Levine, “Materialism and qualia: The explanatory gap,” *Pacific Philosophical Quarterly*, vol. 64, no. October, pp. 354–61, 1983.

[202] M. E. Raichle and M. A. Mintun, “Brain work and brain imaging,” *Annual Review of Neuroscience*, vol. 29, no. 1, pp. 449–476, 2006.

[203] M. E. Raichle, “The brain’s default mode network,” *Annual Review of Neuroscience*, vol. 38, no. 1, pp. 433–447, 2015.

[204] J. I. Gold and M. N. Shadlen, “The neural basis of decision making,” *Annual Review of Neuroscience*, vol. 30, no. 1, pp. 535–574, 2007.

[205] S. Marek, B. Tervo-Clemmens, F. Calabro, and et al., “Reproducible brain-wide association studies require thousands of individuals,” *Nature*, p. 654–660, 2022.

[206] F. Beck and J. C. Eccles, “Quantum aspects of brain activity and the role of consciousness.,” *Proceedings of the National Academy of Sciences*, vol. 89, no. 23, pp. 11357–11361, 1992.

[207] M. Tegmark, “Importance of quantum decoherence in brain processes,” *Phys. Rev. E*, vol. 61, pp. 4194–4206, Apr 2000.

[208] C. Smith, “The ‘hard problem’ and the quantum physicists. part 2: Modern times,” *Brain and Cognition*, vol. 71, no. 2, pp. 54–63, 2009.

[209] J. A. Tuszynski and et al., “Microtubules as sub-cellular memristors,” *Scientific reports*, vol. 10, p. 2108, 2020.

[210] M. Derakhshani, L. Diósi, M. Laubenstein, K. Piscicchia, and C. Curceanu, “At the crossroad of the search for spontaneous radiation and the orch or consciousness theory,” *Physics of Life Reviews*, vol. 42, pp. 8–14, 2022.

[211] M. van Oostrum and et al., “The proteomic landscape of synaptic diversity across brain regions and cell types,” *Cell*, vol. 186, pp. 5411–5427, 2023.

[212] T. Branco and K. Staras, “The probability of neurotransmitter release: variability and feedback control at single synapses,” *Nat Rev Neurosci*, p. 373–383, 2009.

[213] J. Cao, R. J. Cogdell, D. F. Coker, H.-G. Duan, J. Hauer, U. Kleinekathöfer, T. L. C. Jansen, T. Mančal, R. J. D. Miller, J. P. Ogilvie, V. I. Prokhorenko, T. Renger, H.-S. Tan, R. Tempelaar, M. Thorwart, E. Thyrhaug, S. Westenhoff, and D. Zigmantas, “Quantum biology revisited,” *Science Advances*, vol. 6, no. 14, p. eaaz4888, 2020.

[214] J. Summhammer, “Mental intervention in quantum scattering of ions without violating conservation laws,” 2024. arxiv:2406.08601.

[215] B. J. Kagan, A. C. Kitchen, N. T. Tran, F. Habibollahi, M. Khajehnejad, B. J. Parker, A. Bhat, B. Rollo, A. Razi, and K. J. Friston, “In vitro neurons learn and exhibit sentience when embodied in a simulated game-world,” *Neuron*, vol. 110, no. 23, pp. 3952–3969.e8, 2022.

[216] H. Cai, Z. Ao, C. Tian, and et al., “Brain organoid reservoir computing for artificial intelligence,” *Nat Electron*, p. 1032–1039, 2023.

[217] J. Tang, A. LeBel, S. Jain, and et al., “Semantic reconstruction of continuous language from non-invasive brain recordings,” *Nat Neurosci*, p. 858–866, 2023.

[218] Z. Chen, J. Qing, and J. H. Zhou, “Cinematic mindscapes: High-quality video reconstruction from brain activity,” 2023. arxiv:2305.11675.

[219] E. J. Green, “The perception-cognition border: A case for architectural division,” *Philosophical Review*, vol. 129, no. 3, pp. 323–393, 2020.

[220] J. C. Herron and S. Freeman, *Evolutionary Analysis, 5th Edition*. London: Pearson, 2014.

[221] K. Laland, T. Uller, M. Feldman, and et al., “Does evolutionary theory need a rethink?,” *Nature*, p. 161–164, 2014.

[222] D. J. Futuyma, “Evolutionary biology today and the call for an extended synthesis,” *Interface Focus*, vol. 7, no. 5, p. 20160145, 2017.

[223] T. Nagel, *Mind and Cosmos: Why the Materialist Neo-Darwinian Conception of Nature is Almost Certainly False*. New York: Oxford University Press US, 2012.

[224] C. M. Jakobson and D. F. Jarosz, "What has a century of quantitative genetics taught us about nature's genetic tool kit?," *Annual Review of Genetics*, vol. 54, no. 1, pp. 439–464, 2020.

[225] S. A. Kauffman, *A World Beyond Physics, The Emergence and Evolution of Life.* Oxford: Oxford University Press, 2019.

[226] R. Roberts, S. Pop, and L. Prieto-Godino, "Evolution of central neural circuits: state of the art and perspectives," *Nat Rev Neurosci*, p. 725–743, 2022.

[227] D. O. Brink, *Perfectionism and the Common Good: Themes in the Philosophy of T. H. Green.* Oxford: Oxford University Press, 2003.

[228] L. A. Paul, *Transformative Experience.* Oxford: Oxford University Press, 2014.

[229] M. E. P. Seligman, *Flourish: A Visionary New Understanding of Happiness and Well-being.* New York: Free Press, 2011.

[230] V. Hösle, "Einstieg in den objektiven Idealismus," in *Idealismus heute: aktuelle Perspektiven und neue Impulse* (V. Hösle and F. S. Müller, eds.), wbg Academic, 2015.

[231] E. Klein, *Why We're Polarized.* New York: Avid Reader Press, 2020.

[232] K. Pistor, *The Code of Capital: How the Law Creates Wealth and Inequality.* Princeton: Princeton University Press, 2019.

[233] I. Robeyns, *Limitarianism: The Case Against Extreme Wealth.* London: Allen Lane, 2024.

[234] M. Mazzucato, *Mission Economy. A Moonshot Guide to Changing Capitalism.* London: Allen Lane, 2021.

[235] V. Hösle, *Philosophie der Ökologischen Krise: Moskauer Vorträge.* Frankfurt am Main: C.H.Beck, 1994.

[236] M. A. Killingsworth, D. Kahneman, and B. Mellers, "Income and emotional well-being: A conflict resolved," *Proceedings of the National Academy of Sciences*, vol. 120, no. 10, p. e2208661120, 2023.

[237] E. D. Galbraith, C. Barrington-Leigh, S. Miñarro, S. Álvarez Fernández, E. M. N. A. N. Attoh, P. Benyei, L. Calvet-Mir, R. Carmona, R. Chakauya, Z. Chen, F. Chengula, Álvaro Fernández-Llamazares,

D. G. del Amo, M. Glauser, T. Huanca, A. E. Izquierdo, A. B. Junqueira, M. Lanker, X. Li, J. Mariel, M. D. Miara, V. Porcher, A. Porcuna-Ferrer, A. Schlingmann, R. Seidler, U. B. Shrestha, P. Singh, M. Torrents-Ticó, T. Ulambayar, R. Wu, and V. Reyes-García, "High life satisfaction reported among small-scale societies with low incomes," *Proceedings of the National Academy of Sciences*, vol. 121, no. 7, p. e2311703121, 2024.

[238] E. W. Dunn, L. B. Aknin, and M. I. Norton, "Spending money on others promotes happiness," *Science*, vol. 319, no. 5870, pp. 1687–1688, 2008.

[239] D. Meadows, *Thinking in Systems: A Primer.* White River Junction: Chelsea Green Publishing, 2008.

[240] H.-D. Mutschler, *Ästhetik und Metaphysik. Die abgerissene Verbindung.* Darmstadt: wbg Academic, 2023.

[241] T. S. Kuhn, *The Structure of Scientific Revolutions.* Chicago: University of Chicago Press, 1962.