Chapter 12
The Mind-Body Problem

'O blithe little soul, thou, flitting away,
Guest and comrade of this my clay,
Whither now goest thou, to what place
Bare and ghastly and without grace?
Nor, as thy wont was, joke and play.'
Hadrian, on his deathbed (Translated by A. O'Brien-Moore)

12.1 Introduction

What is the relation between the body and the soul, or mind? Is the mind something that exists independently of the body? Or are the body, specifically the brain and nervous system, and the mind merely two different descriptions of the same thing? Our common sense understanding, which is riddled with religious influence, leans towards the first alternative. But then what is the mind? How do the mind and the body interact, if they are in fact fundamentally different things?

If one assumes the former position, one is a dualist and if one assumes a variant of the latter position, one is a monist. There are many versions of monism but basically only two versions of dualism, substance dualism and property dualism.

Dualists and monists face completely different requirements for explanation. If one believes that body and mind are not two different things, but rather one and the same, then what is wanted is an explanation of how mental processes occur in the brain. Such an explanation requires a reduction of the mental to the physical. There are several examples of successful reductions that illuminate how this requirement might be met in this case. One such example is the reduction of genes, hypothesized by Mendel as the carrier of inherited traits. Originally, this concept was a purely theoretical construction. After the discovery of the DNA molecule in the 1950s, one was able to guess that genes were parts of nitrogenous base sequences in the DNA molecule. This was latter confirmed when specific gene sequences were identified along with their phenotypic expressions. As this is an on-going investigation, the details of the reduction are still to be worked out; but one can justly say that the problem of reducing phenotypic traits to heritable genes, now conceived as sequences in DNA molecules, has been solved in principle.

The monists think that an analogous reduction of mental phenomena to physical processes in the brain is at least in principle possible, while others hold that

L.-G. Johansson, *Philosophy of Science for Scientists*, Springer Undergraduate Texts in Philosophy, DOI 10.1007/978-3-319-26551-3_12

reduction, in the sense of identification types of mental states with types of physiological states, is impossible. But this rejection of the possibility of reduction must not be taken as an argument for dualism. One position, eliminative materialism, holds that there is no scientific role for mental phenomena to play; hence they can be eliminated from a scientific description of the human being. The argument justifying this elimination is based on logical and semantic analyses of the reduction concept, as well as analyses of the concept of a mental state.

Nowadays, all monists are materialists; that is, people who think that external reality is material in nature (The concept 'material' should be understood here in terms of modern physics. Since energy and matter are equivalent, according to the well-known formula $E = mc^2$, even energy belongs to the material world.). The opposite thesis–that external reality is mental in nature–is also a form of monism. This latter view is very unusual in recent times, but has had some significant proponents, for instance, Leibniz and Berkeley.

Dualists face a completely different explanatory task; they must explain how it is that the body and the mind can interact. All hitherto known theories of how objects interact are strictly physical in that they describe how physical objects interact via one of the four fundamental natural forces, gravitation, electromagnetism, weak and strong nuclear interaction (Materialism is nowadays identified as the thesis that ultimately all that exists are physical in nature.). If the mind is a non-material object, then no purely physical theory about its interaction with the body can be given. Indeed, the idea that the mind is non-physical, together with the observation that bodily and mental states can causally interact with each other, conflicts with a basic tenet of our scientific thinking, namely, that the physical world is causally closed. Nothing non-physical, if such a thing exists, can influence something physical (For example, abstract things like numbers are non-material things and they cannot possibly be causally related to anything in the physical world, in fact not to anything whatsoever.). This idea is seldom expressed explicitly, probably because most take it to be obvious and without need of conscious expression; but nevertheless, it is fair to say that it is a firmly held tenet of the sciences.

12.2 Substance Dualism

Substance Dualism is the theory that there are two types of substances, or entities: bodies and minds. Property Dualism – an alternative dualist theory most notably championed by Popper[1] and David Chalmers[2] – stands in marked contrast to substance dualism in insisting that there is only one type of substance in the

[1] Karl Popper, 1902–1994, was born in Austria but worked in England beginning in 1945. His central works falls within the philosophy of science (see Sect. 12.2).

[2] Chalmers, D. (1996). *The conscious mind. In search for a fundamental theory*. Oxford: Oxford University Press.

world, matter. Property dualism deserves its name because this single substance has two irreducible types of properties: physical (including chemical and biological) and mental. I shall discuss the latter view in the next section.

Descartes proposed the most well-known substance dualistic theory.[3] He claimed that man is constituted of matter and soul, two different types of substances, each having a fundamental property, or essence. The essence of bodies is extension, and the essence of minds is thought.

Descartes' explanatory task was to show how mind and body interact. Starting from the new mechanical world-view of his time, that all motion is the result of mechanical interaction, push and pull, which occurs when portions of matter, bodies with extension, come into contact. In other words, only things with extension can put a piece of matter into motion. Yet it seems obvious that many of our own bodily motions are caused by our desires and thoughts, our mental processes in our soul, and our soul lacks extension.

Descartes' solution to this problem is to state that the soul is located in the pineal gland and that mental processes set the pineal gland into motion. The motion of this gland is forwarded to the rest of the body via the so-called 'animal spirits', a viscous fluid located in the nerves (Descartes, like all others at the time, believed that the nerves were hollow and filled with 'animal spirits'.).

But why would the soul set just this particular gland into motion? Descartes had two arguments for this. Firstly, since we have only one soul, it must be located in a central part of the brain that has no duplicate, i.e. the pineal gland. Secondly, there is a cavity surrounding the pineal gland, which explains, according to Descartes, how the pineal gland could stimulate motion in the body via the fluid, the animal spirits, that he believed filled both this cavity and the nerves.

A moment of reflection tells us that Descartes' explanation breaks with the principles he himself held: if the mind can set the pineal gland in motion, then there is an exception to the principle that all bodily motion is caused by purely mechanical interactions with other bodies. Thus his theory is self-contradictory.

There are still some people who believe in some version of substance dualism, but none has been able to explain how interactions between mind and matter work.

12.3 Property Dualism

According to Popper, there are three realms: World 1, World 2 and World 3. World 1 is the material world of bodies and physical objects, and World 2 is the matter-independent world of mental processes. World 3 contains such abstract objects as

[3] René Descartes, 1596–1650, was a French philosopher who spent his last year in Sweden working for Queen Kristina. He broke with the scholastic tradition of philosophy, using systematic doubt as a method for arriving at knowledge.

scientific theories, musical works, novels, etc., i.e., cultural products that cannot be identified with physical objects.

According to Popper, entities in worlds 2 and 3 have properties that cannot be explained or reduced to objects and properties in world 1. When a material system becomes sufficiently complicated, new mental, non-material properties arise. Popper has not given any explanation for how or why they arise in a sufficiently complicated system and, without such an explanation, this theory seems to create more problems than it solves.

David Chalmers has in his (1996) proposed a somewhat similar theory. He argues that the mental cannot be reduced to the physical; that is, human (and perhaps even animal) consciousness is a radically different kind of property than the physical properties used to describe the brain. What we need is a completely different kind of theory, one that makes room for both physical and mental properties. Chalmers does not claim to have such a theory. Rather, he claims only that all previous attempts at reduction have failed, and must fail. Thus the only way out is to postulate two fundamentally different types of properties as part of an as yet unknown true theory. The first problem that Chalmers focuses on, in regards to this endeavour, is the so-called 'problem of qualia', which we shall discuss in Sect. 12.6.

12.4 Monism

The principal difficulties of dualism have resulted in that virtually all present-day philosophers are monists as regards the mind-body problem. However, this does not mean that the philosophical debate surrounding this issue has ceased, quite the contrary; there is a great diversity of views within monism. Neither does this mean that one denies the existence of mental properties; rather, it is typically thought that mental properties are, in some sense, dependent on physical properties. Though, some do go further and deny the existence of mental properties. In any case, the least common dominator of monism can be characterized by the following three theses:

- *Supervenience:* If two objects have exactly the same physical properties, then they have exactly the same mental properties (But not vice versa.).[4]
- *Anti-Cartesianism:* Nothing can have only mental properties.
- The physical world is *causally closed*. If x is a physical event, and y is either a cause or an effect of x, then y is also a physical event.

Comments and Illustrations

The first principle may seem empty, since it is not likely that there exist two objects that have exactly the same physical properties. Yet, even if this is the case, it would

[4] There are also other characterisations of supervenience, but this will do for our purposes.

not constitute an objection, since if we reformulate the thesis to its contrapositive, we get the following: if two objects differ in mental properties, then they also differ in physical properties. Even more relevant is the application of this thesis to successive states of a single person: if a person experiences a mental change, then she has undergone a physical change. If, for example, a person were to change mental state by having a new thought, then this person must also undergo a physical change. This is what it means for mental properties to supervene upon physical properties. Supervenience also seems to presume that there are properties, but this is incorrect: those who generally deny the existence of properties, nominalists, can express the supervenience doctrine in terms of classification of states using physical and mental predicates, without assuming that predicates refer to properties.

Anti-cartesianism is the thesis that there doesn't exist any purely mental objects. It implies, among other things, that the soul cannot live after the body dies. It also conflicts with the idea of reincarnation; that a mind, or soul, with certain properties can transfer from one body to another.

At first glance, the physical world's causal closure may seem to conflict with everyday observation: is it not obvious that the physical events we observe give rise to sensations and thoughts? Do not physical events appear to have mental effects? This would be a strong argument if one presupposes that sensations, emotions and thoughts have mental properties only. But if one accepts supervenience, then this argument fails, since the act of having a sensation, emotion or thought is to undergo a change of mental state, and that is a change of physical state. Thus the question of how physical and mental changes relate to one another is a question about the details of monism, for which there are quite different views.

12.5 Monistic Theories

1. Chronologically, the first serious monistic theory was *identity theory*, first proposed by J. J. C. Smart in 1959. According to this theory, when we talk about a mental state, what we are actually talking about is a certain physiological brain state. For example, feeling happy, or perceiving the sky as blue, is nothing more than the brain being in a certain state. Thus the mind-body problem will be solved once we have matched up every mental state with its corresponding physiological brain state. A historic parallel from astronomy will help illustrate how empirical research can result in propositions of identity.

   In oldest times it was believed that the Evening star and the Morning star were two different stars, but the Babylonians discovered that they were actually one and the same planet, Venus. The expressions 'Morning Star' and 'Evening Star' thus denote the same thing. Analogously, identity theory monists hold mental and physical predicates to describe the same events.

   There are two variations of identity theory: *type identity* and *token identity*. Type identity states that for every *type* of mental state there is a neurological/physiological *type of state* that is identical to that type of mental state. This

means that every time I have the sensation 'seeing that the sky is blue' I am in the same type of brain state. Token identity claims that every token of a mental state is identical to some token of a brain state, but it is not always the case that e.g. all instances of 'seeing that the sky is blue' corresponds to exactly the same type of brain state. Thus there can be identity at the token level without there being identity at the type level. Another way of expressing this is the following: physical and mental concepts classify phenomena is different ways, though they do classify the same phenomena; namely, the neurological states of a person.

An influential theory that further develops token identity is Donald Davidson's[5] anomalous monism, but more on this in Sect. 10.7.

2. *Functionalism*. Functionalism was once proposed by Hilary Putnam[6] and many philosophers and researchers in artificial intelligence have followed suit. According to functionalism, a mental term does not refer to any definite event or neural process; rather, it indicates a place in a functional system. Thus mental terms refer to functions in the central nervous system, functions that are carried out by processes in the brain. However, mental terms do not describe these processes in terms of which neurons are active or how they interact, but rather in terms of their input and output signals and what they do; their function. Consequently, the main difference between the functional and identity theories is that a certain function can be executed by different neurons in different ways.

   One can say that the concept of a function and the concept of a process of the brain are both used to describe a certain part of reality, that which goes on inside our heads, but that these concepts classify this part of reality in different ways. Concepts such as brain processes and brain states classify using physiological criteria, whereas functions apply functional criteria. The following example illustrates this point.

   Assume that I believe it to be cold outside. This belief makes it so that I react to a certain input in a certain way. If someone were to ask 'how is the weather?' (input), I would answer 'it is cold' (output). A different output arises if the circumstances are different. If, for some reason, I were to go on a walk, then my belief that it is cold will result in my putting on warm clothes before I go out. Saying about someone that she believes 'It is cold outside.' is a description, in mental terms, of a large set of such input–output functions realized in the body of this person.

   Functionalists do not deny that these input–output functions are performed by nerve cells in the brain, but the state we describe with the expression 'believing that it is cold outside' does not describe any physiologically defined type of brain state. It is fully possible that completely different sequences of states can bring about a connection between an input of the type 'someone asks about the

[5] Donald Davidson: *Essays on Actions and Events*. Clarendon Press, 1980, especially essays 11 and 12.

[6] Putnam (1960, 1967).

weather' and an output of the type 'putting warm clothes on when you go out'. This implies that every system (e.g. a computer) that reacts similarly to a human with respect to some stimuli is, by definition, in the same state of consciousness as a human (Hence research in AI, artificial intelligence, is heavily influenced by functionalist views on the mind.).

A number of versions of functionalism have emerged, what is given above is only a first rough description. For a thorough overview of the different positions, arguments and problems, see http://plato.stanford.edu/entries/functionalism/

3. *Eliminative materialism*. According to this view, propounded by e.g. Paul and Patricia Churchland, our mental vocabulary is misleading, because it presupposes that there exist such entities as sensations, feelings, and beliefs. In other words, use of mental concepts makes us think that they refer to well-defined objects or processes, and this need not be the case. Once we have achieved a complete science of consciousness, it will contain no such concepts, but rather other, scientifically defined concepts. That is to say, we would have eliminated our traditional psychological vocabulary. This theory can be considered a variation of token identity theory, if we add the additional assumption that the everyday mental vocabulary can be eliminated from a real theory.

   Another way of expressing this is the following: if we assume that a reduction (to be explained in the next section) of mental properties to physiological properties is, in principle, impossible, then we will have, if and when this reduction is performed, a number of identity statements of the form 'mental state description M and physiological state description F denote the same state'. These identity statements constitute a sort of psychophysical law. Eliminative materialism denies that such laws exist because the mental state description is empty.

12.6 Three Important Problems for Reductionists

Reductionism in general comprises of at least two theses: (i) concepts in the reduced theory should be defined in terms of concepts from the reduction basis, and (ii) that laws in the reduced theory can be inferred from the base theory. These conditions do not follow from the three monistic principles supervenience, anti-cartesianism and causal closure, thus allowing one to deny that reduction of the mental to the physical is possible and still be a monist. But reductionism about the mental faces severe problems, which is the reason why quite a few philosophers accept monism but are sceptical about reductionism.

Arguments against reductionism are usually based on three purported distinctive features of the mental that are held to raise profound difficulties for reduction. These features are intentionality, i.e., the property that mental acts are directed towards objects (see Sect. 5.2), subjectivity and qualia.

(i) *Intentionality*. Brentano defined the mental sphere in terms of intentionality. All mental phenomena are intentional, whereas no non-mental phenomena are

intentional; intentionality is the distinctive trait of mental phenomena, according to Brentano. As explained in Sect. 5.2, a mental phenomenon is constituted by two aspects, an *act*, which is directed towards an *act-object*; intentionality is this relation of directedness from the act to the act-object. A perception of an object is directed toward the object observed. When we think about a thing, person, event or theory, this thought is directed towards these concrete or abstract objects. When we are mad at someone or something, we feel *about* that person or thing. In all of these cases, the intentionality is indicated by the preposition that points to just that at which the mental process is directed.[7] Indeed, most, if not all, mental processes have this intentional aspect, whereas no material things or processes do.

Possible exceptions to the rule are moods such as euphoria, depression and anxiety. These states are similar to certain emotions, and yet they lack the directedness of those emotions. But one could defend Brentano's stance by saying that these states are not really mental states, they are physiological states, which can be clearly identified by physiological criteria. However, for the present discussion we need not decide the matter.

But how can one reduce the mental to the material? Given that the act of reduction removes an important property of the original phenomenon, must not all forms of reduction involve loss, and so be destined for failure?

Classical reductionism is the thesis that a successful reduction consists in an analysis of the semantics of mental expressions in terms of non-mental categories. For example, functionalists think that the mental state of believing 'It will be snow tomorrow.' cannot be analysed into two separate parts – the act of believing and the content of that belief – but rather that the state description, in its totality, refers to a certain functional state. A certain mental act, such as the expectation that the weather will be nice tomorrow, is a certain functional state, but the same act with different content, e.g. the belief that it will rain tomorrow, is obviously a different mental state. There need not be anything in common, indicated by the psychological word 'believe', to these two states. The intentionality, i.e. the act-object structure, here has, so to speak, disappeared from the analysis. Whether this is a failure or a success of the analysis is debatable.

(ii) Another aspect that disappears in the process of reduction is the purported qualitative aspect of (but not limited to) experiences. When I see a red tomato, my experience of that tomato contains a special chromatic quality, which I recognize and describe as 'the experience of the colour tomato-red'. This special experiential quality seems to be missing when we describe the experience in physical terms. We can perhaps achieve a reduction of the sensation

[7] There is a profound difficulty with this theory, viz., how to analyse the relation act – object in cases of non-existing objects, such as Santa Claus. This difficulty inspired Husserl and Frege to develop their theories in quite distinct ways, which gave rise to the division between 'continental' and 'analytic' philosophy.

of a smell, taste or vision by identifying it with a chain of signals in the brain. For example, we know that the sense of taste operates with five basic tastes (sour, salty, sweet, bitter and umami) and that all tastes can be described as combinations of these components in various magnitudes. We can thus represent a certain taste using a quintuple. For example, the quintuple <1, 4, 7, 3, 8> can be interpreted as '1 part sour, 4 parts salty, 7 parts sweet, 3 parts bitter and 8 parts umami'. It is even possible to identify such quintuples with signals in the part of the brain responsible for the sense of smell. However, the purely qualitative aspect of the sensation, how it *feels* to smell wild strawberries, or how it feels to see red cherries (not the associations we often attribute to the sensation, but the sensation itself), also seems to disappear in the process of reduction, regardless of which form we apply.

We should note that the qualitative aspect plays no role in the use of qualitative words, such as 'red' or 'the smell of wild strawberries'. Each child must learn her native language by interacting with competent language users and the only way to learn the correct use of for example the word 'red' is to observe the situations in which adults indicate the correct and incorrect use of that word through their behaviour. This requires that the child be able to identify the common perceptual traits – e.g., the presence of red objects – in all situations where the word is used correctly. In turn, this requires that the child's is able to compare an actual sensory impression with the memory of past sensory impressions. Thus it is necessary for learning the correct usage of certain words that we experience qualitative differences and similarities between various sensations, and that we compare past and present sensations. The child must be able to remember sensations of colour hues and being able to classify a present sensation as more like one rather than other remembered colour experiences.

However, this is obviously a process that occurs within a single person; there is nothing that guarantees that two people have the same, or even similar, experience when they see a red flower. The only requirement for successful verbal communication is that different people apply category words similarly. Whether different individual's qualitative experiences differ makes no difference as regards their classification.

Many philosophers have taken for granted that humans have similar, qualitative experiences, but such philosophers are immediately faced with the question of what is meant by 'similar' in this context. In order to meaningfully say that different person's experiences are similar, or not, we must have a similarity measure that can be applied interpersonally. But such cannot in principle be had, since we are things that by definition are irreducibly subjective.

If we were to formulate these criteria in terms of the ability to report the same thing and make similar distinctions in language, then we would have transformed the qualitative aspect into a matter of behaviour, which is, as we saw above, something different than the experience itself. But how else should we formulate a criterion determining whether two individuals have

qualitatively similar subjective experiences? If one then abandons the idea of giving any identity criteria, one also abandons the idea that qualitative experiences are comparable in any way.

However, we might be able to say *something* about qualia; namely, we can typically distinguish between cases where qualia are present and cases where they are not. We can easily imagine two people eating wild strawberries, where one of which has a nasty cold. One of them says to other, 'these really taste like wild strawberries', and other responds, 'sorry, but they don't taste of anything at all'.

Is it conceivable that a robot without consciousness could behave in a fashion similar to the people in the above example? Can we conceive of a robot that could carry on a conversation with humans, behave like humans, at least in some situations, without having any sort of consciousness, or self-awareness? Assume that we have succeeded in constructing a prosthetic tongue and nose, which are wired to a computer. These artificial organs can identify tastes and smells almost as well as a human (Suppose that we know a great deal about the senses of taste and smell.). Signals are forwarded to the computer, which processes the information through subroutines, thus generating applicable statements that are sent to the speech organ for communication. Nothing in this example seems impossible; and yet there is nothing that indicates that the computer truly feels the taste of the wild strawberries. Thus in the above example we have a robot that behaves like a human without having consciousness. It is not difficult to generalize this example to the entire spectrum of human behaviour. Doing so would give us a zombie, a robot that behaves like a human but is devoid of an inner life. According to critics of reductionism, this shows that any such reduction looses the most essential aspect of conscious processes: consciousness itself.

Eliminative materialists reply that if we can give an explanation of human behaviour without needing to take into account that some part of our behaviour is intimately related to conscious sensation and thought, then this aspect is irrelevant to scientific theories about consciousness.

The functionalists, on their part, stress that if consciousness has a function, then it can be described in terms of certain input–output functions, which are active when we are conscious and not otherwise. Furthermore, it is the reductionist's goal to identify these functions. The purely qualitative aspect of conscious states, 'how it feels. . .', is irrelevant in both cases.

(iii) All mental states are subjective in the sense that no one other than the person experiencing that state has direct/introspective knowledge of them. Contrarily, all descriptions of brain states are public, since any observer with the requisite knowledge and training can observe them. However, this leads to problems for the identification of mental states with brain states. As regards identity theory, the problem is quite clear. If two descriptions–one in mental terms and another in neuro-physical terms–describe one and the same phenomenon, then this phenomenon must have the same properties regardless of the description, otherwise the two descriptions do not describe the same thing. However, if

> something is public then it is not private, and vice versa. Thus all mental phenomena have a property that all neuro-physical phenomena lack.

In my view, the most serious of the three problems for reductionism is intentionality. Mental directedness appears to be an essential and fundamental property of mental activities. Every monistic theory must therefore be able to say something about what this intentionality amounts to in physiological terms, or else convince us that the act-object distinction is merely a peculiarity of our grammar.

The subjective nature of mental activities does not seem to be as fundamental, or at least not in the same way. Suppose that the strong form of (type) identity theory were true (I do not think that this is particularly likely, but the point is to show that subjectivity is a contingent fact.). Assume further that we have succeeded in pinpointing exactly which systems in the brain are active when we have a specific sensation. One would then be able to know whether or not some person has this sensation by measuring the activity in these systems using an MRI machine for example. This would mean that our states of consciousness are no longer subjective, but rather fully accessible to physiological study.

In regards to the problem of qualia, I doubt it is a real problem for reductionists. Science aims to describe and explain phenomena in objective, i.e., intersubjective, terms. Thus giving a scientific characterization of what the purely qualitative aspect is, seems to conflict with the basic point of science. The purely qualitative content, the 'subjective feel', of the conscious experience, cannot as a matter of principle be described. Why so? Because using words, language, is essentially an intersubjective activity and the word 'qualia' is intended to refer to that which is not intersubjective, but purely subjective. But that is impossible, so the word 'quale' cannot have any reference. Hence, not even in the mental sphere is there anything which we refer to by using the term 'qualia'.

Wittgenstein's discussion of these matters in his *Philosophical Investigations* will perhaps illuminate this issue. In § 304, he discusses pain, or more precisely, the experience of pain, imagining an interlocutor that argues with him:

> 'But you will surely admit that there is a difference between pain-behaviour accompanied by pain and pain-behaviour without any pain?' – Admit it? What greater difference could there be? – 'And yet you again and again reach the conclusion that the sensation itself is a nothing.' – Not at all. It is not a something, but not a nothing either! The conclusion was only that a nothing would serve just as well as a something about which nothing could be said. We have only rejected the grammar which tries to force itself on us here.

I interpret this passage as saying that an individual pain experience cannot be described as being such and so; one cannot say anything more about the experience of pain than that one is, in fact, experiencing it. If so, it cannot be treated as an object for predication.

Suppose I have headache and tell that to someone else. Now why does Wittgenstein say that we cannot say anything about it? Why isn't e.g., 'headache' a description of a certain state? His point was, I think, that if a description is to have any use in a sentence, it must be possible state criteria telling whether an object satisfy this description or not and that these criteria must be such that people

can agree whether they are satisfied or not. But if these criteria in principle cannot be intersubjectively applied, as is the case of qualitative experiences, then such a description has no role to play in our language. It would lack meaning. And one might doubt if it is legitimate to talk about an object about which we cannot say anything.

What one can do, however, is categorize with reference to external objects or circumstances. We can distinguish between experiences of red and experiences of blue because the criteria for this distinction can be formulated in terms of the colour of external objects. The same goes for pain (We can distinguish between pains in different parts of the body, for example.).

The idea that Wittgenstein wants to discard is that, since 'pain' is a noun, it must refer to a thing, an object, about which one can make true or false propositions. According to Wittgenstein, this idea is incorrect. The conclusion is thus that the words we use to talk about inner states cannot function as terms referring to objects of any kind. In Wittgenstein's words we should reject the grammatical idea that nouns must refer to objects.

Some philosophers, such as Chalmers, have objected to the above conclusion by arguing, that we need a better theory: our present understanding of mental phenomena is severely incomplete.

A better theory requires new concepts enabling us to formulate more detailed descriptions of mental states. But Wittgenstein's point remains. Concepts by their very nature require criteria for application that can be intersubjectively determined, and if we by the word 'quale' or 'the quality of a phenomenal experience' try to refer to a certain state of mind that by definition is subjective and not accessible to other persons than the person who is having the experience, then we have failed to refer to anything.

Thus, of the three purported features of the mental, intentionality, subjectivity and the qualitative aspect of conscious states, I think the one to really take care of for reductionists is intentionality. Subjectivity may not be irreducible and I think we should dismiss talk about qualia as a philosophical mistake; talking about qualia is trying to say something where nothing could be said.

12.7 Mental Causes

Light, sound waves and molecules constantly bombard our sense organs, causing sensations to occur in us, whether we are awake or asleep. Sensations are mental states that, together with memories, cause other mental states, such as emotions or expectations.[8] These emotions, along with certain beliefs, cause our actions, or so it

[8] In Brentano's distinction between physical and and mental states, sensations are physical states since they have no act-object structure, which for Brentano was the distinctive trait of mental states. But here we include sensations as pains in the mental.

is generally thought. These actions often have a physical component; we move our arms and legs or emit sound waves. According to this description, physical events cause mental states, which in turn cause other physical events. But how can this be? Recalling what was claimed in Chap. 7, if an event is *the cause* (or *a cause*) of another event, these events are connected to each other by one of the four fundamental types of interactions found in nature. Thus every causal connection between two events must be a connection between physical events; as previously discussed, the physical world is causally closed. However, a mental event is not a physical event, so it seems that we have a problem.

There are three conceivable solutions to this problem. First, one could give up the principle regarding the causal closure of the physical world. Second, one could deny that the concept of a cause is applicable to the connection between thoughts and feelings and our actions, and third, one could say that mental events are in fact identical to physical events.

The first solution has not attracted many supporters for the simple reason that it involves an abandonment of monism. In contrast, the second view is quite common. Many have thought that one should not describe thoughts and feelings as *causes* of our actions, but as *reasons* for our actions. Cause and reasons, on this view, are completely different things, and the concept of a cause in quite simply irrelevant for the description of human *actions* and their connection to our thoughts, opinions, motives and feelings. The problem with this view is that the reasons for our actions become irrelevant for the execution of them; for, if reasons and causes are different things, and if there are physical causes of the observable, behavioural component of our actions, then our reasons do not determine our actions over and above their determination by the prevailing physical. Reasons become redundant.

The third view is that mental events and states are identical to physical events and states where the reasons, i.e., motives and beliefs we propose for our actions, are all physical states of the brain. It may be that one cannot exactly pinpoint *which* physical events are identical to which mental states, but this is not a decisive counter-argument against the idea that mental and physical events are identical, either at the type or token level.

The weaker stance is token identity; it says that every *individual* mental event is identical to some *individual* physical event, but it does not follow that a certain *type* of mental event can be identified with a certain *type* of physical event. If, on a number of different occasions, Lisa is jealous of Karl, then these instances of jealousy are different individual events of the same mental type. It is conceivable that every such individual mental state is identical to some physical state in her brain without every such brain state belonging o the same general type. Indeed, describing Lisa as being jealous at different occasions may correspond to different types of brain states on these different occasions. If such is the case, then reductionism is impossible, since a successful reduction would require psycho-physical laws connecting mental event *types* to physical event *types*. Thus one can be a proponent of the view that tokens of mental events are identical to tokens of physical events and at the same time claim that a reduction of types of mental events to types of physical events is impossible. According to this view, one can

also claim that a mental event is the cause of a physical event, and vice versa. For, if the mental event m is identical to a physical event f (i.e. $m = f$), and f is a cause of another physical event g, then m is a cause of g.

This is the core idea in Donald Davidson's *anomalous monism*. He strongly argues for the claim that the reduction of mental events to physical events is impossible by arguing that there cannot be any strict psycho-physical laws. Davidson admits that one can formulate approximate psychophysical laws, but he thinks this insufficient for achieving a successful reduction. A full discussion of the details of his argument will not be given here; only some of the highlights will be mentioned by way of justifying his non-reductive identity thesis.

One of Davidson's points is that one can distinguish between stated/believed motives and actual motives for actions. People in general assume that the motives for their own actions are good and known to them, and yet others often have strong suspicions that the true motives are quite different from those the agent would give. According to one of psychoanalysis' fundamental assumptions, which most people accept even though they are sceptical about much else in psychoanalysis, many of our psychological states are unconscious. So it is quite possible that an agent's officially stated motive for her action need not be the cause of the action, even if the agent honestly believe so. It is the real motive of an action that is identical to the cause of the action; that is, that motive which actually brought about the action, and nothing else.

12.8 Speculations

One way of trying to understand the relation between the mental and the material, whilst at the same time take into account the distinctive aspects of mental processes, is to characterize the surrounding debate as two different perspectives on the same object. When we use concepts like neuron, brain state and other biological concepts, we assume a third-person, scientific, perspective where the criteria for using terms can be made intersubjectively verifiable. On the other hand, when we use concepts like sensation, emotion and thought, we assume a first-person perspective. Thoughts, feelings and sensation are brain states described, using mental vocabulary, by the subject to whom those brain states belong. It is clear, by definition, that this is a subjective perspective in which we associate a qualitative subjective experience with each of the terms, a subjective experience which cannot be expressed in a third-person perspective. Wittgenstein's argument, referred to above, shows that no intersubjective criteria for these experiences can be given. Unfortunately, common language often is not clear about this distinction.

One way of understanding at least the intentional aspect of sensations is to relate this aspect to the brain's ability for sorting external stimuli. Everyone has experienced situations in which our attention is so directed towards something (a thought, a feeling or some event) that we are more or less oblivious of anything else going on in our immediate surroundings. This is an extreme situation, but such situations do

often occur, though perhaps to a lesser degree. The brain's first task is always to filter the signals coming in through our sensory organs. If we did not have this ability, we would not be able to concentrate our attention on any one thing; we would be, in effect, helpless. For example, when we say that we observe a certain brand of car, or a characteristic smell, we are experiencing an effect of the brain's filtration of external stimuli. Those stimuli that remain make their way to the prefrontal cortex, the centre of consciousness, where they somehow are processed to become a perceptual object, i.e., a representation of the object towards which we direct our attention.

This way of thinking yields, in my opinion, a plausible explanation of the directedness of sensations in physiological terms. However, it does leave one fundamental question unanswered: what more is required for one to be conscious of something? Do various signals have to have certain properties, or do they simply have to reach a particular part of the brain? And what about intentional objects that do not exist?

12.9 The Science of Man

The expression 'science of man' was used by eighteenth century philosophers as referring to studies of human thinking, feeling and deliberating (The word 'man' should here not be understood as contrast to 'woman' but to 'nature'.). It is closely connected to philosophy of mind. The core problem in philosophy of mind is in my view how to integrate descriptions of the mind into a biological description of humans. I'm convinced that the mind, or the soul, is not a separate substance different from the human body. The mind-body relation is not a relation between two different things, but a relation, or perhaps a system of relations between two different ways of *describing* states of affairs and events within a human being. When we say about a certain person that she has a certain thought, desire, or emotion, we apply a mental predicate and say that the person we are talking about satisfies this predicate (See Appendix). When we say about this person that her central nervous system is in a certain physiological state, we say that she satisfies a biological predicate. Hence, when trying to understand the relation between mind and body we should study how we use mental and biological predicates to sort individual person's states into categories and how mental and biological predicates are related.

There is little hope of finding that a certain mental predicate has the same extension as a certain biological predicate, because we apply rationality principles when using mental predicates, as Davidson has shown (see Sect. 12.7), whereas no such principles are operative when we apply biological predicates. Hence, reduction of the mental to the biological seems, in my view, utterly implausible.

If our goal with the study of science of man is to gain better understanding, we should use mental predicates in our studies, because 'understanding' means knowing what beliefs, desires and emotions people has. Many hold that it is difficult,

some say impossible, to formulate reliable laws, or regularities useful for predictions, using mental predicates. This stronger stance is taken by Davidson, who argues against the possibility of strict psycho-physical laws connecting the mental to the physical.

But I see a way out here, inspired by Freud, the inventor of psychoanalysis. Freud had patients whose actions appeared irrational in the sense that they did not seem to follow from the agent's conscious desires and beliefs. Freud's idea was that their actions in fact was driven by unconscious beliefs and desires, they were hidden from the agent's own consciousness by a kind of inner censor. Hence introspection into our own mind is not at all reliable; we often fail to obtain knowledge about our own desires and beliefs (Psychoanalysis is a highly controversial theory, but the notion that one may have unconscious desires and beliefs is generally accepted.). This was, by the way, the same conclusion James Watson (1913), the founder of behaviourism, drew from the complete failure of empirical psychology during nineteenth century. So if we want to retain the belief-desire-action model of understanding humans, we cannot rely on individual subject's reports about their own beliefs and desires. But another strategy is possible. Suppose we accept as a fundamental law: If an agent desires the goal G and has the belief that the best action for achieving the goal G is to perform the action A, then the agent will do A (As I argued in Chap. 10, a general statement may be a law, in spite of the fact that it is, or is part of, an implicit definition of a term used in the formulation of that law.).

This law can easily be refuted if we identify the goal, the action and the belief independently of each other. But all three concepts are intentional, so it is obvious that neither can be directly observed; identifying a particular desire, belief or action must be done indirectly, via interpretations of observable behaviour, such as responses to stimuli, answers to questions etc. Just as we identify a certain charge by observing the motion of a physical body, we may identify a certain belief or desire by observing the outward aspect of the agent's action, his behaviour. It might be more complicated in the mental domain than in electromagnetism, but I see no principal difference. The crucial restriction is that the agent's own reports about his/her desires or beliefs should not *automatically* count as evidence against such an attribution, just because such reports are the results of introspection. Introspection is unreliable.

Daniel Kahneman's book, '*Thinking fast and slow'*, are full of such examples (See more about this book in Sect. 14.1). If, for example, the fast system determines our action, we can describe its operation as that an unconscious belief and desire is formed such that the appropriate action takes place. One could perhaps object by saying that unconscious mind-states cannot be beliefs or desires; such states are by nature conscious. But, again, we have strong evidence that we are not conscious of most of what goes on in our minds; this was a main reason for Freud to invent his theory about unconscious dynamical processes. One could be sceptical about Freud's theory without denying that the distinction between conscious and unconscious mental states is a distinction with a difference.

In fact, the concepts of belief and desire, in ordinary parlance, has a strong functional character; they are names for those states of our minds that guide our

actions; and calling a state a belief state doesn't entail anything about its mode of operation in producing an action, or that it is conscious. So I see no fundamental difficulty in a science of man aiming both at understanding and at prediction.

Many argue that man and society is so much more complex than simple physical or chemical systems so that there is no hope of making reliable predictions. But the premise that man and society is more complex I outrightly reject. Most purely physical, chemical and biological systems have an enormous complexity, as measured for example by the number of degrees of freedom. I see no difference in that respect between a human being, a human being in its environment, or a society and a system studied by natural science.

A second argument for the difference between science of man and of nature is that for ethical reasons we cannot perform controlled experiments in strictly controlled circumstances. This is true, of course, but the same problem is prevalent also in many areas of natural sciences. One way to overcome this is to study partial systems and then construct models for their combination; and this can be done also in the study of man. Another is to use statistical methods, which, as we have seen, gives *some* indications about possible mechanisms and causal structures. And this can be done also in the study of man.

A third argument for the difference between the study of man and of nature is that humans react and change their minds when being studied. This is certainly true. Humans are essentially agents whose beliefs, desires and volitions may change when being described by others, whereas things studied by physics, chemistry or biology are not in this sense agents, they have no thoughts. But I doubt if this really is a difference that have methodological relevance. The fact that people are affected and may change their minds when being subject to actions or descriptions from other people is a typical example of a feedback mechanism. And the same is true in micro physics; observing an object means exchanging photons, i.e. portions of energy with it, and when observing the smallest objects they may change their state, causing difficulties in making predictions. The mechanisms differ, but the fundamental methodological difficulty, to make reliable predictions, is the same.

Still, both in physics and in the study of man we are sometimes able to make predictions of the evolution of ensembles of individuals, particles in the case of physics and people in the case of the study of man.

All in all, I see no fundamental obstacle for future progress in the study of man using mental predicates in the descriptions: And by progress I mean not only to provide understanding of individuals or groups of individuals and societies, but also results with predictive force, at least predictions in terms of probabilities.

Discussion Questions

1. In the film *Alien*, there is a scene where two crewmembers of a spaceship become violently disagreeable and begin to fight. During the fight, one of the crewmembers falls against a sharp corner. To everyone's surprise, the crewmember's wound does not release blood but hydraulic oil! This 'crewmember' was in fact a robot, which the villains had smuggled onto the ship in order to destroy it.

Suppose that this robot behaves just like a human; that it speaks, jokes, eats and so forth. Is it possible that this robot lacks consciousness; that it has no feelings, thoughts or sensations? If your answer is yes, what more do you require of the robot for it to be capable of these experiences?

2. Within computer linguistics, researchers are eagerly investigating technology that can translate one natural language into another, and much headway has been made. In the E.U., for example, computers are now responsible for the brunt of the work in the translation of official documents. It does not seem unreasonable to believe that computers of the future will be able to translate non-fictional texts just as well as a human. Would we then say that such a computers understand the meaning of various words and expressions? If not, why would we say that a human who performs the same task does understand the meaning of the text he or she translates?

Further Reading

Chalmers, D. (1996). *The conscious mind*. Oxford: Oxford University Press.
Churchland, P. M. (1988). *Matter and consciousness*. Cambridge, MA: MIT Press.
Churchland, P. M. (1995). *The engine of reason, the seat of the soul*. Cambridge, MA: MIT Press.
Churchland, P. S. (2013). *Touching a nerve. Our brains, our selves*. New York: W.W. Northon & Co.
Davidson, D. (1980). *Essays on actions and events*. Oxford: Clarendon.
Dennett, D. (1981). *Brainstorms*. Cambridge, MA: MIT Press.
Kim, J. (1998). *Philosophy of mind*. Boulder: Westview Press.
Putnam, H. (1960). *Minds and machines*. Reprinted in Putnam 1975, pp. 362–385.
Putnam, H. (1967). *The nature of mental states*, reprinted in Putnam 1975, pp. 429–440.
Putnam, H. (1975). *Mind, language, and reality*. Cambridge: Cambridge University Press.
Searle, J. (1992). *The rediscovery of mind*. Cambridge, MA: MIT Press.
Watson, J. B. (1913). Psychology as the behaviourist views it. *Psychological Review, 20*, 158–177.
Wittgenstein, L. (1953). *Philosophical investigations*. Oxford: Blackwell.