Unhelpful! Mindsets that I found less than conducive to fully grasp, let alone make progress with, the mind/body problem

Martin Korth*

IVV NWZ, WWU Münster, Wilhelm-Klemm-Str. 10, 48149 Münster (Germany)

* Corresponding author email: dgd@uni-muenster.de

v2, 2025/05/19; v1, 2025/05/16

Concerning the mind/body problem, most people seem to have basic intuitions about the nature of this problem that lie somewhere on a spectrum between what one could call an 'inflated' and a 'deflated' view of subjectivity, experience and human thought.

On the 'inflated' side, people take a strong view of subjectivity, the central importance of phenomenological experiences and often also special human cognitive abilities as so obvious, that they and not some 'scientific poetry' on top of them should be taken as the basis for our understanding of the world. The cogito argument, the (at least perceived) freedom of will and qualia are then seen as clearly supporting this view, and seem to call for a revision of the common physicalist world view, alas without giving any unambiguous hints on how this could be done. For the inflationist there exists something like qualitative information that is intuitively grasped (as quality or idea), and human thought can accordingly be much more than 'symbol pushing', with rich pre-, sub- and unconscious contributions at the bottom as well as access to universal abstract entities at the top. But how to connect this to the scientific concept of physical, quantitative information processing?

On the 'deflated' side, subjectivity is understood to be no more than the access to self-locating information, and phenomenological experience and meaning are just special types of concepts and properly coordinated, physically realized information. For the deflationist there is only quantitative information, so that human thought boils down to physical information processing and meaning to optimized (logico-)mathematical relations, e.g. in a web of symbols or as what remains constant over 'all possible worlds'. The underlying assumption is that not only symbolic structure, but also qualitative features arise from self-organization on the basis of feedback to actions in some form of reinforcement-type learning; only that it is still very unclear how broad and

stable abstractions could self-implement in neural-network type systems, let alone how structural relations from the individual coordination of quantitative information could be uniquely connected with identity, quality or universal elements of meaning. The deflated view is very much in line with current physics and AI research, but does not seem to cover the human experience in all aspects, and is therefore in danger of missing important clues even for the further development of physics and AI research themselves. (The later is really now repeating all of this including all errors; we have to see whether there is anything to the saying that sometimes problems were only solved because young people didn't know that it couldn't be done.) In trying to understand, let alone make progress with, the mind/body problem, I went through several mindsets that, in retrospect, were not very helpful. To remind myself to not fall into the same trap twice, I have written this manuscript.

Unhelpful mindsets

For truly grasping the mind/body problem, I found it first of all very unhelpful, to not grant that the 'deflated' view could indeed provide a valid solution, which then leads to iterating endlessly on the importance of subjectivity, free will and qualia – without being able to make any definite argument that would not require my basic intuition about their importance to hold true in the first place. To me it seems of fundamental importance to really understand how powerful, coherent and helpful the physicalist view is; otherwise one will always be tempted to suggest quasi-religious, for instance eastern or classical idealist 'background stories', with all progress still coming from science alone. One will also always underestimate how easily arguments from subjectivity, free will and qualia are rationally subverted (which never proves the opposite, but still). Here, an important example is the cogito argument, which if I am convinced of the central importance of subjectivity, is an extremely powerful argument against physicalism. But the point is that the physicalist does not question whatever comes after the cogito (qualia, free will, ...) argument, for instance that the cogito then proofs that we are not finite, physical machines. Instead, the physicalist denies the basic intuition that the cogito should not be more properly formulated as 'there are thoughts happening', thus diffusing the seemingly unquestionable existence of a subject into the mere existence of physical information processes. Other examples are statements like 'x is the only possible metaphysics', that issues connected to the measurement problem in quantum mechanics proves physicalism to be wrong, that it is obviously impossible that human thought and culture could have developed from

biophysical evolution, or that human thought without question exceeds the cognitive capabilities of any conceivable physical system.

I found it equally unhelpful, on the other hand, to not grant that the mentioned 'inflated' intuitions are very good reasons to question the physicalist world view. Physics itself is well aware of its unfinished business and always on the search to extend its metaphysical basis (e.g. from solid bodies to quantum objects to maybe strings). It's also not as if physicalism has already proven itself to be valid also for describing mental phenomena by giving a sound explanation of the neural code and how this then produces the phenomena of the above intuitions. It is also quite open how meaning supervenes on physically implemented information, how in general abstract structures are generated on top of neural-network like information processing, and how 'universal' symbol systems like logic, mathematics and maybe also values like truth fit into this picture. Based on our current knowledge, we should indeed see physicalism as the default mode when pressed for practically relevant answers, e.g. concerning medical advice or also the distribution of research money, but there should also be no question that it is well worth and necessary to explore bold new ideas on the mind/body problem. Exploration beyond the boundaries of current science is not bad science, but part of the process; only that the explorers then of course have to adhere to common standards of rationality when putting forward their arguments.

In the middle, we have on the one hand all sorts of arguments which defuse and by this actually entrench the real problem: This includes explanatory pluralism, epiphenomenalism, aspect dualism, etc. These approaches are very powerful in the philosophical boxing ring, but are not at all helpful to build a combined neuroscientific and psychological theory on top of them, as they simply pass over how exactly physical and mental phenomena are interfaced. All are good 'background stories' for the 'opening-up' physicalist, but they don't explain anything new and they can't guide us to new experiments.

On the other hand, the middle is also (densely) populated with theories in which models are developed, like for instance 'theories of consciousness', in which elements of conscious experience and subjectivity are simply assigned to functional elements of a model, often with the claim that this solves the 'hard problem'. The thing is, with such assignments we just get out what we put in; 'defining' the qualitative feeling of pain as being a certain functional relationship does not solve the hard problem, it denies that there is such a problem. The hard problem is not about putting

forward ideas of how certain processes or structures can (ideally) explain the 'functional' part of conscious experience and should thus be seen as corresponding to the qualitative impression; it is to explain why and how this functionality should be connected to specific qualitative experiences: Why should the given physical process have any consciousness experience attached to it at all, and how would then THIS process have THIS specific qualitative impression connected to it? To say that conscious experience is just the 'inside-view' of a physical or informational process is not answering the question; solving the hard problem would mean to enable us to understand which parts of the process or structure give rise to which qualitative features. That's what we need to bridge neuroscience and (phenomenological) psychology; otherwise we can just stick with physicalism and posit qualitative experience as the 'inside-view' of neural correlates: Neither way are we told how exactly the 'what-it-is-like' is connected to the rest. One can see this also from asking whether the proposed model is in any sense 'effective'; does it help us to push forward our knowledge about how brains and minds are related beyond the proposed initial assignment? (There is without question a 'functional' part of qualitative features, as our mind is making use of them to interact with the world by assigning them to parts of the causal network of the material world and/or puts them into relation to other qualitative features, but the conscious experience of those features is not exhausted in this functional part.)

Another variant of this are ideas like enactivism or agent-based models, where qualitative features are assigned to being in certain functional contexts, which if interpreted physically is adding nothing substantially new to physicalism, and if interpreted phenomenlogically, is still unhelpful to build a bridge to neuroscience, as now the other part is missing. We want both a brain embedded in quantitative informational relations AND a mind embedded in qualitative experiences and furthermore know the details of how exactly they belong together.

Once we have come to the point where we allow us to explore ideas beyond physicalism (not because we know that it is wrong, but because physicalism might need to be extended), our approach will most likely still be influenced by our basic intuitions about the problem. More deflationary types will then start from assuming that physical structures give rise to non-material structures due to additional psycho-physical laws. These laws would connect physical and non-physical parts, which make up dualist or panpsychist brains and minds. As a result we most likely arrive at 'weak' subjects as no more than bundles of their (partly non-physical) elements. Here,

we extrapolate up from (current) physics; the challenge is that the required psycho-physical laws look arbitrary – and arbitrarily complex. Nevertheless, this is a fair challenge to take up. What I found unhelpful on this road is the often dualist wish for a non-material mental part without making a commitment to the objective existence of non-physical properties. Assuming that the quality red is a thing, but then arguing for the existence of only physical properties, requires some 'cloudy' soul-like subjectivity 'floating' above the material brain; this is an extremely hard, if not impossible path towards connecting neuroscience and psychology. We already know from psychology that our mind (physical or not) is to a large part a 'bundle' of bodily, sub-conscious, as well as conscious elements, so that although we don't know exactly how monolithic or not our (dualist, panpsychist or idealist) mind is, we want to allow it to have at least certain distinct elements, also content-wise, for which we must then assume ontological independence of whatever we find to be the core of subjectivity – if there is one. (The thing is; if we allow for some non-physical properties, it's quite hard to avoid not going down the whole path to idealistic panpsychism or idealism; this seems to be the reason why especially wishful dualists stick to trying to have their cake and eat it.) It is in general then also unhelpful on the way up from the bottom to ignore that the regularities of the mental world will require some regular mechanisms across the mind/matter divide and also beyond. Many theories connecting quantum theory and consciousness experience for instance give absolute no clue on why we experience a mind/matter divide and how brain and mind activity would be connected in detail. Additionally, I found it unhelpful in these cases to argue for a strong subject arising from a bundle of non-physical properties without proposing any additional elements or mechanisms that promote a weak (bundle-like) subject to a strong (properly free) one. On the way up, not making a commitment to the existence of non-physical properties, and not proposing any laws that cross the perceived mind-matter gap, is delivering less and not more than physicalism.

On the other side, more inflationary types will likely try to extrapolate down from human experience to biology, continuing all the way down to something like 'physical microbes' acting at the very bottom of reality, for instance conceding some sort of mentality even to elementary particles. Natural history then starts from extremely simple subjects ('monads' maybe) that manipulate properties to form the world and everything in it, including proper minds. Non-physical structures are then made from non-physical properties (e.g, the color red), like physical structures are made

from physical properties (e.g., charge), with no need for properties to be 'physical' right from the start and also no need for the pre-existence of complex ideas; simple building blocks will do also in the mental world. Here, 'strong' subjects are the driving force behind the manipulation of even elementary properties; the challenge is how to derive the observed physical laws and the perceived dualism from this setup (the 'emanation problem'). On this way down from life to physics, I found it first of all unhelpful to claim that this somehow implies or even requires that 'all is one' (like in classical idealism, but also modern analogs like analytical idealism, and related ideas like qbism, etc.), without giving detailed mechanisms of how separate properties (and subjects and physical laws) could arise from this; classical idealism died because it was an experimentally empty claim. It is furthermore unhelpful for such theories to give no explanation for the perceived matter/mind gap, i.e. the observed dualism, and to not try to derive the laws of physics etc. in detail from the underlying model, by just jumping from phenomenology to physics, for instance with proposing the existence of some 'inside-view' of physical processes. Also such an inside-view would require the objective existence of the properties which make up the contents of this view, some form of specified subjectivity that acts as the nexus of these experiences, and some form of interaction from there to the details of physical processes and structures. Many idealist and panpsychist models are then again nothing more than good background stories, this time for the anti-physicalist, as they again can't explain anything new and don't guide us to new experiments.

The game is not to come back to existing inflationist or deflationist arguments again and again, or worse to switch back and forth between them. The game is also not to propose inflationist or deflationist ideas, but not go through with the implications of those ideas (e.g. call for dualism, but row back on the objective existence of non-physical properties or argue that 'models of consciousness' or physicalist enactivism solve the hard problem instead of denying it.). The game is not to insist on physicalism, but then fail to deliver on the details of human cognition. The game is not to propose dualist models without detailed psycho-physical laws (ideas for which still remain elusive). The game is not to propose panpsychist models without any explanatory power beyond physicalism. The game is not to propose idealistic models without making clear which kind of basic subjectivity and which non-physical properties are assumed to exist, and without giving a solution to the emanation problem, of how the perceived dualism and the details of our physical laws arrive from this setup. It's choose your hard problem, but dancing around all of them is just

unhelpful.

The game is to propose detailed – though at first maybe fully hypothetical – models, for which we then have to show how they recover the 'phenomenological' structure (though not necessarily the proposed ontology) of our best scientific theories, and from which we then have to derive new experiments, which would not make sense at all based on current physics alone. (Similar to quantum mechanical experiments against the background of classical mechanics.) It is unhelpful to not be willing to stick out your head and propose detailed hypothetical models. It is unhelpful to concentrate on solutions to the mind/body which cover only specific problems in certain subfields. It is unhelpful if the proposed model does not allow to bridge psychology and neuroscience, which amongst others also requires explanations of the connections between subjectivity and physics, the evolutionary purpose of qualia, why we at least perceive free will etc.

More deflationary types could then try to empirically answer questions like: Does mental causation violate energy conservation? Are mental phenomena involved in quantum measurement? What detailed, predictive brain/mind relationships can be identified? More inflationary types could try to empirically answer questions like: Does the build-up from universal properties explain the quantum nature of the micro-scale? Do 'higher-level' natural laws come from this setup, too? Does the 'evolutionary growth' of the physical world explain non-uniformities on the cosmic scale? Does the evolutionary buildup from monads/universals explain the structuring of brains (e.g. as composite of cells)? And both types could try to answer questions like: Is the non-material necessary to coordinate biological processes (e.g., for stable gene expression, cell organization, ontogenesis, evolutionary mechanisms, etc.)? Does the non-material enhance human information processing? (Given evolution, it better does - but if so, how exactly?) Does the brain show too little activity/information content for certain though processes, because parts of them are non-material?

(My) Outlook

To develop my own ideas on the above questions, I follow certain rather inflationary basic convictions. (0. Physicalism might be right, but at the current stage, it is a well worth and necessary venture to ponder alternatives.) 1. The 'phenomenological' structure (unlike the proposed ontology) of our currently best scientific theories has to be taken 'as is'; e.g., no easy dismissal of energy conservation, etc. 2. The evolutionary, continuous development from matter to life and

from simple organisms to humans is also a given. 3. Successful models should allow us to bridge neuroscience and (phenomenological) psychology. 4. There is a genuine 'hard problem' of qualia(, free will, subjectivity, ...), that physicalism cannot solve already for conceptual reasons. 5. At least some mental entities are of universal nature (e.g., math, which science needs) and this gives human thought its power. (This point is motivated not so much by scientific, but philosophical problems with abstract entities, as well as our recent advances and setbacks with artificial intelligence.) I thus believe we have to solve a 'broad' mind/matter, not only a narrow mind/brain problem and that the later cannot be solved independently of the former.

Based on these 'primary' convictions, I have acquired a set of 'secondary' convictions: 1. Qualia must be basic (and as a result universal) entities; otherwise we have to explain how they could be 'generated' by particle positions etc. (and strong emergence just means 'we don't know'). 2. Some form of 'core' subjectivity must be basic, too; otherwise we have to explain how subjects are formed from mind dust our split from a cosmic soul (if you hold inflationary intuitions, this results then in the hitherto unsolved combination problem; for emergence see above). 3. For several reasons (the interaction problem, continuous evolution, ...), the mind/matter gap must be a resulting, not a principal gap, i.e. we should aim at a monist theory; otherwise we have to explain how (especially first) interactions across the gap work. 4. All my convictions could be satisfied with an idealistic theory (though not in the sense of a 'naive' Platonism) – if we can show how modern science fits in ('the emanation problem'). To make it more clear how this could work out in detail, I have proposed 'Model A' or 'A-world' as a first 'consensus model',[2] i.e. a model that tries to satisfy all intentions behind different metaphysical options rather than combine proposed solutions.[1] But this is really just an invitation to develop consensus models, in my case under the label of a progressive, secular, pragmatic idealism. There can be no question that in the not too distant future, I will find this Unhelpful!, too.

References

[1] M. Korth, How to construct consensus models to (maybe) make sense of the mind-body problem, 2025. https://philpapers.org/rec/KORHTC

[2] M. Korth, Information, Intelligence and Idealism, 2025. https://philpapers.org/rec/KORIIA-4