

WITTGENSTEIN, WITTGENSTEINIANISM
AND THE CONTEMPORARY
PHILOSOPHY OF MIND – CONTINUITIES AND CHANGES*

Ansgar Beckermann

For a short period of time in the middle of the last century, at least in Europe, Wittgenstein was the measure of all things in philosophy and especially in the philosophy of mind. The private language argument had shown the conception of the mind going back to Descartes and Locke to be principally flawed – or so the consensus was. Mental phenomena are not essentially private, and there simply cannot be mental states without any observable criteria at all. Anyone who disagreed was in for a difficult time. Yet, only one or two decades later the discussion had moved on considerably. First, the identity theory overcame behaviourism. Second, functionalism superseded the identity theory, thereby paving the way for more specialist approaches such as Fodor's representational theory of mind. Finally, a new, post-Wittgensteinian orthodoxy developed with amazing swiftness. With hindsight, this seems a remarkable phenomenon in the sociology of philosophy. As interesting as it would be, I shall disregard the sociological side here. Instead, I am interested in the questions: What has changed? And, are there good reasons for these changes?

These questions are difficult to answer, not least, because it is far from clear which position Wittgenstein himself held with regard to the mind-body problem. Even today, articles and books are being published in an attempt to come closer to answering this question; but 50 years after Wittgenstein's death we are still far from a generally accepted consensus. Of course, this picture is painted a little bleakly: some points are clear – for instance Wittgenstein's rejection of the picture of an inner world of the mind and an external world of material things, which pervades all areas of philosophy since early modern times. Ac-

According to Descartes, the mind is an immaterial substance of its own – a *res cogitans* – and thinking, feeling and remembering are occurrences in this substance accessible only to this substance itself. The consequence of this picture is that the mind is private. Only the mind itself can know what happens within; others, at best, have indirect access through a kind of inductive inference. They have to infer from the person's behaviour what goes on in a person's mind. This, Wittgenstein says in a famous passage in the *Philosophical Investigations*, is complete nonsense:

In what sense are my sensations *private*? – Well, only I can know whether I am really in pain; another person can only surmise it. – In one way this is wrong, and in another nonsense. If we are using the word “to know” as it is normally used (and how else are we to use it?), then other people very often know when I am in pain. – Yes, but all the same not with the certainty with which I know it myself! – It can't be said of me at all (except perhaps as a joke) that I *know* I am in pain. What is it supposed to mean – except perhaps that I *am* in pain? (PI § 246)

If the epistemic consequences of the Cartesian picture of the mind are nonsensical, the picture itself must be wrong: the mind is not private, but public. But what does this mean? The predominant view in the 50s and 60s was a view that one could call the “criteriological account”¹. According to the proponents of this view Wittgenstein has shown by means of considerations on the meaning of linguistic expressions in general that there can be no mental states without behavioural criteria. *Pain behaviour* is not just a *symptom* of the mental state pain, but a *criterion*. That is to say, pain behaviour is corrigible evidence that somebody is in pain, but for semantic reasons it is, in a certain way, also sufficient evidence. *For semantic reasons* it is true that if a person shows this behaviour and there is no evidence to the contrary, then this person is in pain. There is ample evidence that Wittgenstein held this view. Consider for example this passage from the *Blue Book*:

When we learnt the use of the phrase “so-and-so has toothache” we were pointed out certain kinds of behaviour of those who were said to have toothache. As an instance of these kinds of behaviour let us take holding your cheek. [...] Now one may [...] ask: “How do you know that he has got toothache when he holds his cheek?” The answer to this might be, “I say, *he* has toothache when he holds his cheek because I hold my cheek when I have toothache.”

But what if we went on asking: – “And why do you suppose that toothache corresponds to his holding his cheek just because your toothache corresponds to your holding your cheek?” You will be at a loss to answer this question, and find that here we strike rock bottom, that is we have come down to conventions. (*Blue and Brown Books*, 24)

In other words: not all signs of the presence of pain can be mere symptoms; some must be criteria in the semantic sense, because otherwise we would have no basis for the application of the concept ‘pain’. Obviously, this is exactly one of the points of the private language argument. However, the criteriological interpretation has been criticised in recent years, among other reasons, because it places Wittgenstein in great proximity to behaviourism, a theory which he explicitly rejected in many places.

What may have been his reasons for this rejection? Perhaps, as Hanjo Glock suggests, that the behaviourist is still sticking too closely to the Cartesian picture by construing the mental after the image of the physical.

Wittgenstein's attack on the inner/outer dichotomy is often accused of reducing the inner to the outer, and thereby ignoring the most important aspects of human existence. Ironically, Wittgenstein in turn accuses the inner/outer conception of mistakenly assimilating the mental to the physical. It construes the relationship between mental phenomena and mental terms ‘on the model of’ material ‘object and designation’, and thereby turns the mind into a *realm* of mental entities, states, processes and events, which are just like their physical counterparts, only hidden and more ethereal [...] [T]his tendency is fuelled by the *Augustinian picture of language*, which suggests that all words stand for objects, and all sentences describe something – if not physical entities, then entities of a different kind. (Glock 1996, 175)

Indeed, there are a number of passages that suggest that it was Wittgenstein's opinion that it is a fundamental error to construe the use of mental terms after the model of the use of physical language. He writes, for example, in the *Philosophical Investigations*:

How does the philosophical problem about mental processes and states and about behaviourism arise? – The first step is the one that altogether escapes notice. We talk of processes and states and leave their nature undecided. Sometime perhaps we shall know more about them – we think. But that is just what commits us to a particular way of looking at the matter. For we have a definite concept of what it means to learn to know a process better. (The decisive movement in the conjuring trick has been made, and it was the very one that

we thought quite innocent.) – And now the analogy which was to make us understand our thoughts falls to pieces. So we have to deny the yet uncomprehended process in the yet unexplored medium. And now it looks as if we had denied mental processes. And naturally we don't want to deny them. (PI § 308)

In the second part of the *Philosophical Investigations*, Wittgenstein touches again on his views on behaviourism:

Then psychology treats of behaviour, not of the mind?
What do psychologists record? – What do they observe? Isn't it the behaviour of human beings, in particular their utterances? But *these* are not about behaviour. "I noticed that he was out of humour." Is this a report about his behaviour or his state of mind? ("The sky looks threatening": is this about the present or the future?) Both; not side-by-side, however, but about the one *via* the other. [...] It is like the relation: physical object – sense-impressions. Here we have two different language-games and a complicated relation between them. – If you try to reduce their relations to a *simple* formula you go wrong. (PI, 179 f.)

In my eyes, this remark is rather enigmatic. One thing is clear, however: for Wittgenstein the difference between speaking-of-behaviour and speaking-of-mental-states is similar to the difference of speaking-of-physical-objects and speaking-of-sense-impressions. Here we have two different language games, even if these language games – as he explicitly points out – are closely linked. But, and this is very regrettable, Wittgenstein says very little about how the language game of talking about the mental really works and how it differs from speaking about behaviour and from other more physical language games. The situation gets even more confusing because in the *Remarks on the Philosophy of Psychology I* and in *Zettel* we find a rather peculiar passage, according to which Wittgenstein sees the level of psychology and that of physiology as completely separate.

No supposition seems to me more natural than that there is no process in the brain correlated with associating or with thinking; so that it would be impossible to read off thought-processes from brain-processes. I mean this: if I talk or write there is, I assume, a system of impulses going out from my brain and correlated with my spoken or written thoughts. But why should [...] this order not proceed, so to speak, out of chaos? The case would be like the following – certain kinds of plants multiply by seed, so that a seed always produces a plant of the same kind as that from which it was produced – but *nothing* in the seed corresponds to the plant which comes from it; so that it is impossible to infer the properties or structure of the plant from those of the seed that it comes out of [...] [T]here is no reason why this should not really hold for our thoughts, and hence for our talking and writing. [...] (RPPI 903; see Z 608)

I saw this man years ago; now I have seen him again, I recognize him, I remember his name. And why does there have to be a cause of this remembering in my nervous system? Why must something or other, whatever it may be, be stored-up there *in any form*? Why must a trace have been left behind? Why should there not be a psychological regularity to which *no* physiological regularity corresponds? If this upsets our concepts of causality then it is high time they were upset. (RPPI 905; see Z 610)

If we read this passage from today's perspective, Wittgenstein seems to say no more and no less than that the thesis of the *emergent nature* of mental phenomena appears extremely plausible to him². To that end he is prepared to allow a causality between mental phenomena that is not mediated by physiological processes, even if this brings the concurrent danger that it may appear to count in favour of classical mind-body dualism.

The prejudice in favour of psycho-physical parallelism is also a fruit of the primitive conception of grammar. For when one admits a causality between psychological phenomena, which is not mediated physiologically, one fancies that in doing so one is making an admission of the existence of a soul *alongside* the body, a ghostly mental nature. (RPPI 906; see Z 611)

Of course, all these quotations come from texts written at very different times and occasions. But even if this is so, it can not be disputed that we do not find a coherent account of the relations and the differences between the mental and the physical, between the use of mental terms and the use of physical terms in Wittgenstein's writings. We do find a straightforward rejection of the idea that the mind is a private inner theatre, and we also find hints in direction of the idea that there is a significant difference between the use of the mental and the physical vocabulary. At least, many philosophers understood Wittgenstein this way. And, what is more, in my view a large part of post-Wittgensteinian philosophy can be conceived of as an attempt to spell out the idea of two levels or two language games.

One reason may have been that this idea seems to allow the dissolution of the mind-body problem. This problem is widely held to be the problem "of accounting for the place of mind in a world that is essentially physical"³. At least at first sight there seem to be mental items in the world: pains and thoughts, colour sensations and wishes, consciousness and perhaps even souls. How is this realm of the mental

related to the realm of the physical? Are mental items in fact not so different, but only a special kind of physical items? Or does the mental constitute a special realm of non-physical entities that nonetheless causally interact with the physical? Wittgenstein's views seem to allow the answer: All these questions are ill conceived. One only has to notice that the mental language does work in a way very different from that of the physical language. Mental terms do not denote special states or processes in the way physical terms denote physical states or processes. Indeed, they do not denote at all. And this means that there simply are no mental items about which we can reasonably ask how they fit into an essentially physical world.

This attractive feature of Wittgenstein's views seems to have inspired for example Gilbert Ryle whose work may also be understood as an attempt to elaborate Wittgenstein's two languages account⁴. To be sure, *The Concept of Mind* was published four years before the *Philosophical Investigations*. However, Ryle's thoughts are so similar to the thoughts of the late Wittgenstein, and he was so much influenced by Wittgenstein through conversations and lecture-notes that I do not hesitate to treat Ryle as a Wittgensteinian here. Ryle, however, seems to have been concerned with a particular aspect of the mind-body problem – the question of how the mind can causally interact with the body. He also tries to dissolve this problem by a semantic argument. According to Ryle, mental terms refer to dispositions and not to events, and, that is, not to possible causes. The problem of how the mind causally interacts with the body simply disappears when we only acknowledge that mental explanations are dispositional explanations. I shall come back to this issue soon. But first I would like to say a few words on the general outline of Ryle's argument.

To begin with, Ryle, too, emphatically rejects the concept of mind that has become pervasive since early modern times. For Ryle, too, it is absurd to assume a Cartesian theatre, in which mental objects abound and mental occurrences take place, that only the mind knows about – occurrences, which moreover interact causally with each other and with occurrences in the physical world. According to Ryle, this assumption rests on one big misunderstanding, or, to put it more precisely, it rests on a category mistake. This mistake can briefly be characterised like this: Cartesians, but not only Cartesians, assume that

mental expressions such as 'to remember', 'to think', 'to perceive', 'to believe' and 'to want' refer to (hidden) *internal occurrences within a person's mind*, which *cause* the person's outward behaviour. In reality however, according to Ryle, we do not employ these expressions to refer to some 'shadow actions', which are hidden antecedents of the overt behaviour. Instead we use the mental expressions to characterise the publicly observable actions in a different way. The mentalist thinks mental phenomena consist in enigmatic occurrences behind the observable actions, while in reality mental phenomena are nothing but a manner of organisation of these actions.

Ryle supports his view that mental expressions do not refer to hidden inner occurrences – among other arguments – with an analysis of intelligent and voluntary actions. His opponent, the mentalist, analyses intelligent, or voluntary behaviour like this:

- an action is intelligent if and only if it has been caused by a corresponding thought;
- an action is voluntary if and only if it has been caused by a corresponding act of the will.

However, when we consider carefully under which circumstances we really call an action 'intelligent', a completely different picture emerges. For, normally, we would say that someone acts intelligently if

- he normally does what he does correctly, well and successfully and if
- he is able to discover a mistake in his way of proceeding and to eliminate it, if he is able to repeat successes and improve on them, if he is able to learn from the example of others etc.

That is to say, a closer consideration of our actual use of language shows that we do not call an action intelligent if we can trace it back to a hidden inner process, but rather if the action does not stand alone, but is part of a pattern of actions and abilities.

'Willing', too, according to Ryle, is certainly not a verb that we use to refer to occurrences (acts of the will); for if this were the case, these occurrences would have to be dateable and countable. However:

No one ever says such things as that at 10 a.m. he was occupied in willing this or that, or that he performed five quick and easy volitions and two slow and difficult volitions between midday and lunch-time. An accused person may

admit or deny that he did something, or that he did it on purpose, but he never admits or denies having willed. Nor do the judge and jury require to be satisfied by evidence, which in the nature of the case could never be adduced, that a volition preceded the pulling of the trigger. (Ryle 1963, 63)

Moreover: What properties could acts of the will have? Can we perform them fast or slowly? Can we do more than one at the same time? Can we interrupt an act of will and later pick it up where we left it? The impossibility of finding answers to these questions shows very clearly that in our everyday understanding ‘willing’ does not stand for occurrences or actions. However, what *does* it stand for?

Ryle advocates carefully considering our everyday use of language. It shows that we normally use the adjectives ‘voluntary’ or ‘involuntary’ when we are interested in the question of whether a mistake deserves reproach. A sailor is asked to tie a reef-knot, but he ties a granny-knot. A pupil arrives late for school. These are typical cases in which we ask whether the respective actions were voluntary or not. It depends upon the answer to this question whether we reprimand the sailor or the pupil or even punish them. The decisive question is whether the person who committed the mistake could have avoided it. This, in turn, depends upon whether he or she had the knowledge and the ability to carry out the action properly, and whether external circumstances have prevented him or her from correctly carrying it out. However, both can be detected without recourse to some mysterious acts of will.

Therefore, voluntary actions, contrary to the official characterisation, are better analysed thus:

- a mistaken (wrong) action is voluntary if and only if the agent possesses the knowledge and the ability to perform the action correctly and if he or she is not prevented by external circumstances from the correct performance of the action.

The analysis of intelligent actions as well as the analysis of voluntary actions therefore shows that the official doctrine is mistaken. Both types of actions are *not* characterised by the ‘fact’ that they are caused by hidden internal events in the agent’s mind.

However, if this is so, the question arises how this mistaken impression could have come about. Why are we so prone to make ca-

tegory mistakes when thinking about the mind? Why do we assume that mental expressions refer to events that take place inside people’s heads? According to Ryle, a central reason for the mistaken views of the official doctrine lies in the fact that the official doctrine construes mental explanations as causal explanations, and therefore views mental phenomena as (hidden) causes. In Ryle’s view, however, mental states are really dispositions and hence mental explanations *dispositional explanations*.

Ryle provides a famous analysis of the sentence “He boasted from vanity.”

The statement “he boasted from vanity” ought, on one view, to be construed as saying that “he boasted and the cause of his boasting was the occurrence in him of a particular feeling or impulse of vanity”. On the other view, it is to be construed as saying “he boasted on meeting the stranger and his doing so satisfies the law-like proposition that whenever he finds a chance of securing the admiration and envy of others, he does whatever he thinks will produce this admiration and envy.” (Ryle 1963, 87)

According to Ryle, it is perfectly obvious that the first analysis is totally absurd. When we ask to which type of explanation the statement “he passed his neighbour the salt from politeness” belongs, it is immediately obvious that the causal analysis is not tenable. This is so because a polite person is one who has the disposition not to jump the queue, to let others pass before him, to help without being asked, to avoid making tactless remarks, not to make his hosts uncomfortable through inappropriate dress or inappropriate behaviour, and so forth. Furthermore, the dispositional character of this explanation also shows itself from the fact that it requires supplementing by a causal explanation.

But the general fact that a person is disposed to act in such and such ways in such and such circumstances does not by itself account for his doing a particular thing at a particular moment; any more than the fact that the glass was brittle accounts for its fracture at 10 p.m. As the impact of the stone at 10 p.m. caused the glass to break, so some antecedent of an action causes or occasions the agent to perform it when and where he does so. For example, a man passes his neighbour the salt from politeness; but his politeness is merely his inclination to pass the salt when it is wanted, as well as to perform a thousand other courtesies of the same general kind. So besides the question “for what reason did he pass the salt?” there is the quite different question “what made him pass the salt at that moment to that neighbour?” This question is probably answered by “he heard his neighbour ask for it”, or “he noticed his neighbour’s eye wandering over the table”, or something of the sort. (Ryle 1963, 109)

The emerging picture is this: According to Ryle, dispositional statements like “this plane is brittle” or “John is polite” are encapsulated laws or law-like statements. Within the context of explanation dispositional statements, therefore, never express antecedent conditions and, that is, causes. Their role is the role of statements expressing laws. This, in turn, implies that dispositional explanations are, in a sense, incomplete. In addition to the relevant dispositions, i.e. laws, we also need to know the relevant antecedent conditions and, that is, the causes of the action to be explained.

In Ryle’s eyes there can be no doubt that the explanation “he passed his neighbour the salt from politeness” – just as the explanation “he boasted from vanity” – is, in this sense, a dispositional explanation. And this, in his view, implies that the mental phenomena, to which the overt behaviour is traced back in these explanations, are not mysterious inner processes in the agent’s mind, but dispositional properties, which are just as publicly accessible as are the dispositions of brittleness or of being soluble in water.

With this dispositional analysis Ryle aims at the same thing as with his alternative analyses of intelligent and voluntary behaviour. Firstly, he wishes to show that mental expressions do not refer to hidden inner occurrences in a person’s mind, but to circumstances whose public observability is beyond doubt. Secondly, he wants to show that mental concepts *as dispositional concepts* do not refer to the causes of actions and that mental explanations, therefore, never compete for instance with physiological explanations.

A closer look reveals, however, that in Ryle’s considerations there are at least two accounts of what the use of mental vocabulary amounts to: the pattern account and the dispositional account. According to the pattern account using mental terms to describe and explain behaviour is nothing but regarding the behaviour as an integral part of a wider pattern of behaviour and capacities. The dispositional account is more straightforward. According to the dispositional account, in using mental concepts we do nothing but ascribe certain behavioural dispositions and capacities to the person in question.

In the history of Ryle reception the dispositional account was obviously the more prominent one. However, it has been precisely this aspect of Ryle’s thought that proved to be particularly vulnerable to objections. This is so for three reasons:

- (1) Dispositional predicates are not just used when we characterise a person’s mental life; they are just as much at home in the vocabulary of the natural sciences. Mass or weight, for instance, are classical examples of dispositions. That an object has the weight of 10 kp means, among other things, that if placed on scales it will generate a certain movement of the pointer. That an object has the mass of 10 kg means, among other things, that it experiences a certain amount of acceleration if a certain force acts upon it. Furthermore, natural dispositions are multi-track, too – i.e., they are characterised not just by one occasion-reaction pair but by quite a number of these pairs. Therefore, the dispositional account is not very helpful when we are trying to draw the line between the mental language game and the physical language game. At most, we can distinguish between the language games of events and the language games of dispositions in this way; this distinction, however, stands in an orthogonal relation to that between the mental and the physical language game.
- (2) Ryle’s thesis that a categorical difference exists between dispositional explanations and causal explanations, was highly controversial from the outset. Let us just listen to one of the voices of the diverse chorus of critical contributions to this question.

[...] actually, it is a mistake to suppose that only events may be properly spoken of as causes, for we frequently refer to states, dispositional properties and even the failure of events to occur as causes. For example, given appropriate circumstances, we might speak of a bent rail, an icy track or the failure of the brakeman to signal as the cause of a given train accident. (Gean 1965/1966, 677)

It is possible that Ryle might have felt vindicated in his view through Davidson’s thoughts on the concept of causality according to which the relata of causal relations are events and nothing else. However, many doubts remain.

- (3) The strongest argument against Ryle’s dispositional account arose from developments in the theory of scientific concept formation around that time. As early as 1936/1937 in his paper “Testability and Meaning” Carnap broke with the old thesis of Logical Empiricism that all scientific concepts have to be definable exclusively in observation terms and logical vocabulary. It was the dispositional terms that had been the downfall of this thesis. Carnap

therefore suggested analysing the recalcitrant dispositional concepts by means of so-called reduction sentences. In the final consequence, however, this suggestion proved to be only the first step to a complete dissolution of the empiricist criterion of meaning. Once the step was taken, the view gained hold that most concepts of scientific interest could not be defined in purely observational terms or even be adequately analysed by means of reduction sentences. Rather, according to the new insight, central concepts such as length, mass, temperature and charge are *theoretical concepts*, which receive their meaning on the one hand through their relations to other theoretical concepts and on the other hand through a number of correspondence rules, which – rather loosely – connect certain theoretical concepts with observational terms. An early canonical formulation of this new view can be found in Hempel's *Fundamentals of Concept Formation in Empirical Science*, which was published in 1952.

After this theory, which Carnap endorsed too, had gained fast acceptance, it was natural to conceive of mental terms no longer as dispositional predicates, but also as theoretical terms – especially so, since Ryle had already spoken of multi-track dispositions in this context. This view was defended by, e.g., Fodor and Chihara as well as by Brandt and Kim, who immediately drew a conclusion that is very uncomfortable for the dispositional account: if mental concepts really are theoretical concepts, then they are indeed not different from the concepts of natural science and everything suggests that mental explanations are completely normal causal explanations much like “the iron glowed red because it was heated to 750° C” or “this piece of iron attracts iron filings because it is magnetic”. In other words, if mental concepts really are theoretical concepts, then statements like “John is angry” are *not* encapsulated laws or law-like statements, but statements that express antecedent conditions – and that is, causes – of actions.

This conclusion was strongly supported by the resurrection of realism going back to Carnap's and Tarski's work on pure semantics. Actually, this major shift in the fundamental ideas of Logical Empiricism also took place about 1950. Physical objects were no longer re-

garded as logical constructs out of sense data and unobservable states and properties no longer as logical constructs out of observables. Physical objects and unobservable states and properties were again considered as perfectly *real* though not *given*. In his article “The Mind-Body Problem in the Development of Logical Empiricism” Herbert Feigl writes⁵:

The slogan of Vienna Logical Positivism: “The meaning of a statement is the method of its verification”; and the slogan of Bridgman's operationism: “A concept is synonymous with the set of operations [which determine its applications]” were excellent preventives of the transcendent type of metaphysical speculations. [...] Logical empiricism in its later development, however, had to replace these radical principles by more conservative ones. ... [T]he meaning of scientific statements cannot in general be identified with their confirming evidence. [...] For a [...] very simple example we may refer to the concept of the temperature of a body. As ordinary and scientific commonsense [...] would put it, thermometer (or pyrometer) readings, spectroscopic findings, and other types of measurement merely indicate something about the body in question, namely the intensity of heat which is a state of that body. No matter whether this heat intensity is construed in terms of classical (macro-) thermodynamics or in terms of statistical (micro- or molecular) thermodynamics, it is in any case only *evidenced by but not identical with* those indications. Similarly for psychology: The overt symptoms and behavior that indicate an emotion, like e.g., anxiety, are confirmable and measurable in terms of skin-temperature, endocrine secretions, psychogalvanic reflexes, verbal responses, etc. but must not be confused with the emotion itself. Generally, the “theoretical constructs,” i.e., the hypothetically assumed entities of the sciences cannot be identified with (i.e., explicitly defined in terms of) concepts which apply to the directly perceptible facts as they are manifest in the contexts of ordinary observation or of experimental operations. (Feigl 1950a, 617f.)

One page later Feigl continues:

The realistic correction of positivism consists in the identification of meaning with factual reference. This conforms well with customary usage according to which a statement *means* a state of affairs; and is *true* if that state of affairs is fulfilled (“is real,” “exists”). This is the obvious grammar of “meaning,” “truth,” and “reality.” (Feigl 1950a, 619)

Feigl could have added: According to customary usage, even the meaning of predicates as ‘temperature’ and ‘pain’ cannot be identified with the methods we use in trying to find out what temperature a body has or whether a person is in pain. Even these predicates *refer* – to the internal state of a body or the mental state of a person which are

“evidenced by but not identical with” thermometer readings or certain symptoms and ways to behave⁶.

Perhaps these considerations contributed to the fact that other authors preferred the pattern account, which, as already mentioned, can also be found in Ryle. As early as the 50s and 60s of the last century Melden formulated thoughts in this direction. But, as far as I know, this account was only thoroughly spelled out by Eike von Savigny in his interpretation of the philosophy of the *Philosophical Investigations*. Central to this interpretation is the thesis that with regard to mental states it is not only the behaviour of the person who has the mental state in question that counts, but also the behaviour of the members of the society in which this person lives. Von Savigny believes that Wittgenstein held the following view:

The fact that someone imagines something, expects something, wishes something, feels something, thinks of something or intends to do something etc. does not concern that person in isolation. Rather, this fact consists in that the *patterns* of this person’s individual behaviour are in a certain way embedded in the *pattern* of the social behaviour of the community to which he belongs. (von Savigny 1994, 10 – italics mine)

Let us suppose someone has a headache and he displays the corresponding behaviour: he tries to hold his head still, he presses his hands against his temples, cools the forehead with a wet cloth and retires to a dark room. But not only that, his fellow humans, too, react in a way that we are used to: they offer him aspirin, they slink around noiselessly, they show sympathy towards him etc. This person is in pain. Let us now consider someone who behaves in exactly the same way as this person, but his environment reacts in a completely different manner:

Moaning surprises the people; one does not receive any aspirin, and nobody calls the doctor. What we regard as pain behaviour, would there be treated as if the people were affected by some passing peculiarity. Their behaviour irritates the others, but they put up with it. The person with the headache has not changed in himself. (von Savigny 1994, 11f)

According to von Savigny, Wittgenstein would say, relative to this second environment, that the person no longer has a headache. That someone has a headache means that his own behaviour shows a certain pattern, and that the reactions of his fellow humans have a certain pattern. The extreme – and in my eyes very implausible – ex-

ternalism of this view does not have to concern us here since the basic ideas of the pattern account may well be separated from it. For our purposes it is only important that according to the pattern account the following is true:

- If one says that a person is in pain, one does not say that she is in any particular inner state, but that this person’s behaviour (and the behaviour of his fellow humans) has a certain pattern.
- If one says a person is holding her cheek because she has a toothache, one does not causally explain this behaviour by tracing it back to a cause – the state of pain – but rather one explains this behaviour by pointing out that it forms part of a certain pattern of behaviour.
- If one says that a person simulates pain, one says that the same behaviour – holding one’s cheek – for this person is embedded in a different pattern of behaviour. (An actor who acts being in pain, under certain circumstances behaves differently from someone who really is in pain.) If the same behaviour is embedded in exactly the same pattern of behaviour, it does not make sense to say in one case that the person is in pain and in the other she is not.

Let us recapitulate here. Wittgensteinians and the proponents of the new orthodoxy agree in one point – their rejection of Cartesianism. The mind is not a non-material substance and the mental is not an inner world of occurrences to which only the mind itself has access. Mental terms such as ‘to remember’, ‘to think’, ‘to perceive’, ‘to believe’ and ‘to want’ do not refer to hidden occurrences inside a person or inside a person’s mind. The mental is just as public as the physical. However, here the agreement ends. For the proponents of the new orthodoxy spell out the rejection of Cartesianism differently from the Wittgensteinians. For the new orthodoxy the mental is public because mental concepts are theoretical concepts – concepts with which we ascribe theoretical properties or states to persons and which – as concepts – have the same status as the concepts ‘is magnetic’ or ‘has a mass of 10 kg’. Even if in mental explanations the behaviour of persons is not traced back to hidden mental occurrences, these explanations are every bit as causal as the corresponding physical explanations.

The characteristic view of most Wittgensteinians on the other hand is that this picture is fundamentally wrong. Mental expressions do not refer to any states – not even to theoretical ones. By means of mental expressions we characterise persons whose behaviour shows a certain pattern; but with these expressions we do not ascribe states to these persons which give rise to these patterns. For this reason, mental explanations are certainly not causal explanations. Given these views it is no wonder that the battleground for the fight between Wittgensteinians and the proponents of the new orthodoxy became the question of whether mental explanations are causal explanations or explanations of a completely different kind. What were the reasons which the proponents of the new orthodoxy put forward for their case?

To start with: Wittgensteinians who are committed to the pattern account claim that people who are in pain show a very specific behaviour. But they do not say anything about the causes of this behaviour; sometimes it even seems as if they regard even asking the question of the cause as illegitimate or nonsense. *Prima facie*, however, nothing counts against the view that pain behaviour has a cause, too. And don't we say "He is holding his cheek *because* he is in pain"? So what seems more natural than to assume that with the expression 'pain' we do not refer to a pattern of behaviour, but to the *cause* of this behaviour?

Let us consider the parallel physical case of being magnetic. The behaviour of magnetic objects also forms a characteristic pattern: they attract iron filings in their proximity and induce an electrical current in coils which they pass through, the needle of a compass near them tends to point in their direction etc. However, when we say of an object *a* that it is magnetic then we thereby do not say that *a* shows the behaviour that is characteristic of magnetic objects, but that it has the property that is causally responsible for this pattern of behaviour. Why should this case differ so fundamentally from the case of pain?

Secondly, even proponents of the new orthodoxy do not deny that there is a relatively close relationship between mental states and typical behaviours. But they point out that it is simply not the case that all persons who are in pain show the same behavioural pattern. Compare, e.g., a person who cuts her finger during the usual morning toilet with a person who cuts her finger while hiding in a cave to escape her persecutors. The respective behaviour will be as different

as you can imagine. The simple explanation for this is straight forward: Pain causes behaviour, but it does not do so in isolation, but only in the context of the respective (relevant) beliefs, wishes, ideological attitudes etc. Thus, in the context of different mental states pain may give rise to different patterns of behaviour.

Thirdly, pain itself can obviously be causally influenced in a multitude of ways. On the one hand pains are caused – through stabbing or beating, too much alcohol or muscle cramps and many other things. On the other hand, one can causally fight pain – through painkillers, through a cold compress, sometimes through warming the affected part, through acupuncture or through relaxation techniques etc. How can a proponent of the pattern account integrate this? Only by claiming that stabbings, beatings or too much alcohol causally lead to the person showing the pattern of behaviour characteristic of pain, and that painkillers, a cold compress on the forehead or acupuncture causally lead to the person ceasing to display this behaviour. This seems at least implausible. Causing pain is something different from causing pain behaviour, and fighting pain causally is something different from preventing pain behaviour. If pain was nothing but a certain pattern of behaviour people should get rid of their pain by being paralysed⁷. Remember that in anaesthesia usually three different kinds of drugs are used: one that induces unconsciousness, another by which the muscles get paralysed, and, finally, a third against pain. On the pattern account it is completely incomprehensible what additional role the third kind of drug is supposed to play.

The strongest argument for the new orthodoxy, however, flows from the direct analysis of mental explanations. At least if one follows Mackie's or Lewis' considerations, these have exactly the features that are characteristic of causal explanations. For Mackie a cause of an effect *e* is, at least in principle, an INUS condition for *e* – an insufficient but necessary part of a condition which is itself unnecessary but sufficient for *e*⁸. And indeed, if someone says "John is holding his cheek because he has a strong toothache", we accept this explanation only if we are convinced that the toothache together with other conditions is sufficient for John's holding his cheek and if we furthermore believe that John would not hold his cheek if he had no toothache. If we learn that John would hold his cheek even if he had no ache

whatsoever simply because someone told him to do so, we would immediately reject the explanation given. This also is in complete accordance with David Lewis' counterfactual account of causation⁹.

Moreover, causal explanations provide an answer to the question of why a certain event took place at all. Pattern explanations, however, are not suited to this job. If saying "John is holding his cheek because he has a strong toothache" would just mean "John is holding his cheek, and this behaviour is part of a certain behavioural pattern characteristic of pain" then, after this explanation, we would still not know why John is holding his cheek, because the explanation remains silent on the question what leads to John showing this behavioural pattern. Mental explanations, however, *do* tell us why a person behaves the way he or she does. If I learn that John crossed the road in order to buy something at the grocer's, then I also learn why John did exactly this and not something else, that is to say, I learn why this behaviour took place at all. If the explanation given were a pattern explanation, I would not learn this. Hence, by simple counterposition, this explanation is obviously a causal and not a pattern explanation.

Resorting to the logical connection argument does not help here. As early as in the 60s it was shown with a great number of arguments that the logical connection argument is neither correct nor based on true premises. Davidson, for instance, has demonstrated convincingly that logical relations do not obtain between events, but between descriptions of events, and that for any two events one can always find descriptions that let them appear logically dependent as well as descriptions that make them logically independent. If *c* is the cause of *e*, *c* can always be described by the expression 'the cause of *e*', but of course this does not render the sentence "the cause of *e* is the cause of *e*" false. On the other hand, the considerations concerning the theoretical character of mental expressions have made especially clear that there is a connection that obtains between the mental states and the behaviour that they are meant to explain, but that this connection is by no means so close that it would preclude a causal relationship. And anyway: even if by definition the expression "streptococci infection" meant "infection caused by streptococci", this would not render false the sentence "Streptococci infections are caused by streptococci".

To sum up. It seems to me that in the dispute about the causal character of mental explanations at the end of the 60s and the beginning of the 70s the causalists clearly won the day. At least, this is how it appears among the academic public. Further, it seems to me that it was precisely this that has permitted the new orthodoxy to prevail so quickly over Wittgensteinianism. However, we should not forget to realise the consequences the most remarkable of which consists in the fact that the mind-body problem is on the agenda again. Maybe the question of whether there are mental things like souls or other spooky stuff is obsolete even now. But if mental properties are pretty normal and seemingly causally efficacious properties we cannot ignore the question of how these properties fit into an essentially physical world. And that is exactly the question which has been addressed by most of the work in the philosophy of mind in the last decades.

In my mind, we should also acknowledge that we really have made some progress here, at least in understanding the question itself. However, in the mid-70s something began to happen within the framework of the new orthodoxy which could well be regarded as a return to Cartesianism. The starting point was Thomas Nagel's seminal paper "What is it like to be a bat?". The considerations of Nagel, Jackson, Levine, Chalmers¹⁰ and many others all seem to point to the same result, namely, that at least phenomenal states have characteristic features that in the last consequence are not public, since they are neither tied to typical behaviours nor to causal roles. The idea of the philosophical zombie was born – the idea of a being that in all situations says exactly the same as I say, and does exactly the same as I do, but whose phenomenal states are – on this assumption – either connected with radically different qualia or with none at all.

It seems to me that those who claim that philosophical zombies are possible have strayed a step too far from Wittgensteinianism. For this assumption has a number of consequences that cast doubt on its coherence. One of these consequences has been very clearly spelled out by Levine himself in his 1997 paper "Recent Work on Consciousness". Suppose, there is a creature that has the same functional structure as me, but whose functional states are not connected with any

qualia. Suppose this creature is a functional zombie. According to the assumption, there are states within this creature – let’s call him ‘Zansgar’ – that play the same causal role as my sensations; but not only that, in this creature there are even states that play the same causal role as the beliefs that I have with regard to my sensations. Let us call these states Z-beliefs. Obviously, notwithstanding the differences, there is a great deal of similarity between my beliefs and Zansgar’s Z-beliefs. Zansgar’s Z-beliefs for example will make him say about himself: “Of course my phenomenal states are accompanied by certain qualitative experiences”, even if this is false according to the assumption. Let us assume further, that the states through which Zansgar’s Z-sensations are realised, are one-by-one replaced by states which not only play the right causal roles, but also are accompanied by the qualia belonging to that state. Would Zansgar notice any difference? Or Z-notice a difference? This does not seem to be the case since there is no change in the causal roles of his states. But how can one say under these circumstances that there is a significant difference between Zansgar and me? After all, by the reverse process I could be changed into Zansgar – without noticing anything.

What’s going on? The very intuitive stance, the first-person point of view, that fuelled the pro-zombie intuition, seems now to be undermining it. On the one hand, from within I seem aware of a feature of mental life whose absence I can so clearly conceive of in another. Yet, allowing for that absence in another seems to open up the possibility that its presence or absence makes no discernible difference, and that includes no discernible difference to me. (Levine 1997, 385)

If this is the case and there is no discernible difference whether a certain functional state is accompanied by a certain qualitative experience or not, how could it make sense to distinguish between functional states with and without this experiential quality? This truly seems to be an example of the Wittgensteinian wheel that is not part of the mechanism¹¹.

So, there are good reasons for following the new orthodoxy insofar as it holds the view that speaking about the mental is not different as a matter of principle from speaking about the physical and that therefore mental explanations can be regarded as causal explanations. However, one should become sceptical if the Cartesian picture of the mind creeps in through the back door. Wittgenstein’s Anticartesianism and that of his followers is an achievement we should not fall behind.

References

- Beckermann, A. 2000: “The perennial problem of the reductive explainability of phenomenal consciousness – C.D. Broad on the explanatory gap”. In: T. Metzinger (ed.) *Neural Correlates of Consciousness – Empirical and Conceptual Questions*, Cambridge MA: MIT-Press, 41-55.
- 2001: *Analytische Einführung in die Philosophie des Geistes*. 2nd ed., Berlin/New York: de Gruyter.
- Broad, C.D. 1925: *The Mind and Its Place In Nature*. London: Kegan Paul, Trench, Turbner, & Co.
- Carnap, R. 1936/1937: “Testability and Meaning”, *Philosophy of Science* 3, 419-471, and *Philosophy of Science* 4, 1-40.
- Chalmers, D. 1996: *The Conscious Mind*. Oxford: Oxford University Press.
- Feigl, H. 1950a: “The Mind-Body Problem in the Development of Logical Empiricism”. *Revue Internationale de Philosophie*. Repr. in: H. Feigl & M. Brodbeck (eds.) *Readings in the Philosophy of Science*, New York: Appleton-Century-Crofts 1953, 612-626.
- 1950b: “Existential Hypotheses: realistic versus phenomenalistic interpretations”. *Philosophy of Science* 17.
- Gean, W.D. 1965/1966: “Reasons and Causes”. *Review of Metaphysics* 19, 667-688.
- Glock, H.-J. 1996: *A Wittgenstein Dictionary*. Oxford: Blackwell.
- Hempel, C.G. 1952: *Fundamentals of Concept Formation in Empirical Science*. Chicago: The University of Chicago Press.
- Jackson, F. 1982: “Epiphenomenal Qualia”. *Philosophical Quarterly* 32, 127-136.
- Kim, J. 1996: *Philosophy of Mind*. Boulder CO: Westview Press.
- Levine, J. 1983: “Materialism and Qualia: The Explanatory Gap”. *Pacific Philosophical Quarterly* 64, 354-361.
- 1993: “On Leaving Out What It’s Like”. In: M. Davies and G.W. Humphreys (eds.) *Consciousness: Psychological and Philosophical Essays*. Oxford: Blackwell, 121-136.
- 1997: “Recent Work on Consciousness”. *American Philosophical Quarterly* 34, 379-404.
- Lewis, D. 1986: “Causation”. In: David Lewis *Philosophical Papers, Volume 2*. Oxford: Oxford University Press, 159-172.

- Nagel, T. 1974: "What is it like to be a bat?" *Philosophical Review* 83, 435-450.
- Mackie, J.L. 1965: "Causes and Conditions". *American Philosophical Quarterly* 2, 245-255.
- Ryle, G. 1949: *The Concept of Mind*, Hutchinson 1949 (Reprint: Penguin Books 1963).
- Sellars, W.S. 1948: "Realism and the new way of words". *Philosophy and Phenomenological Research* 8.
- ter Hark, M. 1995: „Wittgenstein und Russell über Psychologie und Fremdpsychisches“. In: Eike von Savigny and Oliver Scholz (eds.) *Wittgenstein über die Seele*. Frankfurt/M.: Suhrkamp, 84-106.
- von Savigny, E. 1994: *Wittgensteins "Philosophische Untersuchungen". Ein Kommentar für Leser. Band I*. 2nd, completely revised and augmented ed., Frankfurt/M.: Klostermann.
- Wittgenstein, L. 1953: *Philosophical Investigations*, Oxford: Blackwell.
- 1967: *Zettel*, Oxford: Blackwell.
 - 1969: *Blue and Brown Books*. Oxford: Blackwell.
 - 1980: *Remarks on the Philosophy of Psychology I*, Oxford: Blackwell.

* I would like to thank Antonia Barke for translating this paper into English. I am very grateful to the audience of my lecture in Bologna and to my colleagues in Bielefeld, especially to Eike von Savigny, Joachim Schulte, Hanjo Glock and Christian Nitz, for very helpful comments on earlier drafts of this paper.

¹ See ter Hark 1995.

² "Emergence" here is to be understood as C.D. Broad developed the concept in 1925: A property *F* of a complex system is emergent if it *cannot be deduced* from the properties of the parts of the system together with their spatial relations. See Beckermann 2000.

³ Kim 1996, 9.

⁴ For the following see Beckermann 2001, sec. 4.1.3.

⁵ See also Feigl (1950b) and Sellars (1948).

⁶ In recent times this line of reasoning has been pushed even further by the work of Kripke and Putnam on the semantics of natural kind terms. Nonetheless, I would like to stress that even this first change in the semantics of theoretical terms was an indispensable step on the way towards the identity theory. Only if we assume that these terms, one way or another, *refer* to states or properties, we can ask whether two of these terms refer to the *same* state or the *same* property.

⁷ Or, what seems even more absurd, by being moved to a different society.

⁸ Mackie 1965, 245-246.

⁹ See Lewis "Causation".

¹⁰ Nagel 1974, Jackson 1982, Levine 1983, 1993, Chalmers 1996.

¹¹ See *PI* § 271.