# Why physics should care about the mind, and how to think about it without worrying about the mind-body problem

Jenann Ismael
Columbia University
Philosophy

Abstract

This paper argues for a division between the aspects of the mind that physics can (and must) cope with, and the aspects that it can't cope with, but can ignore.

It is uncontroversial to say that physics does not have a very finely etched understanding how to fit the mind into its account of the natural world. It is not that physics has any particular problem with the brain and body. These are made of the same stuff as, and obey the same laws as, trees and planets. And it is not that we don't know how to talk about the mind when we can describe it in the vocabulary recognizable from our own experience: we treat it as an information-processing system whose job is to transform the flow of information coming in through perceptual pathways into action. The problem is that it is not obvious how to bring the mind itself under the scope of physical theory and to treat it as *part* of the world. We don't know how to describe the mind in the terms that physical theory itself provides, that is to say, in such a way that it *contains* our experience within it.i

Physics has all kinds of workarounds to avoid focusing on experience. We talk about 'observation', but by 'observation', we often mean 'measurement', so the mind never enters into it. We talk about evidence, but always stated in the language of physics; the positions of pointers on the front of measuring instruments or marks on a photographic plate. Because the evidence for our theories ultimately comes from experience, however, eventually we have to be able to bring the mind itself firmly under the scope of our physical theories and understand how human experience fits into the picture.

**The intrusion of the observer**

Talk of experience has begun to creep into physics in a variety of ways. In quantum mechanics, the contrast between the deterministic dynamics yields and what the observer *sees* has been at the forefront of the theory almost from the beginning.

In the controversies surrounding that status of time, straightforwardly physical questions like whether there is a global present have gotten intertwined with questions about the subjective experience of time. Questions like whether we really have experience of the passage of time, and whether flow is a property of the world or internal to the mind, are used to question the relativistic conception of time.

In discussions of quantum gravity where it is sometimes said that 'space disappears' at the fundamental level, there is a need to understand what to make of the spatial character of our experience. In what sense do we see space and what sort of requirement does that place on physical theory? Does it make the existence of space as an external structure a non-negotiable desideratum for physics, or is there some weaker requirement? ii

It's a sign of maturity of physics that these questions are arising. Experience our ultimate source of information about the world, and although we can get pretty far with a rough and ready division between what is *out there* and what is *in here*, eventually, this is going to come under pressure. What information experience contains precisely depends on details of the perceptual processing between skin and skull.iii Quite generally, before any physical lessons can be drawn from experience, we need to sort out which aspects of experience belong to the world and which belong to the mind. And that demands bring the coupling between mind and world into clearer focus.iv

Precision about none of these things is possible until we say firmly and clearly how to recover our experience in a way that makes a clear division between mind-independent structure in the external environment and mind-dependent structure.

**The pessimistic reaction**

The pessimistic reaction to the intrusion of experience into physics is a kind of horror. The thought is that there are reasons to steer a wide berth from talking about the mind: it's a morass of endless and fruitless debate. Physics is about the movements of material things. If progress of physics depended on resolution of the mind-body problem, it would be a terrible thing. Physics has gotten as far as it has is precisely because it has left the messy business of human experience alone. And it does not help that the people who have been willing to talk about consciousness have often approached from a fringe perspective. A desire to keep physics *physics*, and to avoid the squishiness

of philosophical debate is more than enough reason (one might think) to stay away from talk of experience.

I think the pessimistic reaction is too pessimistic. Physics doesn't stop at the surface of the skin. Some understanding of observation as a physical process is always implicit in bringing evidence to bear on theory and the fact that questions about experience are beginning to infect physics at quite fundamental level suggests it is time to bring them into focus. And an increasing amount is known about the mind in purely scientific terms. Because the explanatory emphases and points of departure for physics are very different than from philosophy, problems and issues that are deeply contested and tend to form focal points for debate in those discussions can be set aside. I'm going to give a quick, opinionated account of what matters and what can be sidelined and then a sketch of how to fill in the outlines of the parts that matter.

**What we can ignore: The mind-body problem and physics**

First: what can be sidelined. The mind-body problem is one of the oldest, and most intractable problems of philosophy. It concerns the relationship between the mind and the body -- between the realm of experience and the realm of matter. The question is whether the progression of thoughts, feelings, perceptions, sensations, that make up our mental lives are things that happen in addition to the physical processes in the brain, or are themselves just some of those physical processes?

The last 40 or 50 years has seen an explosion of scientific progress understanding the human mind, in no small part as the result of the emergence of cognitive science; interdisciplinary study of mind and intelligence, embracing philosophy, psychology, artificial intelligence, neuroscience, linguistics, and anthropology. Its intellectual origins are in the mid-1950s, as behaviorism was finally fell out of fashion and people began developing theories of mind based on complex representations and computational procedures. The field employs a fruitful mixture of methods. Cognitive abilities are typically functionalized and studied. Mechanisms and neural implementations sought for things like the ability to discriminate, categorize, and react to environmental stimuli; the ability to integrate information coming through different perceptual pathways, to take its own states as objects of representation, to organize them, and report them; to impose coherence and consistency constraints; to regulate attention, and control behavior.

In light of all of the progress made understanding the mind in scientific terms, there was a general attitude of optimism that cognitive science might be resolving this age-old philosophical problem. In 1996, David Chalmers threw cold water on that optimism with an article, followed by extremely

influential book in which he argued that none of this evident progress touches the heart of the mind-body problem.v Chalmers separated problems into two classes: Easy and Hard. Easy Problems concerned cognitive abilities like those above that can be characterized in functional terms. Once an ability is characterized in functional terms, computational and neural mechanisms that give rise to the ability can be sought. The ultimate goal is to uncover how the brain supports the ability in question. These kinds of problems, though not actually easy, are at least amenable to scientific understanding. The Hard Problem, according to Chalmers, is the problem of accounting for subjective experience. This one, he argued, is not amenable to scientific understanding, because it concerns a notion of the qualitative character of mental states that is *not* functionally definable. The notion of qualitative character he has in mind is something you are supposed to recognize from your own case:

"When we see… we experience visual sensations: the felt quality of redness, the experience of dark and light, the quality of depth in a visual field. Other experiences go along with perception in different modalities: the sound of a clarinet, the smell of mothballs. Then there are bodily sensations, from pains to orgasms; mental images that are conjured up internally; the felt quality of emotion, and the experience of a stream of conscious thought. What unites all of these states is that there is something it is like to be in them. All of them are states of experience."

The term 'phenomenal consciousness' was coined to refer to the property of there being something it is like for a system to be in a given state. Chalmers argued that no matter how detailed an account we give of the cognitive and behavioral capacities that a system possesses or of the physical mechanisms that underwrite those capacities, that will leave undetermined whether it is phenomenally conscious. Whether it is phenomenally conscious is, he argued, a *further fact*, distinct from any collection of outwardly observable abilities, and one that (moreover) no amount of scientific investigation will settle.

Chalmers collected and organized the best of Descartes arguments, combined them with others that had been floating around in the literature, and provided some of his own, in a powerful case meant to bring the Hard Problem into relief and establish its scientific intractability. Here (briefly) is how the arguments go.vi Suppose we give some functional specification of what it was for a system to have introspectively accessible states and we offer introspective accessibility as an account of what it is for a state to be conscious. We will be met with arguments like this: we can imagine a being (a robot, for example) who had states that were introspectively accessible, but who was not conscious, so consciousness can't just be introspective accessibility. The same will go for global broadcast, informational integration, and any purely functional specification we can give: it

will always seem possible to imagine a creature that satisfied that specification but that wasn't conscious, and that is supposed to show that consciousness can't be a matter of satisfying one of these functional descriptions. Whatever it is to have conscious experience, the argument goes, it is not a matter of having a certain functional organization or cognitive or behavioral capacities, because any such organization, and any collection of such capacities, could be reproduced in a system that wasn't conscious.

Then we have a staring match that ensues here between those think that consciousness has to be some sophisticated, functionally definable notion (maybe a kind of informational integration and introspective accessibility), and those who think it the most obvious thing in the world that you could whatever functional characterization you give could be satisfied by an unconscious robot or a zombie. If these arguments are correct, phenomenal consciousness by its nature falls through any attempt to capture it by linking it to something that can be empirically investigated. That's what makes the Hard Problem *hard*. Whether or not you agree with the arguments, there is no question that they have a powerful intuitive force. They have prompted debate that has produced a mountain of baroque argumentation without showing any signs of resolution.

The nice part about all of this for our purposes is that although Chalmers' own purpose was to reestablish the heart of the mind-body problem as impenetrable to scientific resolution, he managed to isolate it almost surgically making a difference to physics. If there is such a thing as a kind of consciousness that by its nature falls through the net of physical description because it has no functional or causal role of its own - the physicist interested in the role of mind in nature doesn't worry about it.[vii] The physicist interested in representing the role that minds play in the causal fabric of the world can be serenely unconcerned whether phenomenal consciousness really is a kind of magic fairy dust that when sprinkled on certain processes, lights them up from the inside. It is concerned only with the shadow those processes cast in the physical world.[viii] As soon as consciousness matters to physics - i.e., as soon as it makes an observable impact on the motions of material things - it becomes detectable by that impact and integrated into the causal fabric of the world. And then it becomes (as far as physics is concerned) physical. [ix] What this means is that functional interpretations give us everything we need in order to address questions about observation and action as they appear in the problem space of physics.

The second debate that can be sidelined for purposes of the physicist is what we might think of as an analogous Hard Problem of intentionality; again, it is taken as the hallmark of mental states that they represent features of the world, and it has become a focal point of debate in the philosophy of mind to say what it is for one to represent something distinct from itself. Some of the central

arguments that make this an object of philosophical dispute purport to show that no functional account of what it takes for a state to have representational (or 'intentional') content could be right. If anyone offers such an account, they will be met with an argument that the functional description can be satisfied, and there be nothing like full-blooded content present. Searle's Chinese Room argument is the locus classicus of this kind of argument.x

Again, here, there is the scientific question: what functional role do representational states play in whatever happens between sensory impact and movement in a human being? And then there is the philosophical question of whether their having a content is purely a matter of playing that role. The Hard Problem of Intentionality is about bridging the gap between these functionally specifiable notions and some more full-blooded notion we are supposed to know in a first-personal way.

And again, it doesn't matter for physics. In this case, there is less consensus about whether the arguments have the conclusion that intentionality is by its nature non-physical. The reason is that there is more room for things like embedding the human in a social environment, enhancing it with memory and reflective processing, and in general imposing more structure on the setting (cognitive or environmental) in which mental states are used, that might make a difference to the persuasiveness of the arguments. On a wide conception of physical that includes the social environment, it is not nearly as clear that the features of meaning that make a direct reduction to physics difficult can't be ultimately emergent from social interactions. In this sense, it is harder to make the intuitive case that once all of the easy problems are solved, there won't be any Hard residue left. What is making the arguments for the Hard Problem of consciousness so powerfully convincing is the sense that we have immediate awareness of the phenomenal properties of our mental lives, and nothing that we can learn about another person seems to cross the divide. There is nothing analogous to that in the intentionality case.

The nice thing about all of this from the point of view of someone interested in the role of the mind in the physical world is that the arguments that are supposed to establish that some aspect of mind (e.g., Consciousness, Intentionality) is irreducible to physics only at the expense of making it irrelevant to physics. That's bad news if you want to solve the mind body problem. It's good news if you want to do physics without worrying about it. All that you need to care about for physical purposes is those aspects of mind that make a difference to the movements of physical things.

In saying this, I am not saying anything proponents of the Hard Problem in the vein of Chalmers would disagree with. The isolation of the Hard Problem and its separation from the Easy ones, and

the very features of the Hard Problem that make it scientifically intractable also make it irrelevant to physics. Physics can focus on the Easy Problems, which include finding the physical basis for consciousness. It doesn't need to worry about whether the relationship between the physical basis and the first-person accessible phenomenon is analytic entailment, metaphysical necessity, entailment by special psychophysical laws, or something else altogether. The point is just that the features of experience that escape functional characterization in terms of their causal relations to something in the environment, or to the movements of the human body, don't appear in the problem space of physics, and so nothing in the problem space of physics is going to depend on their resolution.xi As soon as Consciousness and Intentionality make a difference in these ways, they become something that matters to physics. But then they also become something that is characterizable in terms of their physical role. There is a kind of closure in the problem space.

**Does Quantum Mechanics make consciousness relevant to physics?**

A few words about the role that consciousness has played in the discussions of the foundations of quantum mechanics will illustrate all of this quite nicely. Any theory has to predict something about the observer's experience if it is going to make testable predictions. In classical contexts, observation stayed mostly out of view. There was a presumption that observation gives us information about the values of local macroscopic variables, and theories were tested by deriving implications for the values of such variables. In quantum mechanics, observation becomes problematic because of a conflict between the linear evolution of the wave function and what an observer sees at the end of a measurement. Linearity entails that an observer coupled to an apparatus carrying out a measurement on a system in a superposition of the measured observable should end up in a superposition of seeing different results, but the observer invariably sees a definite result. The difficulty has brought the coupled interaction between observer and measuring apparatus under careful scrutiny, and discussion of the observer's experience is made explicit in careful presentations since this is ultimately where the conflict occurs.

Although many of those working in quantum foundations think that solving the problem is a matter producing definite pointer states for the measuring apparatus, Shan Gao advocates a mentalistic formulation of the problem that makes reference to the observers experience explicit, since the dynamics on its own is perfectly consistent. As he says, one can see the influential responses to the measurement problem as advocating different kinds of psychophysical linkages. He writes:

"The mentalistic formulation of the measurement problem highlights the important role of psychophysical connection in causing the measurement problem. By this new formulation, we can look at the solutions of the problem from a new angle. In particular, Bohm's theory, Everett's theory and collapse theories correspond to three different forms of psychophysical connection (as well as three different result assumptions)."xii

Each of these three theories assumes some form of psychophysical supervenience.xiii If one assumes psychophysical supervenience, whatever form it takes, there will be necessary connection between a conscious state and its physical basis and the dynamical role of conscious states won't be any different from that of the brain states on which they supervene. This means that the physicist can safely focus on the phenomenal properties without worrying that he's missing something that makes a difference to the dynamics. If psychophysical supervenience is assumed, physics is not going to know the difference between the conscious state and its physical basis, and consciousness is – *for that very reason* – irrelevant to the physics.

There is, however, a small and well-established tradition stemming from Wigner that treats consciousness itself as a physical agent. The thought is a natural one, given the predicament in quantum mechanics. It is that consciousness comes from outside the known physics, and isn't supervenient on anything that falls under the linear dynamics. If the states in which observations terminate come from outside the known physics, they don't need to be governed by Schrodinger's equation and one is free to expand the theory by thinking of consciousness as inducing collapse.

There are two ways of developing the suggestion. One is that there is some new physical quantity or stuff outside the *known* physics, on which mental states supervene, which induces collapse. If this is the suggestion, then there is an extension of the physical ontology, but psychophysical supervenience still holds and the situation can be assimilated to the one above. The more interesting way to develop the suggestion, and the one that Wigner seems to have intended, is to hold that the physical interaction between measured system and human brain is described by Schrodinger evolution, but consciousness itself (not some hitherto unknown physical stuff on which it supervenes) induces collapse. This suggestion denies the supervenience of the mental on the physical and treats *consciousness* as a physical agent in its own right.

As a resolution of the measurement problem, the proposal hasn't found many supporters.xiv But it is interesting as a test case for the claim that physics can ignore the Hard Problem of Consciousness by showing us what happens as soon as consciousness becomes relevant to physics. It amounts to a solution to the Hard Problem.

The Hard Problem was supposed to bet hat there is a kind of consciousness - phenomenal consciousness - that is by its nature undetectable and physically elusive. Whether a system is phenomenally conscious was supposed to be detectable from the inside, only by the system itself. As Chalmers said:

"We know that a theory of consciousness requires the addition of something fundamental to our ontology, as everything in physical theory is compatible with the absence of consciousness."

The proposal that is being entertained here – i.e., that consciousness is a causally active, physical stuff that induces wave-function collapse – amounts to a solution to the denial that everything in physical theory is compatible with the absence of consciousness. If consciousness *itself* induces collapse, consciousness is an ineliminable part of the causal fabric of the physical world, one that can be defined by its causal role, and appears alongside other physical quantities in a unified theory that explains the observable movements of material bodies.xv Even if difficult to detect experimentally, we could treat inducing wave function collapse as a test for the presence of consciousness. And for the physicist, consciousness would now be brought firmly into the realm of physics.

So, again, we have the closure of the problem space. As soon as consciousness makes a difference to the dynamics of material things, it becomes something that matters to physics. But then it also becomes something that is characterizable in terms of its physical role. The physicist needs to think about the human mind insofar as internal processes are part of the causal fabric of the world. The Hard Problem can be ignored because if consciousness enters the problem space of physics, it does so by making a difference to the behavior of physical objects.

**Easy (and interesting) Problems: Is the mind properly characterized in computational or dynamical terms?**

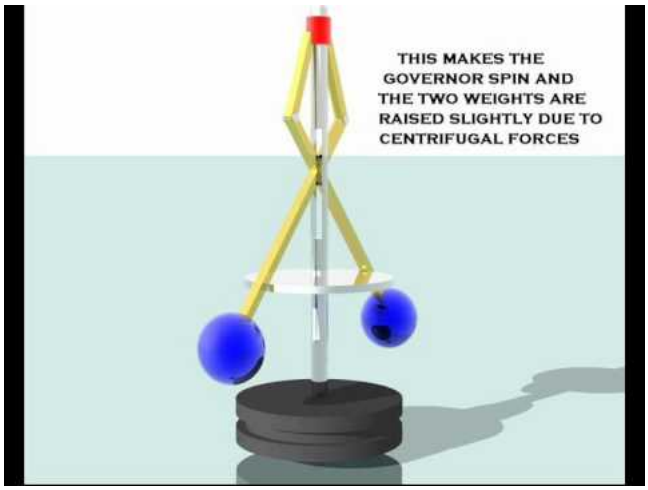
Saying that we can put aside the Hard Problems doesn't mean that the remaining questions about how to fit our mental lives into the general machinery of nature are easy, but there have been some fruitful disputes in cognitive science whose resolution has provided the outlines of a way of thinking about the mind that is helpful from the point of view of someone who is trying to understand how it fits into physics.

One of the disputes is a question about the vocabulary we use to describe the mind. It is the question of whether the mind is properly characterized in computational or dynamical terms.

The answer to this is: both.

There was a time when people thought of the mind just in terms of its conscious part; i.e., in terms of the progression of perceptions, thoughts and feelings of which we are consciously aware, and because those are identified and individuated in representational terms from a first-person perspective, that is the vocabulary we use to describe them. Cognitive science has taught us that there's a lot going on in the mind that falls below the threshold of first-person accessibility, but it continues to describe that activity in terms that come naturally when we are describing our conscious lives: representation, and computation. A spate of challenges in the last couple of decades (starting around 1995) has argued that this whole vocabulary is misplaced who argue that we should use the same straightforwardly dynamical vocabulary to describe the mind that we use to describe planets and pendula.[xvi] Their reasons have to do with specific issues about the role of time and the complexity of certain kinds of causal interactions. The reason that it is an interesting dispute for our purposes is that it forces us to get clear on what the vocabulary of representation and computation is doing, and how it relates to the dynamical vocabulary. To illustrate the difference between the representation-&-computation-based description, on the one hand, and a straightforward dynamical description, on the other van Gelder (who is one of the primary figures advocating the dynamical vocabulary) describes a device whose job is to keep constant the speed of a flywheel to which some machinery is connected. The device is called a Watt Governor, because it was invented by James Watt in 1788. There is a tendency for the speed to fluctuate (because of varying steam pressures and workloads). To smooth things out the amount of steam entering the pistons is controlled by a throttle valve. How might such control be achieved?

1. One way would be to program a device to measure the speed of the flywheel, compare this to some desired speed, measure the steam pressure, calculate any change in pressure needed to maintain the desired speed, adjust the throttle valve accordingly, then start again,
2. Watt's solution was to instead geared a vertical spindle into the flywheel and attach two hinged arms to the spindle. To the end of each arm, attach a metal ball. Link the arms to the throttle valve so that the higher the arms swing out, the less steam is allowed through. As the spindle turns, centrifugal force causes the arms to fly out. The faster it turns, the higher the arms fly out. But this now reduces steam flow, causing the engine to slow down and the arms to fall. This, of course, opens the valve and allows more steam to flow. Properly calibrated, it maintains engine speed smoothly despite wide variations in pressure, workload and so on.[xvii] Here's what it looks like




In (1) that there is (measurement)-computation-action cycle in which the environment is probed, internal representations created, computations performed, and an action selected. In (2), there are no representations (except in a very deflated sense), and no distinct sequence of manipulations to identify with the steps in a computational process. There's just an ongoing process of continuous reciprocal causation in which the Governor (here, the agent) is coupled to the rest of the system (the engine).

The way that we model a system like that is that we write down a state-space for the engine as a whole, and look for a set of differential equations that describe trajectories through that space.[xviii] The space may have some interesting structure (attractors and so on) that we use to understand systemic behavior, but what we don't do is try to pull out one component (the Governor) and describe its exchanges with the rest of the system in terms of representation and computation.[xix]

The claim of those who advocate this sort of description for the mind is that this sort of process is much closer to the true profile of agent-environment interactions than is the traditional vision of a simple perception-computation-action sequence. The kind of interaction that they have in mind is like a baseball player running for a fly ball, keeping visual track of the ball and moving his body in a way that is directly responsive to perceived position, maybe keeping half an eye on what's going on in the bases so that his actions are guided in a particular kind of unmediated way by an ongoing

signal from the environment.

There's a lot to say here, and the cognitive scientific literature contains a lot of helpful and often fascinating details that makes a convincing case that the representation-&-computation style explanation is not particularly well suited to explaining the aspects of cognition that are closely tied to ongoing stimulus. But not all cognitive activity is like that. There is lots of cognitive activity carried out in the absence of any constant, lawful and reliable signal from the local environment. This is where the vocabulary of representation and computation really comes into play.

If we look at biological systems from the very simple, to the more complex, we can see a line of development in which there is an increasing amount of activity between stimulus and response. More and more internal activity decoupled from the environment and designed to support the uptake of information and its use to guide behavior. In the human being large amounts of neural machinery are devoted not to the direct control of action but to the trafficking and routing of information within the brain. If we treat the brain as *just one more factor* in the complex overall web of causal influences, writing down dynamical equations that describe the coupled evolution of agent and environment, we have a single vocabulary that integrates it smoothly into the rest of physics, but we obscure something important.

The key to understanding what we might think of as the *intelligence*-based route to evolutionary success has to do with our ability to exploit information. In systems like human beings, behavior is not keyed directly to an environmental stimulus, but depends on the maintenance of many bodies of information and complex goal structures. Systems in which complex information flow plays a key role tend to exhibit a kind of complex articulation in which behavioral flexibility comes from being able to quickly and cheaply alter the inner flow of information in a wide variety of ways. That articulation is revealed in the computational description, but is (for reasons that I'll talk about below) camouflaged in the purely dynamical description. It would not be wrong to say that the computational description highlights a level of functional organization that explains the point of the low level activity in the brain, and that is crucial to understanding the distinctive kinds of flexibility and control characteristic of truly mindful engagements with the world.xx

So the lesson of this dispute is that to treat the brain as the principal seat of information-processing activity, is to recognize that nature discovered the utility of information long before Silicon Valley, and used that insight to build machines capable of highly complex, flexible behavior. In representing the low-level dynamics of brain and body the dynamical vocabulary is still applicable,

but we get a special kind of insight by seeing how the high-level dynamics guides the flow of information through the mind. The right approach to understanding how the human being fits into physics will be a pragmatic pluralism in which charting the flow of information is *as important* as the low-level dynamics, and in which some *high-level* dynamical features lead a double life as elements in an information-processing economy. It's the information-processing economy that explains the distinctive kinds of behavioral flexibility that human beings exhibit. Actions are no longer spontaneous reactions to stimuli, but temporally extended plans that respond to complex stores of information, with flexible goal structures.[xxi] The way that human experience fits into this is that we're able to identify, at least functional analogues of our conscious mental lives in a way that explains why observation is a source of information about the world, and how human agency (in the guise of action guided by decision procedures that drawn on all of that internally stored information) is possible.

Now, turning to why the 'information-processing economy' is concealed by the dynamical description and that leads us to a second dispute that has led to some helpful insight.

**Is the mind a computer?**

The answer to the now hackneyed question of whether the mind is a computer is: in one sense no, and in one sense yes.

The early uses of the computer analogy in trying to understand the mind emphasized computation (in a particularly narrow sense) and turned out to be limited in a number of ways that probably should have been obvious from the beginning: the formal theory of computation as exemplified in Turing Machine Computationalism is defined only for discrete state machines, and digitality is crucial to many classical results in the theory of computability. Neither of these is a general feature of the kind of information-processing that the brain performs. But there is a less formal notion of computation which is tied to the much more general idea of automated information processing and semantically sensible transitions between representational elements. In that much looser sense, the mind *is* a computer, but the notion of computation doesn't do much work. Almost any physical process can be thought of as a computation in that sense.[xxii]

There is a very different lesson that we can learn from computers, which has much more to do with how the high-level functional organization of a computer relates to the bit –level description of the hardware. A typical modern computer can be thought of as having a state represented by a vector giving the bit-values of all the locations in its memory and in its registers, and all processes in the

computer can be modeled as trajectories through the machine's state space. In practice, software engineers don't think that way at all. They find it more useful to think of various persisting sub-components (strings, arrays, trees, networks, databases, stored programs) as having their own changing states, which interact with one another. In a standard computer, for example, we find multiple databases, procedures and operations, and the information-processing power of the device lies in the fact that these can be rapidly and cheaply reconfigured; much more rapidly and cheaply reconfigured than its mechanical components. This way of parsing the activity isn't imposed; it emerges very naturally from the high-level functional organization that captures the contours of processes in a way that highlights what is crucial to what we use computers for: to process information.

If you asked about the relationship between the high-level functional organization and the bit-level description of the hardware, you might have thought that there must be some sort of discernible correspondence between components and operations at the level of functional organization, and processes defined over bit-level components.

But it doesn't work like that. Computers give us insight into the complex ways in which higher level entities and processes defined over them) can be realized in lower level hard-ware by giving us concrete examples of (running) *virtual machine*s. A virtual machine is a generic world for a functional duplicate of a real or hypothetical machine made not of mechanical parts, but of virtual components. Examples of virtual machines include a word processing system, a simulation of an earthquake, or a growing population, a video game, a calculator, or a chess player. These are all specialized virtual machines that perform specific functions. There are also platform virtual machines (like operating systems) that are capable of supporting many specialized virtual machines. The cool thing about virtual machines is that they can exhibit dynamical behavior, as this variety attests, very different from the physical hardware in which they are implemented. In so doing they show us how layers of structure supports high-level functionality, in a way that defies the expectation of reduction or anything like a simple, visualizable supervenience relationship.

In computers, a combination of hardware and software technology produces of a complex web of causal (or, if you like, virtuo-causal) relationships between elements displayed on your screen when the machine is running. The support for that network of relationships is all of the layers of accreted structure that developed in stages, sometimes over decades. These include a plethora of interacting software or hybrid hardware-software sub-systems, including: schedulers, device drivers, file management systems, memory management systems, compilers, interpreters, interrupt handlers, caches, programmable firmware stores, error-correcting memory, wired and wireless network

interfaces, network protocol handlers, email systems, web browsers, and many more. People added to these structures and built on top of platforms once they were in place, without knowing how to relate the structures they built to what was below. The relationship between a running virtual machine and the physical machine on which it is implemented may be no more transparent than all of these levels of accreted structure. And there won't typically be a discernible structural correspondence the two. There won't be a fixed correspondence between components of the virtual machine and components of the physical machine (files stored by a word-program, for instance, won't have a fixed location in computer memory). There won't be processes, available on inspection of the hardware that look like the processes executed by the machine. xxiii The structure of the virtual machine, moreover, can change significantly without structural changes occurring at the physical level though the physical states of millions of switches may need to change to alter conditional connections. Indeed, that is what explains the power of these devices to quickly and cheaply alter the flow of information.

Evolution has had far more time to discover these cheap and flexible ways of routing information. And it has also had more time to build up layers of virtual machines running on virtual machines. If one is thinking in these terms, it is natural to think of our own experience as a kind of high level control interface with no direct structural correspondence to anything that goes on in the brain.

The suggestion here – which I take from Aaron Sloman, but which is implicit in the very familiar idea that the mind is the software of the brain – is that the right kind of functionalism is virtual machine functionalism.xxiv And the right way to think of the relationship between mental activity and brain activity is implementation: not reduction, not correspondence, none of the much simpler, visualizable relationships that philosophers have sometimes thought have to hold if the mind is part of the physical world. It's a relationship of vastly more complexity than the imagination by itself can penetrate.

**Details of implementation**

Once we have introduced the notion of a virtual machine, we can speak freely of mental processes in the vocabulary that comes naturally, treating as a machine that processes and stores information, bringing it to bear on the determination of behavior, integrate mind into the dynamics in a rather smooth fashion, because it speaks the language of mechanisms, and dynamics. xxv , So conceived, the mind becomes both an object of empirical investigation and one that is integrated into physics, so that we can understand how observation and reason-guided action fit into the larger dynamical framework. But the difficult details of implementation can be left aside to be

sorted out by others, which would be bound to be enormously complex, and tend to obscure the high level functional relationships that matter. It was the information-processing economy that was selected for, because it is the architecture that determines the causal-informational exchanges with the environment.

**In Sum**

Here, then, is an opinionated list of the progress that has been made about either resolving or sidelining some of the disputes in the literature:

- The dispute between dynamical and computational theories of mind is resolved by separating aspects of cognition that are closely tied to we have an event stream, and a mind that is 'wired into', or coupled to, the event stream, in such a way that it is receiving information from the stream, processing it, and then intervening in a way that produces sensory feedback. The more or less continuous processes that involve continuous, reciprocal causation and that are well-described by dynamical systems approaches (e.g., the football player running for a long pass), and the more cognitive internal processes that are less tightly coupled, more sequential, and better modeled by talking about information, representation, and computation. Real-time perceptual input from the environment (running for a fly-ball) from those that are more
- The dispute about whether the mind is a computer is resolved by distinguishing different notions of computation, saying in some senses it is not, and in some senses it is.
- How do we map talk of representations and computation into a causal map of the brain? This is where computers can shed the most light by defusing very simple ideas about how the information-processing economy of the mind is related to the low level dynamical description.

Here is a list of what has not been resolved, but can be isolated (I think) from the ways in which human experience should matter to physics

- The so called Hard Problem of consciousness;
- The analogous Hard Problem of intentionality;
- Details about implementation that don't matter for architecture.

---

i Although sometimes used in a narrow way to refer to sensations or perceptual offerings, 'experience' is used here to mean the full introspectively accessible mental life.

ii "Completeness, Supervenience, and Ontology", Tim W E Maudlin 2007 *J. Phys. A: Math. Theor.* **40** 3151. Ismael, J., "Do You *See* Space? How to Recover the Visible and Tangible Reality of Space (Without Space)", *Philosophy Beyond Space-time 2*, edited by Christian Wuthrich and Nick Huggett, Oxford University Press, forthcoming.

iii More precisely between skin and whatever features of the brain support perceptual awareness.

iv In cosmology the questions about the relations between what our theories say about the world and the experience of the observer are coming under pressure as well, because of a disparity between the scope of the theory and the information that is even in principle available from within space-time. The issues here treat the observer as a generic embedded system. The mind doesn't enter in a specific way.
v Chalmers, D. 1995. "Facing up to the problem of consciousness". *Journal of Consciousness Studies*, 2: 200–19, Chalmers, D. 1996. *The Conscious Mind*. Oxford: Oxford University Press.

vi There are six canonical arguments. Some of them deal with epistemic possibility (the Open Question argument, Frank Jackson's Black and White Mary argument), some with metaphysical possibility (the zombie argument, the modal argument). The literature surrounding them is now enormous. See also Chalmers "Consciousness and Its Place in Nature" d in (S. Stich and F. Warfield, eds.) *Blackwell Guide to the Philosophy of Mind* (Blackwell, 2003), and for an overview of the wider literature (with bibliography) see Van Gulick, Robert, "Consciousness", *The Stanford Encyclopedia of Philosophy* (Spring 2018 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/spr2018/entries/consciousness/>. I have also written about Chalmers' arguments (The Situated Self, Oxford University Press, 2006/7, particularly Part II). I don't think that anybody has a satisfying answer to the arguments, and they may not be not answerable in the terms that proponents demand, but it is better to go around the problem through it. It would be a very bad outcome if science were diverted from pulling the mind into its own subject matter because of the deadlock of the philosophical debate.

vii Our own Conscious states can be *associated* with brain states identified in virtue of the fact that they play a certain functional role, and thereby *associated* with states that are uncontroversially physical and integrated into the causal web. The arguments are supposed to establish that the association cannot be one of identity since the functional organization can be reproduced in a non-conscious system. So there is agreement that our conscious states can be picked out by extensional definition that lets us associate them with brain states. The issue for proponents of the Hard Problem specifically concerns claims of identity between consciousness and any proposed physical basis.
viii And to say that (again) is not to agree that that is what phenomenal consciousness is. It is just to say that we can ignore the philosophical debate, because proponents of the Hard Problem succeed in establishing that consciousness is not understandable in physical terms, only at the expense of making it irrelevant to physics.

ix At least in the early articles, Chalmers agrees. Defending his own view that there should be a devoted science of consciousness that looks for psychophysical laws he writes:
"Certain features of the world need to be taken as fundamental by any scientific theory. A theory of matter can still explain allsorts of facts about matter, by showing how they are consequences of the basic laws. The same goes for a theory of experience. *This position qualifies as a variety of dualism, as it postulates basic properties over and above the properties invoked by physics.*" Emphasis mine.

xAgain, there is a large literature. Discussion occurs in the disputes surrounding the naturalization of content. See Adams, Fred and Aizawa, Ken, "Causal Theories of Mental Content", *The Stanford Encyclopedia of Philosophy* (Summer 2017 Edition), Edward N. Zalta (ed.), URL = https://plato.stanford.edu/archives/sum2017/entries/content-causal/ for an overview and bibliography.

xi Note that this isn't behaviorism. Behaviorists held that behavior could be explained by reflex and conditioning. This view recognizes a rich internal life of perception, thought, feelings and volitions. It is simply agnostic about any purported feature of mental life that makes no even potential detectable impact on behavior.

xii"Why Mind Matters in Quantum Mechanics", http://philsci-archive.pitt.edu/15910/. See also Gao, S. (2017). The measurement problem revisited. Synthese, DOI 10.1007/s11229-

017-1476-y. and Oldofredi, A. (2019). Some remarks on the mentalistic reformulation of the measurement problem: a reply to S. Gao. Synthese, https://doi.org/10.1007/s11229-019-02101-3.

xiiiThere are different brands of supervenience, corresponding to different strengths of modal connection between a brain state and its phenomenal properties. There is analytic entailment, metaphysical necessity, even the dualist's special brand of psychophysical necessity. The fine philosophical distinctions between these different brands of necessity can be more or less ignored by the physicist who cares about dynamical roles. Any form of necessary link between the conscious state and its physical basis will make the conscious state and physical basis interchangeable from the physicist's point of view. The most interesting discussions of consciousness in the physics literature are avowedly physicalist and suggest that the explanatory direction might go from physics to consciousness. Penrose, for example, has suggested that the right account of conscious thought might come from consideration of the quantum properties of matter and the way such properties help us understand how deterministic but non-computable state transitions might occur in real physical systems Penrose, R. (1994). Shadows of the Mind. Oxford: Oxford University Press. Hameroff, S. R. & Penrose, R. 1996b Conscious events as orchestrated spacetime selections. J. Conscious. Stud. 3, 36–53.

xivThis is for a variety of reasons. In its original formulation, the proposal was vague. Which systems exactly are associated with consciousness, and exactly when does it enter the dynamics and induce collapse? As Bell famously asked: "Was the wave function of the world waiting to jump for thousands of millions of years until a single-celled living creature appeared? Or did it have to wait a little longer, for some better qualified system ... with a PhD?" (Bell, J. S. (1990), "Against "measurement"," *Physics World* 3 (8): 33–40. Reprinted in Bell (2004), 213–231., 34). Most importantly, however, there is increasing evidence that no collapse occurs. See, for example, K. G. Johnson, J. D. Wong-Campos, B. Neyenhuis, J. Mizrahi and C.

Monroe (2017). Ultrafast Creation of Large Schr dinger Cat States of an Atom," Nature Communications 8, article number 697;doi: 10.1038/s41467-017-00682-6, for recent developments.

xv And there's the question of whether it is even properly thought of as dualism. Chalmers and Kelvin McQueen have recently revived this sort of view, and Chalmers himself seems to continue to regard it as a dualist view, though I'm puzzled why, in light of his given reasons for calling his original position dualist. He writes, "Certain features of the world need to be taken as fundamental by any scientific theory. A theory of matter can still explain all sorts of facts about matter, by showing how they are consequences of the basic laws. The same goes for a theory of experience. *This position qualifies as a variety of dualism, as it postulates basic properties over and above the properties invoked by physics.*" Emphasis mine. (op.cit. 1995)

xvi See van Gelder, T. (1995). What Might Cognition Be, If Not Computation? Journal of Philosophy, XCII (7), 345-381, van Gelder, T., & Port, R. (1995). It's About Time: An Overview of the Dynamical Approach to Cognition. In R. Port & T. v. Gelder (Eds.), Mind as Motion: Explorations in the Dynamics of Cognition (pp. 1-44). Cambridge, MA: MIT Press. Related arguments are found in Thelen, E., & Smith, L. (1994). A Dynamic Systems Approach to the Development of Cognition and Action. Cambridge, MA: MIT Press, Kelso, S. (1995) Dynamic Patterns Cambridge, MA: MIT Press, Varela, F., Thompson, E., & Rosch, E. (1991). The Embodied Mind. Cambridge, MA: MIT Press and in Wheeler, M. (1994). From Activation to Activity. Artificial Intelligence and the Simulation of Behavior (AISB) Quarterly, 87, 36-42. Beer, R., & Gallagher, J. C. (1992). Evolving dynamical neural networks for adaptive behavior. Adaptive Behavior, 1, 91-122, Wheeler, M. *op. cit.* p.36-42. For discussion see Keijzer, F. & Bem, S. (1996). Behavioral Systems Interpreted as Autonomous Agents and as Coupled Dynamical Systems: A Criticism. Philosophical Psychology, 9, 323-46, Clark & Toribio *op. cit.* and Clark, A. & Grush, R. (To appear). Towards a Cognitive Robotics. Adaptive Behavior.

xvii Van Gelder, *op. cit.*, p. 347-350

xviii We could use representational vocabulary to describe this only if 'representation' meant something like 'information-bearing state', in which case the representational description would add nothing to the

dynamical description.

xix Total state explanation for any dynamical model that emphasizes the fact that all aspects of a system are changing simultaneously and invites us to understand the behavior of the system in terms of the possible sequences of changes in total state over time. Trajectories through state spaces populated by attractors, repellors and so on reflect motion in a space of total states, i.e., states that *assign* values to all systemic variables and parameters.

xx Where we confront especially complex interactive causal webs, however, it does indeed become harder to isolate the syntactic vehicles required by the computational approach.

xxi Natural functional interpretations of properties such as rationality and intentionality will then emerge if we get our designs right. Consciousness might not have a single functional interpretation, but a collection of related notions (introspective accessibility, global broadcast, informational integration). We won't have solutions to the Hard Problems, but we will have everything we need in order to address questions about observation and action as they appear in the problem space of physics.

xxii A common worry about the proposal to recognize analog computation is that everything then turns out to be *some* kind of computer and hence the thesis that the brain computes is empty. The response is that we can narrow the focus by looking at the special class of computations whose implementations support flexible behavior and reason-guided action. And we can draw on the intuitive notion of computation as semantically sensible state-transitions without insisting that the notion of computation needs to do anything more than heuristic work.

xxiii Where there are no fixed physical correlates for virtual machine entities, the processes that operate on them will not be recognizable in the low-level dynamics. Detection will be possible where there are more or less reliable physical correlates.

xxiv Sloman, A. (2010), 'Phenomenal and Access Consciousness and the "Hard" Problem: A View from the Designer Stance', *Int. J. Of Machine Consciousness* 2(1), 117–169. Sloman, A. (2002), Architecture-based conceptions of mind, *in* P. Gardenfors, K. Kijania-Placek & J. Wolenski, eds, 'In the Scope of Logic, Methodology, and Philosophy of Science (Vol II)', Synthese Library Vol. 316, Kluwer, Dordrecht, pp. 403–427. Sloman, A. and Chrisley, R., "Virtual Machines and Consciousness, In *Journal of Consciousness Studies*, 10, No. 4-5, 2003. The idea was and which was also developed by John Pollock. "What Am I? Virtual Machines and the Mind/Body Problem", Philosophy and Phenomenological Research, Vol. 76, No. 2 (Mar., 2008), pp. 237-309

xxv And the hope is that natural functional interpretations of properties such as rationality and intentionality will then emerge if we get our designs right. The pre-theoretic notion of consciousness might separate into a collection of functionally definable notions (introspective accessibility, global broadcast, informational integration).