Wittgenstein and Sellars on Intentionality

Stefan Geoffrey Heinrich Brandt

University College, University of Oxford

Submitted for the degree of Doctor of Philosophy

Trinity Term 2011

Word count (excluding bibliography): 74 529

Acknowledgements

My greatest debt in writing this thesis is to John Hyman. He has been an excellent supervisor and has kept me constantly on my toes with a remarkable skill at raising objections that regularly took me completely by surprise. I am also very grateful to Robert Brandom who made it possible for me to visit the University of Pittsburgh in the Spring Term of 2010 and to study the Wilfrid S. Sellars Collection at the Archives for Scientific Philosophy, and who has been more than generous with his time during my visit, discussing my work with me every week.

Further thanks go to Bill Child for his valuable feedback during my transfer and confirmation of status and to Christian Kietzmann for valuable comments on the fourth chapter. I am grateful to my mother Emily Clarke-Brandt for proofreading chapters I, II and IV.

While writing this thesis I was financially supported by the AHRC, the Department of Philosophy at Oxford University and by my college. I also received a Hanfling Scholarship and I am grateful to the Department of Philosophy and the Hanfling family for this support.

Finally, I wish to thank Anna Flammiger for her support and forbearance.

Oxford, 4 September 2011

Contents

INTRODUCTION **3**

I. INTENTIONALITY IN THE TRACTATUS **10**

2. PICTURING 13
2.1 Ontology *15*
2.2 Language *16*
2.3 Philosophy of mind *25*
2.4 Intentionality *33*
3. INTERNAL RELATIONS AND CRITICISM OF RUSSELL 35
4. CONCLUDING REMARKS 41

II. THE LATER WITTGENSTEIN ON THOUGHT AND MEANING **43**

1. INTRODUCTION 43
2. THE LANGUAGE OF THOUGHT 44
2.1 Thoughts as the 'last interpretation' *46*
2.2 The location of thought *50*
3. THE DUAL-PROCESS CONCEPTION 55
3.1 Phenomenology *59*
3.2 Explanatory futility *62*
3.3 The metaphysics of meaning and understanding *70*
4. MEANING, USE AND PRACTICE 75
4.1 Meaning scepticism *84*
5. CONCLUSION 93

III. INTENTIONALITY **96**

1. THE REJECTION OF THE 'PICTORIAL RELATIONSHIP' 96
2. A NON-RELATIONAL ACCOUNT OF INTENTIONALITY 112
2.1 Sellars on meaning *115*
2.2 Sentences and facts *130*
2.3 The intentionality of the mind *140*
3. THOUGHTS AND OBJECTS 141
3.1 Crane's objection *153*
4. CONCLUSION 155

IV. SELLARS ON THINKING **157**

1. INTRODUCTION 157
2. THE MYTH OF THE GIVEN 161
3. CONCEPT EMPIRICISM 163
4. PSYCHOLOGICAL NOMINALISM 166
5. PERCEPTION AND SELF-KNOWLEDGE 172
6. THINKING 174
7. WITTGENSTEIN'S CRITICISM 199

CONCLUSION **203**

BIBLIOGRAPHY **210**

Introduction

It is one of the central features, if not *the* central feature, of our mental lives that we can represent things. We do so when we believe something, desire something, expect something, intend something and think or speak about something. All these phenomena are, as it is sometimes put, directed upon an object. The objects we represent are, of course, enormously diverse and may have nothing in common except being the object at which some representation is directed. They include concrete and abstract objects, real and fictional ones, as well as things that are not commonly called objects, such as events, properties, states of affairs and locations. The property of being directed at something, in the sense of representing that thing, is called *intentionality*. Possessing intentionality is the same thing as having a content or – in the case of linguistic representations – having a meaning.

There are a number of philosophical puzzles surrounding the phenomenon of intentionality. How do certain events states and objects, such as thoughts, beliefs and linguistic signs, come to possess intentionality? How is it that certain things in nature, such as the sounds and marks we produce when using language, come to represent things, while other things do not, such as the marks we leave in the sand when walking across it? How can different thoughts by different thinkers and different acts of speech in different languages be directed at the same object and have the same content? What is a content? Is it distinct from or identical with the things represented? Is it something abstract or concrete? Is it something to which our representations are related or is it something they instantiate or exemplify? What is the relation between intentional phenomena, such as beliefs, thoughts and utterances, and the objects, events and states of affairs at which they are directed? How can we explain that intentional phenomena can be directed at something that does not exist and represent something that is not the case?

In this essay I present the answers that Ludwig Wittgenstein and Wilfrid Sellars gave to at least some of these questions. In particular, I argue that Wittgenstein and Sellars both embrace what Sellars calls a 'non-relational' account of intentionality (cf. SM, ix): Intentional acts and states, such as utterances, thoughts and beliefs, should be understood in terms of their *role* in action, speech and reasoning rather than in terms of their *relation* to items in extra-mental and extra-linguistic reality.

The *first* chapter is devoted to the account of intentionality given in Wittgenstein's first work, the *Tractatus Logico-Philosophicus.* This account is clearly 'relational' in the sense criticized by Sellars and the later Wittgenstein. Accordingly, the purpose of this chapter is to form a backdrop against which the non-relational accounts of intentionality of Sellars and the later Wittgenstein can be seen more clearly.

The early Wittgenstein understood all intentionality in terms of linguistic intentionality and linguistic intentionality in terms of the notion of depiction. According to the *Tractatus* sentences picture states of affairs by being ultimately made up of simple names whose meaning are ontologically basic simple objects making up states of affairs. A specific sentence pictures a specific state of affairs if and only if the simple names (i) share a 'logical form' with the objects, which are their meaning, and (ii) stand in a 'pictorial relationship' to these objects, which is established through underlying acts of thought. These acts of thought are considered to be analogous to the sentences expressing them, consisting of mental signs corresponding to the signs of our public languages and belonging to what could be called a 'language of thought'.

The relational character of the Tractarian account of intentionality becomes apparent in Wittgenstein's conception of the 'pictorial relationship'. He thought of it as an internal relation between language and world, which is a necessary condition of sense and relates linguistic and other representational items to their content.

I begin the *second chapter* by discussing the later Wittgenstein's criticism of two fundamental mistakes one is prone to make when thinking about intentionality and linguistic meaning. These mistakes, which, as I show in the first chapter, constitute the core of the philosophy of mind behind the semantics of the *Tractatus*, are, first, the idea that thinking consists in the operation of mental signs within a language of thought, and second, in the idea that the meaningful use of language depends on underlying mental processes of meaning and understanding. I argue that the first idea is incompatible with two fundamental insights of the later Wittgenstein: (i) thinking, unlike speaking, cannot be interpreted; (ii) asking with what part of her body someone produced a thought, unlike asking with what part of her body someone produced linguistic signs, does not make sense. The second idea, i.e. the idea that the meaningful use of language depends on underlying mental processes of meaning and understanding, rests on a false conception of the nature of these phenomena. What we mean by what we say is determined by how we *use* the symbols we are employing, not by anything accompanying this use. Understanding, on the other hand, is not some kind of occurrence or process accompanying the use of signs either, it is an ability.

I close the chapter with a discussion of the later Wittgenstein's identification of meaning with use. In particular, I argue that both the meaning of sentence and word types as well as the meaning of particular tokens of these types on the occasion of their use is determined by their rule-governed use within a linguistic practice. It does not depend on underlying acts of thought.

The *third chapter* forms the core of this essay. It has three parts. In the first part I discuss the later Wittgenstein's criticism of the Tractarian 'pictorial relationship', which is supposed to relate intentional phenomena to the states of affairs they represent by connecting simple names with the simple objects in reality that are their meaning. I try to

show that according to the later Wittgenstein the idea of such a relationship rests on a confusion of the specification of the meaning of a linguistic expression with its use. When we use an expression to speak about an object, event or state of affairs we relate that expression to an item in the world. When we specify the meaning of an expression, on the other hand, we give a rule for its use by relating it to another expression or – if we give an ostensive definition – a sample, such as, for example, a coloured piece of cloth. However, we are not relating it to an item outside language that is its meaning. Because the early Wittgenstein failed to properly distinguish between the intra-linguistic relation between an expression, whose meaning is being specified, and another expression or sample used to specify this meaning, and the language-world relation between empirical sentences and the objects, events and states of affairs they describe, he was led to embrace the idea that there are necessary existing simple objects corresponding to the simple elements of language.

In the second part of the chapter I try to illuminate the later Wittgenstein's criticism of the pictorial relationship by discussing Wilfrid Sellars's 'non-relational' account of such meaning-specifying statements as '…means---' and '…stands for---'. According to Sellars these statements do not relate expressions to items that are their meaning but classify their role in 'language games'. I show that Sellars's analysis provides the resources to explain how sentential expressions can come to stand for facts and propositions and how sub-sentential expressions can come to stand for concepts, without committing us to a Tractarian or a Platonistic ontology.

I conclude the chapter with a discussion of the later Wittgenstein's account of the relationship between intentional acts and states, such as acts of speech, thoughts, beliefs and desires, and the objects at which they are directed. I suggest that according him this relationship is explained by the conceptual or 'grammatical' fact that we may use the expressions we use to specify the content of the intentional act or state to speak about the

object at which it is directed. It is not explained by an internal 'pictorial relation' between language and mind, on the one hand, and the world on the other.

I begin the *fourth chapter* by briefly addressing the fact that the later Wittgenstein never developed a philosophical theory of intentional mental acts and states. In particular he never developed a theory of thinking that could be said to replace the mistaken theory he seems to have presupposed when writing the *Tractatus*. I then turn to a fairly extensive account of Sellars's theory of occurrent thoughts. Sellars thinks that thoughts are causally efficacious inner goings-on which play a role in the 'game of reasoning', which is analogous to the role their linguistic expression plays in 'language games'. This account of thinking may at first sight appear to be similar to the Tractarian idea that thinking consists in the operation of mental signs and may therefore appear to be susceptible to the same kinds of objections. I conclude the chapter by arguing that Sellars's account of thinking is immune to the kinds of objections the later Wittgenstein levelled against the Tractarian idea that thinking consists in the operation of mental signs in a mental symbolism.

In the *conclusion* I briefly highlight a couple of points on which I have doubts about the validity of Sellars's theory. However, I do not take myself to be raising definitive objections to Sellars's views; I merely try to highlight some issues on which I think further research is required.

Finally I wish to point out a few of themes that are missing from this essay. *First*, limitations of space did not allow me to include a discussion of the historical influence of Wittgenstein on Sellars. A few remarks in this introduction will therefore have to suffice.

Although there is a significant amount of agreement between Sellars and the later Wittgenstein, Sellars's explicit discussions of Wittgenstein are mostly concerned with the *Tractatus* (cf. e.g. NS, SM, TC, NAO). It is therefore unclear to what extent the agreement

is coincidental and to what extent it is a matter of influence. Still, I think it is quite clear that Wittgenstein's influence is greater than Sellars tends to acknowledge in his writings. For, first Sellars knew Wittgenstein's later work well before the publication of the *Philosophical Investigations*. The Wilfrid S. Sellars Collection at the Archives for Scientific Philosophy at the University of Pittsburgh contains a copy of the *Blue Book and Brown Books* from Sellars's library, which Sellars signed and dated in 1937. Second, at least in the 1960s Sellars seems to have known Wittgenstein's later work well and gave graduate classes on it.[1] Third, at least sometimes, a mostly unacknowledged influence of the *Investigations* on Sellars is quite easily discernible in his writings.

A particularly interesting case is his 'Some Reflections on Language Games' (SRLG). In this essay Wittgenstein is not mentioned by name but merely alluded to in the title. Still, it reveals a profound appreciation of *Philosophical Investigations* by Sellars immediately after its publication. It was published in July 1954, a year after Sellars signed and dated his own copy of the *Investigations* (May 15th 1953). And it contains a detailed and sophisticated discussion of linguistic rules and rule-following, long before the importance of this topic was recognized in the wider philosophical community after the publication of Saul Kripke's *Wittgenstein on Rules and Private Languages* (1982; first publishes in Block 1981).

The *second* omission I wish to mention in this introduction is my failure to discuss Sellars's theory of picturing. Although Sellars rejected both the idea that an internal relation to items in the non-linguistic order is a necessary condition of sense for all sentences and the idea that sentences stand in a relation to their content, he still thought that it was a necessary condition for the possibility of *empirical* meaning that the signs that are used to express empirical propositions occur in patterns of use that are isomorphic with

[1] This was related to me by Jay Rosenberg.

the empirical goings-on they represent.[2] Only if this is the case, Sellars seems to think, can we project the signs we are using onto our environment, and only if we can do so, can we view them as carrying empirical content. I think there may well be some truth to this idea, but the details of Sellars's theory are so complicated and so closely related to a variety of other topics, such as predication, truth and the goals of scientific inquiry, that I have chosen to avoid discussing it.

Finally, I have to apologize for not discussing more secondary literature. The amount of primary literature I try to cover made it impossible to do this within the space available and forced me to take stances on interpretative issues more implicitly than I would have liked to.

[2] For Sellars's clearest exposition of this idea see his BBK.

I. Intentionality in the Tractatus

1. Introduction

Both the later Wittgenstein and Wilfrid Sellars embrace what Sellars calls a 'non-relational' account of intentionality (cf. e.g. SM, ix). Intentional acts and states, such as utterances, thoughts and beliefs, should be understood in terms of their *role* in action, speech and reasoning rather than in terms of their *relation* to items in extra-mental and extra-linguistic reality.

Understanding this view requires grasping the backdrop against which it acquires its plausibility. The present chapter is intended to provide this backdrop by discussing the *relational* account of intentionality the early Wittgenstein developed in his *Tractatus logico-philosophicus*. Turning to the *Tractatus* rather than giving a historical or systematic survey of the different forms relational accounts of intentionality have taken or might take, has two advantages. First, most of what the later Wittgenstein has to say about intentionality should be understood as a critical examination of his earlier views. Second, Sellars sees the theory of intentionality developed in the Tractatus as containing significant insights and as being clearly superior to other relational accounts of intentionality.

According to the so-called 'picture theory', which constitutes the account of intentionality provided in the *Tractatus*, mind and language represent reality by being isomorphic to it or, as Wittgenstein puts it, by sharing its 'logical form'. This idea of an isomorphism or shared 'logical form' between mind and language, on the one hand, and the world, on the other, is, according to Sellars, similar to a thesis found in Thomas Aquinas, namely that 'knowledge involves an isomorphism of the knower with the known' and that this isomorphism is a result of the fact that 'the senses in their way and the intellect in its way are informed by the natures of external objects and events' (BBK, 43). Sellars thinks this thesis compares favourably with the views of Cartesians and some early

20th century British and American realists, who think that mental acts do not differ intrinsically but only extrinsically by standing in different relations to either contents, as Cartesians believe, or extra-mental objects and states of affairs, as realists believe (cf. BBK, 41-2).[1] Sellars suggests that in the Cartesian tradition this latter view often led to idealism and scepticism as it became very difficult to explain the relation between the contents to which mental acts are related and which have only 'being-for-mind' or an 'objective' mode of being, and the objects, events and states of affairs in extra-mental reality at which they are directed (cf. ibid). Realists, on the other hand, are tempted to postulate a variety of ontologically dubious entities, such as subsistent but not existent unicorns and golden mountains, in order to account for the possibility of false representations (cf. ibid). Sellars sees it as a great advantage of the picture theory (and of Aquinas's conception of the mind) over other relational accounts of intentionality that it steers clear of these difficulties.

My exposition of the theory of intentionality of the *Tractatus* will proceed in three steps. I will begin by presenting the picture theory and its metaphysical presuppositions. I will then discuss the philosophy of mind which I take to be implicit in this theory. Finally I will turn to the early Wittgenstein's claim that the relation between picture and pictured is *internal* by discussing his criticism of an account of belief and desire Bertrand Russell

[1] It is rather unclear to whom Sellars is referring when he speaks about the 'early stages of contemporary British and American realism' in (BBK 41). The only name he mentions is Bertrand Russell's (ibid., 42); but it seems to me that he is primarily thinking about a – now mostly forgotten – group of American philosophers who cooperatively developed a form of direct realism in critical reaction to the pragmatic idealism of Josiah Royce and who called themselves 'the new realists'. The most influential members of this group (whom Royce called 'the six little realists'; cf. Kuklick 2001, 203) were Ralph Barton Perry, Edwin Bissell Holt and William Pepperrell Montague. (The others were Walter T. Marvin, Walter B. Pitkin and Edward Gleason Spaulding). The views of these philosophers, in particular as they are developed in their collectively written book *The New Realism* (Holt, *et al*. 1912), seem to fit the description Sellars gives of the realism he is criticising in BBK rather well. Furthermore it is likely that these philosophers figure somewhat more prominently in Sellars's mind than one would expect from a philosopher of his generation, since it was in reaction to their views that Sellars's father, Roy Wood Sellars, first developed his brand of 'critical realism' (first in his *Critical Realism* of 1916; cf. also R. W. Sellars 1969, 43-55); a view Wilfrid Sellars went on to refine in his own work on the philosophy of perception.
For a good discussion of the historical development that led from the 'new realists' to Sellars see Kuklick (2002) chapter 11. (I am grateful to Robert Brandom for making me aware of Sellars's relation to the 'new realists').

developed in his *The Analysis of Mind* (1921), and which can be found in the writings Wittgenstein produced shortly after his return to philosophy in 1929. This discussion will hopefully illuminate and vindicate Sellars's contention that the idea of an isomorphism between language and reality propounded in the Tractatus constitutes a significant advance over other relational accounts of intentionality.

Since Cora Diamond's seminal paper of 1988 'Throwing Away the Ladder: How to Read the *Tractatus*' a lot of attention has been paid to the meta-philosophical aim of the *Tractatus*, the status of its apparent metaphysical doctrines and its relation to Wittgenstein's later work. Much of the discussion has centred around the distinction drawn in the *Tractatus* between what can be said and what can only be shown. This resulted in a focus on the question of whether the early Wittgenstein was committed to a set of substantial views in metaphysics and the philosophy of language, which according to his own doctrines can only be shown but not expressed in meaningful propositions, or whether he practiced some form of 'philosophical therapy' which involved writing a mostly nonsensical piece of work.

Two or possibly three different strands can be discerned in the literature on the *Tractatus*.[2] *First*, there is the so-called 'traditional', 'orthodox' or 'metaphysical' reading whose most important contemporary proponent is Peter Hacker. According to this interpretation the author of the *Tractatus* tries to convey a variety of ineffable metaphysical truths and develops a theory about the nature of representation and the relation between language and world, which it is appropriate to call 'metaphysical'. According to the *second* 'anti-metaphysical' school Wittgenstein did not offer any theory about the connection between language and world, nor was he interested in making any pronouncements about the metaphysical structure of the latter, he merely investigated the logical structure of

[2] For this point and the subsequent distinctions see McGinn (2010, 495-96); cf. also McGinn (2006, ch. 1).

language. Prominent exponents of this school include Hide Ishiguro, Brian McGuinness, Rush Rhees, Peter Winch and more recently Marie McGinn. Finally there is the so-called 'resolute reading', a further variant of the anti-metaphysical reading, which claims that most of the sentences of the *Tractatus* are plain nonsense; their only purpose is the therapeutic aim of leading the reader away from the kind of confusions one is prone to when doing philosophy. The most prominent proponents of this view are Cora Diamond and James Conant.

I believe more than enough ink has already been spilled in this debate and I will therefore not engage in it in this chapter. Still, I cannot avoid taking sides. Any interpretation of central doctrines of the *Tractatus* will almost inevitably end up in one of the camps I have just characterized. As will become clear, my interpretation of the picture theory belongs in the 'traditional' or 'metaphysical' camp. However, I will not try to give a general defence of the traditional reading. The support, if any, this chapter gives to it is at most indirect: Insofar as my interpretation of the picture theory appears convincing, it will provide a reason for the kind of understanding of the *Tractatus* as a whole, by which it is informed.

2. Picturing

For the early Wittgenstein the capacity to be false was the most intriguing feature of intentional phenomena. Unlike in the case of true representations, there seems to be nothing in reality that corresponds to false representations. Since there does not appear to be any intrinsic difference between true and false representations, the possibility of false representations seems to require that *both* true and false representations are in some sense independent of the reality they represent. On the other hand, this *independence requirement* has to be reconciled with the early Wittgenstein's conviction that (a) what a true representation represents simply *is* the fact that makes it true, rather than, say, a

content in Frege's 'third realm' of ideas, and that (b) even false representations appear to be in some sense *about* reality. For we find out that they are false by finding out what is actually the case. False propositions like 'Lenin was an ardent supporter of the Tsar' are in this respect different from mere nonsense propositions like 'Green ideas sleep furiously'. As Wittgenstein sees it, only the former but not the latter can be compared to reality; only they, like true representations, can be 'laid against reality like a measure' (TLP 2.1512).

Wittgenstein tried to account for these phenomena with his so-called 'picture theory'. According to the *Tractatus* we represent reality to ourselves by picturing it (cf. TLP 2.1). Any representation is a picture of a possible state of affairs: 'A picture depicts reality by representing a possibility of existence and non-existence of states of affairs' (TLP 2.201).

Thus, understanding the nature of representation is approached through understanding the nature of depiction. For Wittgenstein this meant, in particular, *linguistic* depiction. Understanding the nature of intentional *mental* states or events, on the other hand, is, according to Wittgenstein, not part of philosophy but of psychology; a discipline he considered to be wholly distinct from the former (cf. TLP 4.1211). As a consequence he tries to stay clear – though not entirely successfully, as I will try to show – of issues belonging to the philosophy of mind and tries to explain the nature of intentionality entirely in terms of linguistic intentionality, i.e. in terms of linguistic picturing. He thereby adheres to a thought expressed in the preface of the work that it is only in language that we can draw a limit to what has content:

> [T]he aim of the book is to draw a limit to thought, or rather – not to thought, but to the expression of thoughts: for in order to be able to draw a limit to thought, we should have to find both sides of the limit thinkable (i.e. we should have to be able to think what cannot be thought).
>
> It will therefore only be in language that the limit can be drawn… (TLP, 3-4)

Hence, giving an account of intentionality consists in giving an account of how *language* represents the world.

According to the Wittgenstein of the Tractatus meaningful sentences can picture reality because they share a 'logical form' with the states of affairs they represent. Only because there is a fundamental isomorphism between language and reality is it possible for the former to picture the latter. Before I delve into the picture theory, I will, therefore, try to give a rough sketch of the ontology of the *Tractatus*. This will allow a clearer understanding of how language pictures reality by sharing its 'logical form'.

2.1 Ontology

Wittgenstein famously opens the *Tractatus* with an ontological remark about the fundamental constitution of the world: 'The world is all that is the case' (TLP 1). This claim he holds to be identical with the claim that the world is constituted by the totality of facts: 'For the totality of facts determines what is the case and also whatever is not the case' (TLP 1.12). Facts are then identified with existent states of affairs: 'What is the case – a fact – is the existence of states of affairs' (TLP 2).

The smallest ontological units are objects. They are non-composite, necessarily existing entities that make up the substance of the world (cf. TLP 2.021-2.0211). They enter into relationships with one another and thereby constitute facts or states of affairs. It is essential to objects that they can occur in certain states of affairs and actually do occur in one (cf. TLP 2.0121 and 2.0131). This essential or 'internal' property (cf. TLP 2.01231) of objects Wittgenstein calls their 'form' or 'logical form': 'The possibility of its occurring in states of affairs is the form of an object' (TLP 2.0141).

If we know all the objects we know all possible, but not all actual, states of affairs. 'If all objects are given, then at same time all possible states of affairs are also given.' (TLP 2.0124) Wittgenstein rejects the idea that reality is solely constituted by objects. The world is rather constituted by facts or states of affairs, i.e. combinations of objects: 'The totality of existing states of affairs is the world' (TLP 2.04).

Objects form the substance of the world. Their logical form determines the range of states of affairs in which they can occur; and the totality of actual states of affairs, i.e. the totality of facts, constitutes the world. Language is able to picture this world because its logical structure mirrors the ontological structure of reality.

2.2 Language

Wittgenstein tries to meet the demand that our representations be in some sense *independent* from the state of affairs they represent, which the possibility of false representations appears to make on any theory of intentionality, by stipulating that all intentional phenomena must have the capacity to be *both* true *or* false. In the case of language this means that any meaningful proposition must be *bipolar*, i.e. that it must be able to be either true or false. If a proposition is necessarily true or necessarily false it is, according to the *Tractatus*, a tautology or a contradiction, i.e. a proposition whose truth-value depends on its logical form alone. Logical truths and falsehoods, however, are senseless, they do not represent anything, i.e. they do not possess intentionality. They merely show us how logical connectives and quantifiers function in our language. Bipolar propositions, on the other hand, combine two apparently contradictory features Wittgenstein thought were essential and distinctive of all intentional phenomena: On the one hand their capacity to be true *or* false makes them independent from the state of affairs they represent. On the other hand, they are intimately related to the represented state of affairs. When they are true, what they say is *identical* with the fact that makes them true. And even when they are false, their content is *identical* with the state of affairs that would make them true if it obtained. Even though what a false proposition says is not the case, there is a sense in which it too reaches 'right out to [reality]' and can be 'laid against reality like a measure' (TLP 2.1511-2.1512).

The purpose of the picture theory is to explain how bipolar propositions can be, in the sense described, both independent of and essentially connected to the reality they represent. Before I turn to this theory, however, I will present two of its main elements, which were already in place before Wittgenstein arrived at its core idea.[3] First, Wittgenstein held that although the propositions '~p' and 'p' have an opposite meaning they represent the same reality. The very same fact that makes 'p' true makes '~p' false. To think otherwise, Wittgenstein believed, is to be misled by the surface structure of our language. The negation symbol does not stand for any object or property as the words 'Stravinsky' and 'red' stand for the composer of the *Firebird* and the favourite colour of philosophers. It does not signify any 'material function' (cf. TLP 5.44), which is what predicates do, but an operation (TLP 5.234). It reverses the sense of the sentence to which it is attached. 'P' expresses agreement with the fact that p, while '~p' expresses disagreement with the same fact. Thus, positive as well as negative propositions are directed at the same reality: 'The propositions 'p' and '~p' have opposite sense, but there corresponds to them one and same reality' (TLP 4.0621).

The second element is the idea that the way a proposition represents a state of affairs cannot be modelled on the way a name stands for an object. If sentences were name-tags for facts, it would be very unclear what a false or a negative sentence stood for. If we assume that a true proposition stands for a positive fact, we would have to assume that a false proposition would either stand for nothing, which according to Wittgenstein would make it meaningless, or would represent a negative fact. A negative fact, however, seems to be an ontologically dubious entity.

[3] Cf. Ammereller (2001b, 64-5); Hacker (1987, 56-7).

As we will see both these elements are of a piece with Wittgenstein's 'fundamental idea' or 'Grundgedanke' that logical constants do not represent (cf. TLP 4.0312). However, in order to see this connection we should now turn to the picture theory itself.

The core idea of the picture theory is of great simplicity: Propositions or sentences in use are pictures of reality.[4] In a notebook entry of 1914 (NB, 7) Wittgenstein compares the use of propositions with the use of toy cars to model traffic accidents, as it was apparently practice in Parisian courts of law at the time. The crash of a car and a truck, for example, was illustrated using a toy car and a toy truck. The toy truck represented the real truck and the toy car the real car and together they formed a complex state of affairs which mirrored the arrangement of real vehicles on the street. The representational relationship between the toys and the real vehicles Wittgenstein called 'pictorial relationship': 'The pictorial relationship consists of the correlations of the picture's elements with things' (TLP 2.1514). Toys and cars can be correlated with each other because they share a 'pictorial form'; the toy car enters into relationships with other toys that mirror the arrangements, in which cars can be found with other objects on the street. Both the actual accident and its 'toy-picture' are facts (cf. TLP 3.14). Hence, when the picture is true, the pictorial relationship is a relationship between facts.

Language, according to Wittgenstein, functions the same way. Sentences in use or propositions are truth-functional constructions out of elementary propositions. The elementary propositions are concatenations of names, which stand for the simple objects that form the substance of the world. Propositions are facts; just like the states of affairs they represent, when they are true. And just like the toy cars and real cars share a pictorial form, so do names and objects. Their pictorial form is what Wittgenstein calls their 'logical

[4] Wittgenstein calls sentences in use 'propositions'. When he abstracts from their employment as representations of states of affairs, he uses the term 'propositional sign'. At the level of the individual words he makes a similar distinction. There he distinguishes between 'signs', i.e. the 'sign-designs' (as Carnap calls them) which are employed in speech and writing, and 'symbols', i.e. words considered as bearers of meaning.

form'. If two entities share nothing more than their combinatorial possibilities they share a 'logical form'. The logical form is the minimum, picture and pictured must have in common. Thus, every picture shares at least a logical form with what it pictures: 'Every picture is at the same time a logical one. (On the other hand, not every picture is, for example, a spatial one)' (TLP 2.182).

Now we can better understand how propositions represent the world: A complex proposition pictures a complex state of affairs by being truth-functionally constructed out of elementary propositions that represent the elementary states of affairs that constitute the complex state of affair. The elementary propositions are concatenations of names, which stand for the objects that make up the simple states of affairs.

The connection between language and reality is made through the names occurring in elementary propositions. They, as ultimate simples of language, stand for, but do not picture, the ultimate simples of reality, i.e. simple objects. Only what is logically complex can be pictured, simples can only be named. Because the names share the logical form of the objects they represent, their combinatorial possibilities in propositions exactly match the combinatorial possibilities of the objects within states of affairs. The range of possible propositions is thereby limited to the range of possible states of affairs (cf. TLP 3.02). Every meaningful proposition, regardless of its truth, *must be* reducible to a set of elementary propositions which are concatenations of names. These names, in turn, *must* stand for objects (cf. TLP 3.328). Since the names share the combinatorial possibilities of the objects they stand for whatever the proposition says is possible, even if it is not actual.

We are finally in a position to appreciate the importance of Wittgenstein's *Grundgedanke* that logical connectives do not represent. If they did, then, if we, for example, negated a sentence, the resulting truth-functional compound would represent a different state of affairs containing one more object (namely the object corresponding to

the negation sign) than the negated sentence. Similarly, in the case of conjunction we would represent a state of affairs entirely different from those represented in the conjuncts, as it contains more objects than the sum of the objects spoken about in the conjuncts. As in the case of names, where every different name stands for a different object, the assumption that connectives represent leads to the idea that to every difference in truth-functional composition there exists a difference in extra-linguistic reality. It would be unclear how the truth and falsity of sentences could be made out by comparison with reality, since the negation of every false sentence would correspond to a different state of affairs than the negated sentence. However, given the *Grundgedanke* and the idea of a shared logical form of object and name, we can now appreciate how even the false sentence is connected to reality. Wittgenstein therefore concludes that even the false proposition is 'laid against reality like a measure' (TLP 2.1512).

A shared logical form of picture and pictured alone, however, does not suffice to fully determine which state of affairs is being pictured. In order for the picturing fact to enter into a pictorial relationship with the pictured state of affairs the elements of the picturing fact, in the case of language the names, have to be correlated with the objects in reality they stand for, i.e. the objects that make up the depicted state of affairs. 'These correlations', Wittgenstein claims, 'are, as it were, the feelers of the picture's elements, with which the picture touches reality' (TLP 2.1515). Only once these correlations are established does the picturing fact become a picture proper, for only then does it enter into a 'pictorial relationship' with a specific state of affairs. These correlations are therefore part of the picture (cf. TLP 2.1513).

To better understand the role of these correlations, it is useful to look at what Wittgenstein has to say about the inscriptions and sounds that constitute public languages. I

have already pointed out[5] that Wittgenstein distinguishes between 'propositional signs' and 'propositions' (cf. TLP 3.12) and between 'signs' and 'symbols' (cf. TLP 3.32). As I understand him, propositional signs are sentences simply considered as inscriptions or noises and signs are words simply considered as items of a specific sign design. Propositions on the other hand are sentences considered as bearers of meaning and symbols are the meaningful elements that constitute them. I take it that on its own the *use* of elementary signs and propositional signs only establishes their logical form. In order for a propositional sign to become genuinely meaningful the names occurring in it have to be correlated with specific objects. These correlations establish the 'pictorial relationship' between propositional sign and pictured state of affairs; and only when this relationship is established becomes the propositional sign a proposition (cf. TLP 3.12).

There are two reasons why these correlations are needed. First, a name is characterized by its logical form, i.e. its combinatorial possibility within a sentence. It shares its logical form with the object it deputizes for. This alone, however, does not determine the exact object the name stands for, as there could be many objects with the same logical form. Thus, if only the logical form of a simple sign is determined we get something akin to an unbound variable ranging over all those objects that share its logical form (cf. TLP 3.13). Therefore, in order for such a variable to become a proper name – and for the propositional sign to become a proposition – it has to be correlated with an object.[6] Second, when a linguistic picture is true, then both the picture and what it depicts are facts. So, the question

[5] cf. footnote 4.

[6] This reading doesn't fit well with the third sentence of TLP 3.13 which states that '[a] proposition… does not actually contain its sense, but does contain the possibility of expressing it'. This seems to suggest that the correlations between names and objects are not part of a proposition since a proposition only contains 'the possibility of expressing its sense' but does not actually express it. This, however, seems inconsistent with Wittgenstein's claim in TLP 3.12 that 'a proposition is a propositional sign in its projective relation to the world', which seems to state clearly that the correlations between names and objects are part of the proposition. The tension is removed if we substitute 'propositional sign' for 'proposition' in TLP 3.13, as Kenny suggests (1984, 4-5).

naturally arises, what makes one the picture of the other. It seems to be the fact that only the picture is 'projected' onto what it pictures (cf. TLP 3.11-3.13).

But what, then, is the nature of these correlations between the non-logical elements of the picture and the objects in reality they deputize for? How is the projection of a propositional sign onto a fact achieved? All the *Tractatus* says about this is the following:

> We use the perceptible sign of a proposition (spoken or written, etc.) as a projection of a possible situation.
> The method of projection is to think the sense of the proposition. (TLP 3.11)[7]

Thus, it is by a psychological act of thinking or meaning that the pictorial relationship, i.e. the connection between language and reality, is ultimately established. TLP 3.11 is one of the few passages were the psychological presuppositions of the *Tractatus* become apparent. Contrary to the views of the author of the *Philosophical Investigations*, who believed that the meaning of a proposition is fully determined by its rule-governed use, the early Wittgenstein assumed that only the 'logical form' of a sentence is determined by its use (cf. TLP 3.327) and further psychological acts of 'thinking the sense of the proposition' were needed to make a propositional sign meaningful. When at the beginning of the *Blue Book*, Wittgenstein describes the temptation to assume 'certain definite mental processes' behind any meaningful use of linguistic signs, without which these signs would have no 'life' (BB, 4), he may very well describe a temptation to which his earlier self succumbed:

> It seems that there are *certain definite* mental processes bound up with the working of language, processes through which alone language can function. I mean the processes of understanding and meaning. The signs of our language seem dead without these processes…Thus, if you are asked what is the relation between a name and the thing it names, you will be inclined to answer that the relation is a psychological one, and perhaps when you say this you think in particular of the mechanism of association. -- We are tempted to think that the action of language consists of two parts; an inorganic part, the handling of signs, and an organic part, which we may call understanding these signs, meaning them, thinking. (BB, 3)

[7] Following Peter Winch (1987) I corrected the translation from 'The method of projection is to think of the sense of the proposition to 'The method of projection is to think the sense of the proposition', which strikes me as closer to the German original ('Die Projektionsmethode ist das Denken des Satz-Sinnes').

However, interpreting the *Tractatus* as committed to this mentalistic picture of language is quite controversial. Interpreters like Cora Diamond and Peter Winch,[8] who see a fundamental continuity between the work of the early and late Wittgenstein, consider it to be severely mistaken. Accordingly, Winch suggests a very different interpretation of TLP 3.11, in particular its second sentence 'The method of projection is to think the sense of the proposition.' He contends that this sentence should not be understood as claiming (a) that a proposition only stands in a projective relation to the world if it is accompanied by a mental act of 'thinking its sense', but rather as saying (b) that thinking the sense of a proposition means using it in its projective relation to the world. According to Winch the method of projection is already fully determined by the use of a sentence and an additional mental act of meaning is not required. The second sentence of TLP 3.11 should not be interpreted as explaining the method of projection, which has already been explained by the preceding sentence, but as explaining what it is 'to think the sense of a proposition'.

Peter Hacker has forcefully objected to this interpretation.[9] Here I would like to highlight a couple of points that I think count against Winch's reading of TLP 3.11. First, Winch claims that the sentence in question is ambiguous between reading (a) and (b). I do not think that this is accurate as (b) constitutes a forced and unnatural reading of the German original. The German sentence begins with 'Die Projektionsmethode', which not only picks up the phrase 'als Projektion' from the previous sentence and is naturally read as explaining it, but which also indicates by its prominent position at the beginning of the sentence that it (i.e. the method of projection) is the thing to be explained by the sentence. Had it been Wittgenstein's intention to explain what it means to think the sense of a proposition rather than to explain the method of projection, he would almost certainly have used a different word order.

[8] Cf. Diamond (2002) and Winch (1987).
[9] Cf. (Hacker 2001, 170-90). See also Ammereller (2001a, 127ff).

Second, Winch cannot explain why Wittgenstein talks about the need to correlate the elements of pictures with the elements of the pictured states of affairs. He does not address at all the question of what distinguishes one name from another.[10] As we have seen, it is plausible to assume that there are many objects of the same logical form and that therefore a name that is not linked to a specific object through a mental act of meaning is akin to an unbound variable ranging over all the objects that share its logical form. However, if Winch assumes that there are no mental acts of meaning then the meaning of a name has to be determined by its logico-syntactical role alone. That means every different name must have a different logico-syntactical role. And even if this might not be completely out of the question, it has the highly problematic consequence that we cannot speak about any two objects in the world which have the same logical form; for every meaningful name stands for an object which shares its logical form and if the logical form of every name we use is different, then so must be the form of every object we speak about. This seems to directly contradict TLP 2.0233, which appears to suggest that there can be different objects with the same logical form: 'If two objects have the same logical form, the only distinction between them, apart from their external properties, is that they are different.'

I believe these arguments make it sufficiently clear that Winch's interpretation of TLP 3.11 is incorrect and that I am right in claiming that the *Tractatus* assumed mental acts of meaning or thinking that connect language with reality.[11] Wittgenstein's claim that there is a relation of projection which connects propositional signs with the states of affairs they represent should be seen as one of the few places in the *Tractatus* where the psychological presuppositions of the picture theory become apparent. I will, therefore, use it as my point of departure for a brief exploration of the philosophy of mind behind the account of

[10] This is pointed out in (Hacker 2001, 171-2).
[11] This interpretation is also advocated in Ammereller (2001a), Hacker (2001, 170-90), and Malcolm (1986).

intentionality of the *Tractatus*. The focus of this exploration will be the acts of thinking that are said to project propositional signs onto reality.

2.3 Philosophy of mind

The psychological hinterland of the *Tractatus* is notoriously obscure. The claim that the pictorial relationship between linguistic pictures and the states of affairs they represent is established through mental acts of 'thinking the sense of the proposition' does not tell us much about the philosophy of mind underpinning the tractarian account of intentionality. We do not know anything about the nature of these mental acts, about their structure and about their connection to reality.

The *Tractatus* says very little about thinking (or 'Denken' in German) but there are some remarks on the nature of thought ('Gedanke'), maybe some insight can be gained from those. Hans-Johann Glock thinks we can find three apparently disparate notions of thought (or 'Gedanke') in the picture theory:[12]

(a) In TLP 3 Wittgenstein states that a thought is 'a logical picture of facts', i.e. it is a picture whose only pictorial form is logical form (cf. TLP 2.181-2.19). Glock claims this suggests that thoughts are 'optimally abstract pictures', conceived independently of any specific medium of representation.

(b) TLP 3.5 states that a thought is a propositional sign, which is 'applied and thought out'. And TLP 4 states correspondingly that a '[thought] is a proposition with a sense'. According to Glock, this suggests that a thought is a sentence in use.

(c) However, Wittgenstein also writes that '[i]n a proposition a thought finds an expression that can be perceived by the senses' (TLP 3.1), which, Glock claims,

[12] For these distinctions cf. Glock (2006, 141).

> suggests that a thought is a mental occurrence which is isomorphic with the proposition expressing it.

Glock suggests that in all these instances Wittgenstein is talking about mental acts of thinking within a language of thought.[13] He thinks TLP 3 suggests that a thought is a mental act within a maximally abstract mental language that does not rely on any specific medium of representation and is shared by all competent speakers, regardless of what language they speak. Accordingly, he interprets TLP 3.5 as speaking about a propositional sign in the language of thought; that is why Wittgenstein speaks of a 'propositional sign, applied and *thought out'*. And on the assumption that TLP 3.1 speaks about propositions in a public language Glock reads it as suggesting that 'the logical analysis of propositions in any sign-language will reveal the structure of the underlying proposition in the language of thought' (Glock 2006, 141).

I agree with Glock that the early Wittgenstein embraced the idea of a language of thought, but I do not think he speaks about it in the remarks Glock discusses, nor in any other passage of the *Tractatus* which mentions *Gedanken* (thoughts). The noun 'thought' and its German equivalent 'Gedanke' are commonly used in a *psychological* or a *semantic* sense. In its psychological sense it refers to acts of thinking; in its semantic sense, in which Frege famously uses it, 'Gedanke' refers to the content of these acts. Contrary to Glock I believe that within the *Tractatus* Wittgenstein uses 'Gedanke' fairly consistently in a way which is more akin to its 'Fregean' use than to its psychological use.[14]

I am suggesting that Wittgenstein uses 'Gedanken' in a sense that is *akin* to its Fregean sense rather than that he uses it *in* its Fregean sense, because there are significant

[13] Glock (2006, 140-43).
[14] A similar suggestion has recently been made by Joachim Schulte (2010, 330-34), who otherwise gives a very different interpretation of the passages of the *Tractatus* I am discussing, as he does not believe that there is a philosophy of mind of the *Tractatus*. I hope this chapter succeeds in making this contention implausible.

differences between Frege's and Wittgenstein's conception of thoughts. First, Frege took *Gedanken* to be abstract objects existing independently of the propositions expressing them. The *Tractatus* is clearly not sympathetic to this view. Wittgenstein believes that when '[we] picture facts to ourselves' (TLP 2.1) by means of propositions then there exist only the propositions and the facts they depict and not some third thing, an abstract object, mediating between the two.

Second, for Frege *Gedanken* are the senses of propositions. Wittgenstein, on the other hand, frequently uses 'sense' or 'Sinn' in a sense, which resists this identification. For example, in TLP 2.201 he states that '[a] picture represents a possible situation in logical space', and then – a few remarks later – goes on to say: 'What a picture represents is its sense' (TLP .2.221). This suggests that for Wittgenstein the sense of a proposition is the possible situation or state of affairs it represents. This is confirmed by TLP 4.03, where Wittgenstein first says that a proposition communicates a sense and then goes on to say that what a proposition communicates is a situation with which it is essentially connected. Wittgenstein's conception of 'sense' as equivalent to 'state of affairs' or 'situation' comes out most clearly in TLP 4.1211:

> [One] proposition 'fa' shows that the object a occurs in its sense, two propositions 'fa' and 'ga' show that the same object is mentioned in both of them.

The talk about 'objects' occurring in the sense of propositions shows that 'sense' is not used to refer to something that represents, but to the thing represented. However, when Wittgenstein says in TLP 3 that a *Gedanke* is a 'logical picture of facts', he uses 'Gedanke' clearly to refer to something that represents, i.e. a picture, and not to refer to something represented, i.e. a state of affairs. Hence, in this sense of 'sense', thoughts or *Gedanken* are not senses.

However, Wittgenstein also states in TLP 2.222 that the 'agreement or disagreement' of the sense of a picture 'with reality constitutes its truth or falsity'. And in TLP 4.0621 he says: 'The propositions 'p' and '~p' have opposite sense, but there corresponds to them one and the same reality'. Here he does not use 'sense' to refer to states of affairs but rather seems to speak about the properties of propositions in virtue of which they picture whatever they picture.

I believe Wittgenstein's use of 'Gedanke' is akin to Frege's because in this second sense of 'sense' or 'Sinn', 'having a sense' means the same as 'expressing a thought'. This is because Wittgenstein uses 'Gedanke' or 'thought' in the *Tractaus* to speak about what all pictures that picture the same state of affairs, i.e. that have the same sense in both senses of 'sense', have in common; namely a specific logical form and a specific pictorial relation to a state of affairs.

This interpretation of 'Gedanke' makes all the remarks Glock highlights perfectly compatible. First, when Wittgenstein says that a *Gedanke* is 'a logical picture of facts' (TLP 3), he is not saying that a thought is a sentence in an 'optimally abstract mental language' (Glock 2006, 140), but that it is whatever all those pictures picturing the same fact have in common. Second, TLP 3.5 does not speak about a proposition in the language of thought when it says that an 'applied, thought, propositional sign is a thought'. The term 'propositional sign' was introduced in TLP 3.12 to refer to the signs of a public language and it would be odd if Wittgenstein suddenly used it to refer to the signs of a language of thought. Furthermore it would be in clear violation of his claim in TLP 4.1121 that he is studying sign-languages rather than thought-processes. What he is saying in TLP 3.5 is that a propositional sign which is projected onto reality, i.e. a proposition, is an instance of a thought. Since every picture is also a logical picture (cf. TLP 2.182) and a thought is a logical picture this should not be surprising.

Finally, the claim in TLP 3.11 that '[i]n a proposition a thought finds an expression that can be perceived by the senses', does not have to mean that in language an inner episode finds overt expression. The notion of 'expression' in play here could be the same as the one we employ when we say that a sentence *expresses* a certain idea, which just means that it has a specific meaning.[15]

Where does this leave us in our aim to excavate the philosophy of mind and in particular the conception of thinking behind the *Tractatus*? It seems that all we can learn from the *Tractatus* itself is that sign languages have to be accompanied by acts of thinking or meaning in order to be meaningful; the nature of these acts remains unexplained. However, Glock is right in attributing to the early Wittgenstein the idea that thinking is a kind of language. We get a glimpse of Wittgenstein's view when we look at TLP 4.1121 where he asks rhetorically 'Does not my study of sign-language correspond to the study of thought-processes…?', and at TLP 5.542 where he states that '[i]t is clear,…, that 'A believes that p', 'A thinks that p', and 'A says p' are of the form '"p" says p'…',[16] but it only becomes fully apparent when we look at the *Notebooks* and a letter to Russell. On September 9th 1916 Wittgenstein made the following notebook entry:

> Now it is becoming clear why I thought that thinking and language were the same. For thinking is a kind of language. For a thought too is, of course, a logical picture of the proposition, and therefore it just is a kind of proposition. (NB, 82)

[15] It should be noted that in his earlier *Wittgenstein Dictionary* (Glock 1996) Glock suggests that in the *Tractatus* Wittgenstein fluctuates between the semantic and the psychological senses of 'thought' (Glock 1996, 357-58). However, his only example of the psychological use of 'thought' stems from a letter to Russell (discussed below) and thus does not count against my claim that *within* the Tractatus Wittgenstein uses 'thought' consistently in its semantic sense.

[16] I corrected Pears and McGuinnes translation from '"A has the thought that p"' to '"A thinks that p"' which is closer to the German original ('"A denkt p"') and corroborates my suggestion that the *noun* 'thought' in the Tractatus should always be understood in its semantic sense.

Here he uses 'thinking' (*Denken*), which suggests that he is speaking about thoughts in the psychological and not the semantic sense. However, as Winch points out,[17] the notebooks have to be used with care. They record a wrestling with the problems Wittgenstein thought he had finally dealt with when he wrote the *Tractatus* and when they contain ideas which do not occur in the published work this may be because he had abandoned them. Still, this does not seem to apply to his views on thinking. In a letter to Russell from August 19th 1919, which Wittgenstein wrote after the completion of the *Tractatus,* he repeats his claim that thinking is a kind of language:

> I don't know *what* the constituents of a thought are but I know *that* it must have such constituents which correspond to the words of Language [*sic*]. Again the kind of relation of the constituents of the thought and of the pictured fact is irrelevant. It would be a matter of psychology to find out.
>
> "Does a Gedanke consist of words?" No! But of psychical constituents that have the same sort of relation to reality as words. What those constituents are I don't know. (NB, 130-31)

The talk about 'psychical constituents' and 'psychology' indicates that Wittgenstein is speaking about *Gedanken* in the psychological sense.[18] If we look at these remarks together with the statement in the *Tractatus* that acts of thinking connect propositional signs in use with the facts they represent, we get at least a rough idea of the early Wittgenstein's conception of thinking and its relation to public language: Thoughts are occurrences in a mental language which are constituted by mental signs corresponding to the signs of public languages and give the latter their meaning by connecting them with the facts they are about. They underlie our use of sign-languages and establish what one might call the *intentional relation* between a proposition and what it represents, by establishing the pictorial relationship between propositional sign and state of affairs.

---

[17] Cf. Winch (1987, 4).

[18] This could, of course, be seen as suggesting that TLP 3, 4 and 4.001 also speak about *Gedanken* in the psychological sense, as Russell's questions concern these passages (cf. Russell letter of 13.08.1919; CL, 124-26). This would cast doubt on my suggestion that *within* the *Tractatus* 'Gedanke' is used only in its semantic sense. However, I do not think this is strong evidence. Wittgenstein points out in his letter, which he wrote while detained in an Italian POW camp, that he did not have a copy of the *Tractatus* and could not check which passages Russell was referring to. Furthermore, in his letter Russell uses 'Gedanke' quite clearly in its psychological sense.

But how is the intentional relation between a thought and fact established? If a thought is conceived as a proposition in a language of thought that is equivalent to a proposition in a public language, just constituted by psychical rather than acoustical or visual elements, then it seems that its elements too have to be correlated with the elements of the represented fact through a relation of projection. However, if this relation is established, as in the case of public languages, by acts of 'thinking the sense of the proposition' (TLP 3.11), we are faced with a vicious regress. In order for an act of thinking to represent a state of affairs (i.e. to possess intentionality) it would have to be accompanied by a further act of thinking its sense, and this act of thinking would have to be again accompanied by a further act of thinking and so on *ad infinitum*. Vexed by this problem Anthony Kenny suggests that the elements of a thought have their meaning conferred on them by the 'metaphysical I' of TLP 5.641, the *I* or *self* that 'does not belong to the world' (5.632).[19] Kenny says:

> In the thought itself, perhaps, we can distinguish between the particular mental configuration, studiable by psychology, and the significance or intentionality of that configuration, conferred by the metaphysical self. Thought… will have the right metaphysical multiplicity to depict the facts; but its multiplicity gives it only the *possibility* of depicting; that it actually does depict depends on the meaning of its elements, and that is given by the extra-psychological will giving those elements a use, an application. (Kenny 1984, 8-9)

I do not think it can be made intelligible how the metaphysical self, which lies outside of the world, can give something within the world, namely a thought-element, a 'use' or 'application'. It strikes me as quite inappropriate to say of the elements of a thought that they have a use or application. Nothing gives our thoughts their meaning or intentionality. As Wittgenstein points out in TLP 3.323 one propositional sign can have different meanings in different contexts; however if one thought has different intentional content than another it simply is a different thought. The idea of 'conferring' meaning onto

[19] The same suggestion is made by Peter Hacker in Hacker (1986, 75 & 100-104).

thought-elements makes no sense, for thoughts cannot be individuated independently of their content. As Wittgenstein later put it, a thought cannot be variously interpreted for 'it is the last interpretation' (cf. BB 34).[20]

In the light of these difficulties Norman Malcolm suggests that the early Wittgenstein believed that the simple elements of our thoughts are *intrinsically* intentional, i.e. are intrinsically connected to the simple elements in reality they stand for.[21] The whole thoughts, which are constituted by these intrinsically intentional thought-elements, are intrinsically intentional as well. Acts of thought, according to this interpretation, are structurally analogous to fully analyzed propositions, but unlike the latter do not derive their intentional content from further underlying intentional phenomena but possess it independently of anything else, or, in other words, *intrinsically*.

The problem with Malcolm's solution is that it merely labels the difference between acts of thought, on the hand, and sentences of public languages, on the other, without explaining it. Saying that the *intentional relation* between thoughts and the states of affairs they are about is intrinsic and accordingly, unlike the intentional relation between propositional signs and states of affairs, not mediated by a further intentional act, does not explain how thoughts come to represent. And since according to the *Tractatus* the intentionality of propositions is dependent on the intentionality of the underlying acts of thought, Malcolm's solution also fails to explain how language acquires its intentionality.

I think the awkwardness of both Kenny's and Malcolm's proposal is due to the fact that they are trying to provide a solution within the tractarian framework to a problem Wittgenstein did not see when writing the *Tractatus*. Wittgenstein simply thought that by giving an account of how language represents reality he *ipso facto* provided an account of how acts of thought represent reality. He did not notice that his picture theory of language

[20] For the above critique of Kenny see Malcolm (1986, 77-78).
[21] Malcolm (1986, 75-76).

relied on the availability of an independent account of intentionality of the acts of thought that are supposed to establish the 'pictorial relationship' between picture and pictured.[22]

2.4 Intentionality

Before I come to the next section of this chapter I will now briefly recapitulate what we have learned so far about the account of intentionality developed in the *Tractatus*. In line with his pronouncement in the preface that it is only in language that a limit can be drawn to what has content, Wittgenstein tries to understand all intentional phenomena in terms of linguistic representations. Propositions possess intentionality because they are pictures of possible states affairs. Any meaningful proposition is either an elementary proposition and consists of a concatenation of names, or it is truth-functionally constructed out of elementary propositions. The 'pictorial relationship' between propositions and the states of affairs they depict is established by the names that occur in the elementary propositions. As the ultimately simple elements of language these names stand for but do not depict the ultimate simples of reality, namely the sempiternal objects that, according to Wittgenstein, 'make up the substance of the world' (TLP 2.021). The names establish this pictorial relationship between picturing proposition and pictured state of affairs (a) by being projected through acts of thinking, or, as we might say, meaning, onto specific objects within the represented state of affairs and (b) by sharing the 'logical form' of the objects they stand for; which means that their combinatorial possibilities within propositions exactly match the combinatorial possibilities of the correlated objects within states of affairs.

Insofar as the early Wittgenstein had views in the philosophy of mind, he conceived of the intentionality of the mental as analogous to the intentionality of language. It is

[22] It should be noted that Kenny and Malcolm are aware of the philosophical shortcomings of the views they attribute to Wittgenstein. My point here is merely that it is a questionable interpretative strategy to ascribe to an author with hardly any textual evidence a philosophically dubious solution to a problem he might not have seen.

impossible to find in the *Tractatus* an account of propositional attitudes, but as we have seen there is implicit in it an account of occurrent thoughts. Acts of thought are the occurrence of propositions in a mental language of thought. It seems to have escaped the early Wittgenstein that he did not have an answer to the question of how these propositions in the language of thought are projected onto reality. As I will show in the next chapter, this oversight is significant, for it is one of the main reasons for his later rejection of the idea that thinking occurs in a mental symbolism.

However, the author of the *Tractatus* succeeds in explaining both how propositions can be sufficiently independent from reality to allow for the possibility of false propositions, and how they can nevertheless be directly about the states of affairs they represent. On the one hand, the shared *logical form* of names and objects, i.e. the fact that the combinatorial possibilities of names in propositions exactly matches the combinatorial possibilities the objects they deputize for have within states of affairs, ensures that propositions are *bipolar*. Every true proposition represents a combination of objects that could have been otherwise and every false proposition represents a combination of objects that could have occurred. On the other hand the fact that every name stands in a *pictorial relation* to an object in reality ensures that even the false sentence 'reaches right out to [reality]' (TLP 2.1511); for it speaks about objects that actually do occur in reality, just not in the combination depicted by the false proposition.

In the next section of this chapter I will try to illuminate this last claim that the pictorial relationship between picture and pictured ensures that the picture 'reaches right out to [reality]'. A claim Wittgenstein seems to repeat when he says that the pictorial relationship is an *internal relation* (cf. TLP 3.24) and that the correlations between names and objects that constitute it 'are, as it were, the feelers of the picture's elements, with which the picture touches reality' (TLP 2.1515). I think the best way to illuminate these

metaphors is by looking at Wittgenstein's criticism of a theory of belief and desire Bertrand Russell developed in *The Analysis of Mind*.[23]

3. Internal relations and criticism of Russell

In order to fully appreciate Wittgenstein's claim that the relation between an intentional state and the state of affairs it represents is *internal*, we should first clarify what he has in mind when he speaks about internal relations. An internal relation is usually considered to be a relation that is in some sense constitutive of or essential to its relata or, if that amounts to anything different, at least holds necessarily between them. F.H. Bradley, for example, with whose conception of internal relations Wittgenstein may have been familiar through the mediation of his teachers Russell and Moore and who notoriously believed that all relations were, in his words, 'intrinsical', characterizes internal relations as follows:

> 'A relation must at both ends affect, and pass into, the being of its terms' (Bradley 1893, 364).
>
> 'Every relation essentially penetrates the being of its terms, and is, in this sense, intrinsical' (ibid., 392).

These remarks suggest that for Bradley internal relations were essential to or constitutive of its relata. Wittgenstein, on the other hand, only claims that internal relations hold necessarily between their relata. In TLP 4.123 he discusses internal relations in the course of an attempt to define the notion of an internal property:

> A property is internal if it is unthinkable that its object should not possess it. (This shade of blue and that one stand, eo ipso, in the internal relation of lighter to darker. It is unthinkable that these two objects should not stand in this relation.)

The sudden shift from talk about internal properties to talk about internal relations in this quote is irritating. However, it does seem fairly unproblematic in the light of Wittgenstein's use of the notion of an internal relation in the rest of the *Tractatus.* As far as

[23] Russell (1921); see especially chapters I, III, X, and XII.

I can see, whenever he talks about internal relations he is really concerned with *internal relational properties*.[24] In particular when he claims that pictures stand in an internal pictorial relationship to what they depict, what he seems to be saying is that it is a necessary property of any picture that it stands in a specific relation to a state of affairs. After all, he also claims that every picture 'includes the pictorial relationship' to the state of affairs it depicts and that only once this relationship is established does it become a picture (cf. TLP 2.1514).

Furthermore, it seems reasonable to assume that unlike the internal relationships Bradley talks about the pictorial relationship neither 'penetrates the being' of both of its relata, nor is it a necessary property of both of them, but is only a necessary or maybe essential property of the picture, not of the pictured. It may, for example, be a necessary or essential property of a photograph of Bill Clinton that it stands in a specific (causal) relationship to him, but it is not a necessary property of Clinton nor does it 'affect his being', that a specific photograph has been taken of him. When I talk about 'internal relations' in what follows I should, therefore, be understood as talking about internal relational properties, i.e. relational properties of an object or state of affairs, in which that object or state of affairs stands necessarily.

In the *Tractatus* Wittgenstein mentions two kinds of internal relations. The first mention of 'internal relation' is in TLP 3.24, where he says that there is an internal relation between a proposition about a complex and a proposition about a constituent of the complex: 'A proposition about a complex stands in an internal relation to a proposition about a constituent of the complex'. Since propositions are facts this statement amounts to the claim that a complex fact stands in an internal relation to the simple facts that constitute it. Examples of these sorts of internal relational properties are easy to imagine. The

[24] The difference between a relation and a relational property can best be explained by means of an example: *Fatherhood* is a relation, whereas *being the father of Susie* is a relational property. For the difference between relations and relational properties cf. Moore (1919, 44-45).

complex fact of George and Maria dancing stands in an internal relation to the simpler fact of Maria dancing. Without Maria dancing the complex state of affairs of George and Maria dancing could not obtain. Furthermore there is a necessary relation between the corresponding complex proposition 'George and Maria dance' and its truth-functional component 'Maria dances'; from 'George and Maria dance' its truth-functional component 'Maria dances' follows necessarily.

The pictorial relationship between picture and pictured, however, is a different kind of internal relationship. It is a relationship that holds not between a complex fact and its constituent facts but between two equally complex facts that share a logical form. In TLP 4.014 Wittgenstein claims that the pictorial relationship is an internal relationship between facts of 'a common logical pattern', a phrase that should be understood as equivalent to 'shared logical form':

> A gramophone record, the musical idea, the written notes, and the sound-waves, all stand to one another in the same internal relation of depicting that holds between language and the world. They are all constructed according to a common logical pattern.

In order to appreciate the significance of the claim that the pictorial relationship between picture and pictured is internal, it is helpful to contrast it with a view that takes this relationship to be external. A good starting point for this is to look at Wittgenstein's criticism of the causal theory of belief and desire Russell developed in *The Analysis of Mind*. Briefly after his return to philosophy and at a time at which he still believed in the picture theory[25] Wittgenstein criticized this theory sharply. His critique can be found in lectures he gave at Cambridge in the Lent term of 1930 (LWL) and in writings of the same time that were later published in his *Philosophical Remarks* (PR).

[25] For this point see Hacker (1996, 30) and Kenny (2006, 96-97). See also Wittgenstein's talk about pictures in (PR, 57 & 63) and (LWL, 1 & 4).

In *The Analysis of Mind* Russell conceived of the relationship between a belief and the state of affairs it is about and between a desire and the fact that fulfils it as an *external* causal relationship. According to *The Analysis of Mind* a conscious desire is accompanied by a belief as to its purpose, that is, the state of affairs that will lead to its fulfilment. The fulfilment of a desire consists in quiescence or the cessation of discomfort. Our belief about what will bring quiescence is a hypothesis based on inductive inference from earlier cases of the same feeling of desire. A desire is a feeling of discomfort (like hunger) which causes a 'behaviour cycle' which terminates in quiescence. We know about our desires in the same way we know about other people's desire, namely by inference from past patterns of behaviour. Similar distinctions apply to belief: Russell distinguishes what is believed, a current state of the believer, from the objective of our belief, a fact or event in the extra-mental world. What I believe is not the same as the fact that makes my belief true. After all, it might seem as though I could be spatially or temporarily far removed from the fact or event that makes my belief true, but I cannot be spatially of temporarily removed from what I believe.

Russell connects desires and beliefs and what fulfils or verifies them through a *tertium quid*: I verify a belief if I feel that I recognize its objective, i.e. the fact that makes it true, I fulfil my desire if I achieve quiescence. Wittgenstein thinks that this view leads to a dilemma. We are led to the *first* horn of the dilemma, if we assume that for every belief and every desire we have to be aware of our feeling of recognition or satisfaction as the correct feeling for that belief or that desire. This horn consists of an infinite regress: If the verification of my belief that *p* is accompanied by a feeling of recognition, how do I know that it is the right sort of feeling and not, for example, the feeling that accompanies the verification of my belief that *q*? It seems that I have to recognize my feeling of recognition as the kind of feeling of recognition that accompanies the verification of the belief that *p*

rather than the belief that *q* or any other belief. But this second act of recognition is itself a feeling of recognition and has again to be recognized as the feeling of recognition for the feeling of recognition that accompanies the verification of my belief that *p* rather than the feeling of recognition for the feeling of recognition that accompanies the verification of my belief that *q*; and so on *ad infinitum*. In his Cambridge lectures of Lent term 1930 Wittgenstein makes this point by speaking about expectations and their fulfilment:

> On Russell's view you need a tertium quid besides the expectation and the fact fulfilling it; so if I expect x and x happens, something else is needed, e.g. something that happens in my head, to link expectation and fulfilment. But how do I know that it is the right something? Do I, on the same principle need a fourth something? If so, we have an infinite regress, and I can never know that my expectation has been fulfilled. (LWL, 9)

We are led to the *other horn* of the dilemma if we abandon the idea that there are feelings of recognition or satisfaction that are correct or incorrect for a specific belief or desire and that have themselves to be recognized as correct or incorrect. We then have to accept whatever gives us some not further specified feeling of recognition or satisfaction as the objective of our belief or desire. But this seems wrong; if I wish for a steak and quell my feeling of hunger by eating pasta, I do not fulfil my desire for a steak. And if I expect to receive a letter but I get an e-mail instead and I am satisfied with it, my expectation of a letter will not have been fulfilled. Wittgenstein expresses this line of thought in the following remarks:

> I believe Russell's theory amounts to the following: I give someone an order and I am happy with what he then does, then he has carried out my order.
> (If I wanted to eat an apple, and someone punched me in the stomach, taking away my appetite, then it was this punch that I originally wanted.) (PR, 64)

From a *Tractarian* point of view, the core of the problems with Russell's causal theory of belief and desire is that causal relations are external whereas the *pictorial relationship* between intentional phenomena and the state of affairs they are representing is internal. What we believe is, according to Russell, different from the objective of our belief. For

example, the objective of my belief that the Eiffel-Tower is the highest building in Paris is a fact obtaining in the French capital which is far removed from me and my current state of belief. Russell, therefore, concluded that what I actually believe is a *surrogate* of the objective, namely a content existing within my mind. And the relation between the surrogate and the actual fact we can only discover by inductive reasoning, i.e. by discovering that a certain specific state of belief regularly leads to an act of recognition when confronted with a certain kind of fact; for example when we realize that our belief that the sky is blue always leads to an act of recognition when we see a blue sky.

This, according to Wittgenstein, misconstrues the relationship between mind and world. If I believe truly that London lies on the river Thames, then I believe this very fact and not some mental substitute for it. If what I believe is a surrogate of the actual fact and the relation of this surrogate to the actual fact is an external causal relation, I cannot judge whether my belief and the objective of the belief correspond to each other or not. It is no longer possible to compare my belief with the fact it is supposedly about:

> It is such *comparison* which is left out in Russell's theory. And comparison doesn't consist in confronting the representation with what it represents and through this confrontation experiencing a phenomenon, which, as I have said, itself could not be described in advance.
>
> (Experience decides whether a proposition is true or false, but not its sense.)

The sentence in parenthesis is particularly significant. According to Russell a belief is about whatever causes a feeling of recognition; the feeling of recognition, therefore, determines what I believe. This, according to Wittgenstein, gets things the wrong way round. I can only *re*cognize a fact as corresponding to my belief, if what I believe is determined independently of this fact.

4. Concluding remarks

Now we are in a position to appreciate Sellars's claim that explaining intentionality in terms of the idea that the mind is 'informed by the natures of external objects' and therefore broadly speaking 'isomorphic' (cf. BBK, 41-42) to them has significant advantages over other relational accounts of intentionality, which explain intentionality in terms of a relation between a mental state and a content rather than in terms of a relation between a mental state and the fact that makes it true. Unlike Russell, who thinks that a belief acquires its intentionality by standing in a relation to a content that is a 'present occurrence in the believer' (Russell 1921, 233) and is distinct from the 'objective' of the belief, i.e. the fact that makes it true, Wittgenstein's picture theory with its claim that the *pictorial relationship* between picture and pictured is internal to the picture does not run the risk of severing the link between intentional phenomena and the facts that make them true. Nor does Wittgenstein have to postulate 'subsistent' or 'negative' facts which correspond to false representations, as, according to Sellars, some 'extreme realists' (cf. BBK, 42) are forced to do. As we have seen, according to the picture theory there does not have to be any fact in the world that corresponds to a false picture.

However, the early Wittgenstein is not entirely successful in avoiding this latter problem which, according to Sellars, plagues many relational accounts of intentionality with a realist bent. For although Wittgenstein does not have to admit the existence of negative facts, like England being north of Scotland, or complex subsistent entities, like golden mountains or the fountain of youth, the picture theory still forces him to postulate the existence of sempiternal objects which correspond to the names occurring in elementary propositions and to assume a metaphysical structure of the extra-mental world which ineffably corresponds to the logical structure of our language. These two assumptions constitute a metaphysical leap of faith and, as I will show in the third chapter,

Wittgenstein will only be able to abandon them once he relinquishes the idea that intentionality is constituted by a relation between representings and the things represented, i.e. once he embraces what Sellars calls a 'non-relational' account of intentionality.

However, before I come to a discussion of the later Wittgenstein's rejection of a relational account of intentionality I want to discuss two assumptions which are intimately related with this account as developed in the *Tractatus*. These are, first, the idea that thinking takes place within a language of thought and is constituted by symbolic mental processes, and, secondly, the idea that there are mental acts which establish an intentional relation – in the language of the *Tractatus* a 'pictorial relationship' – between our linguistic representations and the states of affairs they represent. These two assumptions are subjected to thorough and devastating criticism in Wittgenstein's later work. This criticism will be the topic of the next chapter.

II. The Later Wittgenstein on Thought and Meaning

1. Introduction

In this chapter I shall discuss two mistakes the later Wittgenstein thinks one is prone to make when thinking about intentionality and which are at the heart of the account of intentionality given in the *Tractatus*. The first of these mistakes is the idea that thinking takes place within what might be called a 'language of thought' and consists in the operation of mental signs just as the use of public languages consists in the operation of acoustical and visual signs. The second mistake is the idea that the use of public languages consists of two parallel processes, the public use of linguistic signs and underlying mental acts or processes of meaning and understanding these signs. I shall call this understanding of language the 'dual-process conception of meaning'.[1]

In the previous chapter I have tried to show that Wittgenstein made both of these mistakes in the *Tractatus*. First, he embraced the idea that thinking takes place within a language of thought, which consists of mental signs resembling – presumably in their function – the signs of public languages. This conception of thinking is only hinted at in the *Tractatus* itself, when Wittgenstein asks rhetorically: 'Does not my study of sign-language correspond to the study of thought-processes…?' (TLP 4.1121), but is stated explicitly in a letter to Russell which Wittgenstein wrote shortly after the completion of his early masterpiece: '"Does a Gedanke consist of words?" No! But of psychical constituents that have the same sort of relation to reality as words' (NB, 130-31).

Second, he endorsed a form of the dual-process conception of meaning. For he argued that a propositional sign only comes to be a bearer of meaning if it is projected onto the states of affairs, which it is supposed to represent, by acts of '[thinking] the sense of the

[1] This expression is taken from P. M. S. Hacker (cf. 1990, 318-26).

proposition' (TLP 3.11). These acts of 'thinking the sense of the proposition' seem to play in the system of the Tractatus exactly the role acts or processes of meaning and understanding play in the dual-process conception of language, which is the focus of pointed attacks in Wittgenstein's later work.

After a detailed exposition of the later Wittgenstein's criticism of the idea of a language of thought and the dual-process conception, I shall close the chapter with a discussion of his positive conception of linguistic meaning and the correlated cognitive phenomena of understanding and meaning something by one's use of words.

2. The language of thought

The account of the psychological *hinterland* of the *Tractatus* I gave in the previous chapter is obviously rather schematic. This should not be surprising given that it was Wittgenstein's explicit purpose in that work to discuss the nature of *linguistic* representation rather than *mental* representation (cf. TLP, 3-4). The fact that he supposed thinking took place in a language of thought which is constituted by the use of mental signs and that he assumed that the meaning of 'propositional signs' depended on underlying mental acts of 'thinking the sense of the proposition' they express does not license the assumption that his account of thinking and its relation to language would have looked anything like the elaborate theories of contemporary cognitive scientists and philosophers, such as Jerry Fodor, who believe in a 'language of thought' and a 'representational theory of the mind'.[2] Nor is the account of thinking and its relation to language that stands behind the semantic doctrines of the *Tractatus* clearly identifiable as the central target of Wittgenstein's later criticism of Cartesian conceptions of the mind and its cognitive capacities.

[2] The *locus classicus* for Fodor's views is Fodor (1975), for a more recent version of his views see Fodor (2008).

One is tempted to assume that at the time he wrote the *Tractatus* Wittgenstein was wedded to an Augustinian picture of language, i.e. to a picture of language as it is expressed in the quote from Augustine opening the *Philosophical Investigations,*[3] not only in that he thought of the activity of naming something as being fundamental to all use of language, but also because he assumed that the acquisition of language presupposed a conceptually structured awareness of the world, i.e. the possession of some kind of language of thought:

> Augustine describes the learning of human language as if the child came into a foreign country and did not understand the language of the country; that is, as if he already had a language, only not this one. Or again, as if the child could already *think*, only not yet speak. And "think" would here mean something like "talk to itself". (PI §32)

The tractarian account of thinking and its relation to language seems very close to this beguiling picture of language acquisition. However, ascribing Augustinian beliefs about language learning to the early Wittgenstein is not licensed by what can be uncovered from the *Notebooks* and the *Tractatus* itself. We do not know what the author of the *Tractatus* thought about language acquisition and whether he thought it presupposed that the linguistic neophyte already possessed a language of thought. Nor do we know whether the young Wittgenstein was subject to other confusions concerning thinking and its relation to language, which are criticised in his later work. Did he believe that thinking was not only prior to but also independent of the ability to use language? Did he think that it was a process in the brain? Did he assume that it was essentially private and independent of public behaviour? It is not clear whether the early Wittgenstein would have answered any

[3] 'When grown-ups named some object, and at the same time turned towards it, I perceived this, and I grasped that the thing was signified by the sound they uttered, since they meant to point *it* out. This, however, I gathered from their gestures, the natural language of all peoples, the language that by means of facial expression and the play of eyes, of the movement of the limbs and the tone of voice, indicates the affections of the soul when it desires, or clings to, or rejects, or recoils from, something. In this way, little by little, I learnt to understand what things the words, which I heard uttered in their respective places in various sentences, signified. And once I got my tongue around these signs, I used them to express my wishes' (PI §1).

of these questions affirmatively had he considered them, albeit positive answers to at least some of them might appear congenial to his views.

Whatever he may have thought about these questions, two things seem to be reasonably clear by now. *First*, the early Wittgenstein believed that thinking is a kind of speaking, which consists in the operation of 'mental signs' that correspond to the signs of public languages; and, *second*, he believed in a form of the dual-process conception of language, i.e. he thought that the meaningful use of language, whether written or spoken, consists of essentially two parallel processes, operating with signs and mental acts of thinking the sense of these signs. It is on his criticism of the first of these two ideas that I shall concentrate in the present section of this chapter; the criticism of the second idea will be the topic of the next section.

2.1 Thoughts as the 'last interpretation'

In Wittgenstein's later work we can find two particularly profound objections to the thesis that thinking takes place within an inner language-like symbolism. The *first* of these objections is encapsulated in the following quote from *Philosophical Grammar* (PG):

> [The] point is that one has to read off from a thought that it is the thought that such and such is the case. If one can't read it off… then it is of no logical interest…
>
> If a thought is observed there can be no further question of an understanding; for if the thought is seen it must be recognized as a thought with a certain content; it doesn't need to be interpreted!—That really is how it is; when we are thinking, there isn't any interpretation going on. (PG, 143-44)

I take the thought hinted at in this remark to be the following: In the case of linguistic utterances or any other kind of symbolic representation we can always ask how the symbols that we employ are meant, in other words, how we ought to interpret or understand them. In the case of symbolic representation a gap can always open up between the symbol and what it means. In the case of thoughts, on the other hand, this gap cannot open up. That is why Wittgenstein says that a thought 'doesn't need to be interpreted' and

that one can ‘read off from a thought that it is the thought that such and such is the case’. Being aware of someone’s thinking a certain thought – whether it is oneself or someone else – is the same thing as being aware of the content of her thought. This is different in the case of linguistic utterances: I might be aware of someone’s uttering a certain sentence without being aware of the meaning of this sentence, i.e. without understanding it.

This difference between symbolic representation and thinking becomes clearer when we compare what one is doing when one reports someone’s utterance and when one reports someone’s act of thought. We typically report what someone said in two different ways; either by using indirect speech, as when we say ‘Paul said that the English are coming’, or verbatim by the use of quotation, as in ‘Paul said: ‘The English are coming!’’. We can see three interesting differences between these two kinds of report, which may not always be heeded in ordinary discourse but are nevertheless implicit in it. First, if these reports express knowledge, it is knowledge of a different kind. The first kind of report expresses knowledge of the content of what someone said, whereas the second expresses knowledge of the actual words used. We often have both types of knowledge of what someone said but they are nevertheless independent of each other; I can have the former without the latter, for example when I remember what someone said without remembering the exact words; and I can have the latter without the former, for example when I exactly remember the words of someone’s utterance but do not understand the utterance. Second, the latter but not the former kind of report is false if it does not repeat the exact words of the reported utterance – at least when the report is couched in the same language as the reported utterance. If I said ‘Jim said: ‘It is going to rain tonight’’ and Jim really said ‘It will rain tonight’ I have made, strictly speaking, a false statement; but if I reported his utterance with the words ‘Jim said that it is going to rain tonight’ my report is true. Third, if I report someone’s nonsensical utterance, such as ‘Green ideas sleep furiously’, by the

use of quotation, my utterance, contrary to the one I quote, makes sense. The statement 'Jack said: 'Green ideas sleep furiously' is perfectly alright because it only mentions a nonsensical string of words but does not use them. However if I report the same utterance in *oratio obliqua* by saying 'Jack said that green ideas sleep furiously' I myself produce a nonsensical utterance, since I am myself *using* the nonsensical string of words contained in the original utterance.

I take it that the reason for these differences is that the first kind of report tells us, roughly, that someone said something with a particular content; whereas the second tells us that someone said something which consisted of particular words. Accordingly, only the former kind of report can express knowledge of meaning, and only the latter kind of report can express knowledge of the exact words uttered.[4] And since only the latter reports the exact words that were uttered, only the latter can be false if it does not contain those exact words. Finally, since only the former kind of report tries to give the content or meaning of the reported utterance, it alone will fail to report anything if that utterance made no sense, if – in other words – it had no meaning.

In principle we can make these two kinds of reports for every utterance. Furthermore, we can even make these two types of reports for 'utterances' performed in 'inner speech'. We can report what words or sentences we or someone else imagined, and we can report how these words or sentences were meant or understood. And since we can make these two different types of report for both public utterances and utterances performed in 'inner speech' we can, in principle, have two different kinds of knowledge of both of these types of utterance, knowledge of their meaning and knowledge of the words uttered. One might

[4] It should be noted that when I say that only reports of the form 'X said that p' can express knowledge of the meaning of what is reported, whereas reports of the form 'X said: 'p'' merely express knowledge of the words constituting the reported utterance, I am using the phrase 'expression of knowledge' in a very narrow sense. I use it to mean, roughly, 'expression of the proposition that is said to be known'. 'Expression of knowledge' in this sense is to be distinguished from 'manifestation of knowledge'. For by saying: 'X said: 'p'', I can clearly manifest my knowledge of what it is that X said, i.e., unless there is strong evidence to the contrary, I thereby show that I have such knowledge. However, I am not 'expressing' that knowledge in the sense in which I have used this term above.

think that one could not have knowledge of the words oneself imagined without having knowledge of their meaning. This, however, is just as wrong as thinking that whenever one utters a sentence in a public language one knows what one is saying.

In the case of acts of thought these two types of knowledge – knowledge of the content of the thought and knowledge of its vehicle – are not available, not even in one's own case. We can, of course, know that someone is thinking something without knowing what she is thinking. But, first, we can only have this kind of knowledge of another person's thoughts. When we know that we ourselves are thinking something, we also know what we are thinking. And, second, we cannot know the *specific* thought someone is thinking without knowing the content of that thought, whereas we *can* know what specific sentence someone is uttering without knowing the meaning of that sentence. Since knowing that someone is thinking a specific thought implies that one knows what she is thinking, whereas knowing what sentence someone is uttering or imagining does not entail knowing its meaning, not even in one's own case, Wittgenstein concludes that 'if a thought is observed there can be no further question of an understanding' (ibid.) and that thoughts are the 'last interpretation' (cf. BB, 34).

This insight casts doubt on the tractarian assumption that thoughts are symbolic processes in the mind similar to utterances in a public language. Symbolic processes in the mind, such as the utterance of sentences in 'inner speech' can be interpreted or understood. Unlike thought they do not constitute a 'last interpretation'. Contrary to what Wittgenstein seems to have assumed when writing the *Tractatus,* our concept of an act of thought is not the concept of a mental event consisting of 'psychical constituents that have the same sort of relation to reality as words' (NB, 131).

2.2 The location of thought

The second argument against the idea that thinking is constituted by the occurrence of inner signs, I want to discuss, is contained in the following remark from the *Investigations*:

> The chair is thinking to itself…
>
> Where? In one of its parts? Or outside its body; in the air around it? Or not *anywhere* at all? But then what is the difference between this chair's talking silently to itself and another one's doing so, next to it? – But then how is it with man: where does *he* talk to himself? How come that this question seems senseless; and that no specification of a place is necessary, except just that this man is talking to himself silently? Whereas the question of *where* the chair talks to itself seems to demand an answer. – The reason is: we want to know *how* the chair is supposed to be like a human being; whether, for instance, its head is at the top of the back, and so on. (PI §361)

What Wittgenstein is saying here is that if we want to ascribe acts of thought or inner speech to a chair (or any other inanimate object),[5] even if we want to do so only metaphorically, we have to specify in what way the chair (or other inanimate object) is supposed to resemble a human being. This is why we might ask for the location of the thoughts that are ascribed to a chair. We do *not* do so because we know the location of acts of thought within the human body, e.g. the brain, and now want to know where to find the corresponding location within the chair. Asking in which part of her body some human being thought that p does not make sense. Our thoughts are not located in any part of our body. When we ask where someone had the thought that p, we are not asking in which part of her body she thought that p, but where she was when she thought that p, for example in the car, under the shower or on her way to work. Furthermore, if we are presented with a chair that is sufficiently similar to a human being to warrant the ascription of mental acts to it, for example in a fairytale (cf. PI §281) or Disney movie, we do not ask for the location of these thoughts either.

Wittgenstein gives his reason for why it does not make sense to ask in which part of her body a person thinks, in the immediately preceding paragraph:

[5] Cf. e.g. the discussion of the ascription of perception to pots, PI §282, and of sensations to stones PI §§284, 288.

> We say only of a human being and what is like one that it thinks. (PI §360)

This echoes a remark Wittgenstein makes earlier in his discussion of a private language for our sensations (the so-called 'private language argument'):

> "But doesn't what you say amount to this: that there is no pain, for example, without *pain-behaviour*?" – It amounts to this: that only of a human being and what resembles (behaves like) a living human being can one say: it has sensations; it sees; is blind; hears; is deaf; is conscious or unconscious. (PI §281)

Acts of thought and other intentional mental episodes and states, such as perceptions, beliefs, intentions or desires, as well as non-intentional mental episodes and states, such as feelings of pain or pleasure, are only attributable to the person as a whole and not to any part of her. It is the person as a whole who thinks that it is time for dinner, desires a steak and intends to take a second helping and not her brain, her stomach or her hands. Similarly, it is the person as a whole who sees the mountains on the horizon and feels the pain in her left foot and not her eyes or her foot.

However, if acts of thought consisted in the operation of mental signs, as Wittgenstein thought when he wrote the *Tractatus*, asking for the bodily location of an act of thought should make sense. After all, whenever we produce some sign, be it in speech, in writing or in sign-language, we do so with a specific part of our body, i.e. with the mouth and vocal chords or with our hands. If we, for example, claim that Jack uttered 'p', it makes sense to ask whether he did so with his hand in writing or his mouth in speech. On the other hand, asking whether someone thought with his heart or with his brain makes no sense, except in metaphorical contexts. For neither does a person's brain have thoughts nor does a person think with his brain in the sense in which he writes with his hand and speaks with his mouth. For we cannot use our brain in the same way in which we can use our hand or our mouth, i.e. we cannot exercise intentional control over its activities. Accordingly, we cannot say that acts of thought consist in the operation of mental signs as utterances

consist in the operation of acoustical or visual signs. And this implies that thinking is not a matter of operating mental signs.

There are two complementary objections to this argument. First, one might claim that there are certain states that are only attributable to a person as whole and not to any part of her which nevertheless can be said to have a specific location within a person's body: One can say only of a whole person that she feels an itch in her right arm or a pain in her left toe, but this does not mean that the itch is not in her arm or the pain not in her toe. Second, one might claim that there are certain acts, namely speech acts, which can only be attributed to the whole person but which are nevertheless realized by the production of specific linguistic signs.

It is certainly true that we can only say of a whole person that she *feels* or *has* a pain or an itch and that we cannot say it of any part of her body. And although it has been denied by a considerable number of philosophers[6] including – perhaps – Wittgenstein himself,[7] it seems at least *prima facie* true that a pain or an itch is generally located in the part of the body where we feel it. However, contrary to thoughts, itches and pains *can* be attributed to parts of the human body, for example in such sentences as 'My toe is hurting' or 'My arm is itching'. It is only the states of *having* or *feeling* a pain or an itch that are only attributable to the person as a whole. Hence it makes sense to ask where someone's pain is located, whereas it does not make sense to ask about the location of her thought.

The second objection does not cut much ice either. First, the idea that thinking takes place within an inner symbolism is the idea that thinking is *identical* with the occurrence of inner symbols, just as an utterance is identical with the occurrence of linguistic signs. However, a speech act, such as an assertion, cannot be identified with the utterance that is its vehicle. Whatever one's specific account of assertion is, one will presumably agree that

[6] For example R. Descartes, W. James, B. Russell, C. D. Broad, G. Ryle, D. Lewis and D. Armstrong. For this list and references cf. Hyman (2003, 3-10).
[7] This is argued in Hyman (2003, 12-15).

making an assertion is an intentional action. Accordingly, when we attribute a specific assertion to someone we are not only claiming that she produced linguistic signs with a specific meaning, but we are also saying that she did so intentionally. Hence we are attributing to her a mental state, which, just as an act of thought, can only be attributed to the whole person.

Second, attributing an act of thought unlike attributing a speech act does not even imply that certain signs were produced. Had it been discovered upon his death that Immanuel Kant's body was filled with sawdust (cf. OC §281), this would not have been a reason to withdraw one's attribution of his philosophical thoughts to him. On the other hand, if we were to discover that he did not write the three *Critiques* this would be a reason to do so. Thus, unlike attributing a speech act, attributing a specific thought to a person does not imply anything about the presence of some kind of vehicle within her body, let alone about the specific character of that vehicle; e.g. that it consists of language-like symbols, which are combined in the same way the symbols of a public language are, as the claim that thinking takes place within a language of thought suggests.

Wittgenstein goes even so far as to suggest that thinking might take place without a vehicle. In *Zettel* §608 he claims:

> 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 the *system* continue further in the direction of the centre? Why should this order not proceed, so to speak, out of chaos?

And in the next section he concludes:

> It is thus perfectly possible that certain psychological phenomena *cannot* be investigated physiologically, because physiologically nothing corresponds to them. (Z §609)

Wittgenstein claims that this idea, i.e. the idea that there are no physiological processes which correspond to the mental process or activity of thinking, does not imply mind-body

dualism, i.e. it does not imply that there are processes corresponding to thought in some non-physical mind-substance. He appears to think that assuming a mind-substance within which our thoughts are realized is the result of the same kind of confused and misguided thinking as assuming that there have to be intracranial processes which serve as the vehicle of our acts of thought:

> The prejudice in favour of psychophysical parallelism is a fruit of primitive interpretations of our concepts. For if one allows a causality between psychological phenomena which is not mediated physiologically, one thinks one is making profession that there exists a soul *side by side* with the body, a ghostly soul-nature. (Z §611)

I believe Wittgenstein is right in claiming that it is possible that there are no physiological processes corresponding to our mental activities. Furthermore it is illuminating to point out that admitting this possibility is not the same thing as admitting the possibility of mind-body dualism. Acts of thought could indeed occur without any vehicle, physiological or otherwise. However, the intimated suggestion that that is what actually happens, i.e. that when we think there are actually no physiological processes occurring in our brain, which correspond to our thoughts, I believe to be misguided. As Anthony Kenny points out in discussing the above quoted passages, much progress has been made within the history of science by inquiring into the vehicles of capacities.[8] Wittgenstein's suggestion that our power of thought and its exercise could have no vehicle simply seems like a dogmatic rejection of the possibility of a scientific study of the neurological mechanisms underpinning our mental lives. The problematic character of Wittgenstein's conclusion is aptly illustrated by the analogy with which he illuminates it. After suggesting 'no supposition seems… more natural than that there is no process in the brain correlated with associating or with thinking' he claims:

[8] cf. Kenny (2009, 262).

> 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 — this can only be done from the *history* of the seed. So an organism might come into being even out of something quite amorphous, as it were causelessly… (Z §608)

Nowadays we know that there is something in the seed that 'corresponds to the plant which comes from it', namely its DNA. Similarly we ought to expect that there is something in our brains that corresponds to our mental goings-on. Wittgenstein is right in claiming that this is not entailed by our concepts of thought or belief, but it certainly appears to be a rather sound empirical conjecture.

3. The dual-process conception

The second fundamental flaw the later Wittgenstein finds in the kind of picture of intentionality developed in the *Tractatus* is the idea that the signs of our language are given their sense not merely by their role within the system of symbols to which they belong, i.e. by their *logical form*, but also have to be projected onto the reality they represent by underlying acts of meaning. The Tractarian idea that the intentionality of propositional signs is partly determined by mental acts of 'thinking the sense of the proposition', which establish a *pictorial relation* between the linguistic picture and what it depicts, is an application to the specifics of the picture theory of a highly intuitive and philosophically tempting conception of the relationship between our mental life and its expression in language. Wittgenstein gives succinct expression to this idea in a passage from the *Blue Book* already mentioned in the previous chapter:

> It seems that there are *certain definite* mental processes bound up with the working of language, processes through which alone language can function. I mean the processes of understanding and meaning. The signs of our language seem dead without these processes; and it might seem that the only function of the signs is to induce such processes… Thus, if you are asked what is the relation between a name and the thing it names, you will be inclined to answer that the relation is a psychological one, and perhaps when you say this you think in particular of the mechanism of association. -- We are tempted to think that the action of language consists of two parts; an inorganic part, the handling of signs, and an organic part, which we may call understanding these signs, meaning them, thinking. (BB, 3)

I take it that the purpose of this passage is not only to give expression to a particular view about the relationship between names and their bearers, i.e. the Tractarian view that this relation is established by a psychological act of association or projection, but is primarily intended to articulate a very general view about what makes linguistic signs meaningful. According to the 'dual-process conception', as it is characterized in this passage, the difference between the meaningful sounds and marks we produce in speech and writing and meaningless sounds and marks, such as the sounds we emit when humming a tune and the marks we leave in the sand when walking across a beach, is that the former but not the latter are accompanied by or issue from certain semantically significant mental events, states or processes. It is in particular the mental acts, states or processes of meaning and understanding that, according to this conception, turn the signs of language into bearers of meaning. Furthermore, the dual-process conception does not only provide a natural explanation of what makes signs meaningful but also provides very intuitive answers to such questions as 'What is difference between meaning Napoleon I and meaning Napoleon III by one's use of 'Napoleon'?' or 'Why can only a person but not a parrot *claim* or *assert that* Napoleon is impetuous by her use of the words 'Napoleon is impetuous'?' According to the dual process conception it is certain specifically semantic mental phenomena that make our, but not the parrot's, utterance of 'Napoleon is impetuous' a claim or an assertion and determine whether we refer to Napoleon I or Napoleon III.

The processes, acts or states of meaning and understanding, as they are understood by the dual-process conception, can be conceived as merely running alongside overt speech, but are, I think, most naturally understood as its cause or effect. When we conceive of meaning and understanding in this latter way, what distinguishes meaningful speech from the meaningless production of noise is not merely that it involves the production of sounds that are part of some language and are arranged in a manner that counts as grammatical in

that language, but that this production is *caused by* mental acts, states or processes of meaning on the part of the speaker and *causes* mental acts, states or processes of understanding on the part of those who follow with comprehension what is being said.

In this respect the dual process conception of meaningful speech is analogous to a certain form of the causal theory of action. In PI §621 Wittgenstein famously asks: '[What] is left over if I subtract the fact that my arm rises from the fact that I raise my arm?' According to the causal theory of action what remains is some mental event, which is the cause of the event of my arm rising. Donald Davidson, for example, claims in his famous paper 'Actions, Reasons, and Causes'[9] that all intentional actions are caused by the 'onslaught'[10] of what he calls a 'primary reason'; where a 'primary reason' consists of a desire (or in Davidson's terminology a 'pro-attitude') to perform an action of a specific type and a belief that the action one performs is of that type.[11] If we now ask 'What is left over if I subtract the fact that I uttered 'p' from the fact that I said that p?' where 'saying that p' is a matter of uttering 'p' with meaning, the dual process conception provides an answer exactly parallel to the answer given by the causal theory of action to the corresponding question about raising one's arm. What remains, according to the dual-process conception, is an inner act, state or process of a specific semantic character, namely an act, state or process of meaning. And if there is an audience who understands what I am saying she will be subject to corresponding mental states, events or processes of understanding.

The dual-process conception as presented so far – and as Wittgenstein appears to have endorsed it in the *Tractatus* – not only provides an intuitive picture of the meaningful use of language, it also has a clear explanatory purport. It aims to explain what is distinctive about the meaningful use of language, as contrasted with, say, the mere parroting use of

[9] Reprinted in Davidson (2001, 3-19).
[10] Cf. ibid, 12.
[11] Cf. ibid, 5.

words or the production of sounds or marks, which have no meaning, just as the causal theory of action tries to explain what is distinctive about those bodily movements that can be attributed to a person's agency. It ultimately seeks not merely to describe the meaningful use of language but to *explain* how we can endow 'mere dead signs' with intentional content, i.e. how we can use signs to say something.

In his later work Wittgenstein thoroughly rejects the dual-process conception of language as well as the analogous idea that intentional action is distinguished from non-intentional behaviour by the presence of underlying mental processes. As he puts it: 'The intention *with which* one acts does not 'accompany' the action anymore than a thought 'accompanies' speech' (PPF §280, italics in the original).

His objections to the dual-process conception can be roughly divided into three different kinds. *First*, it does not fit with what might be called the *phenomenology* of language use. When we look at what happens when we are speaking meaningfully or when we understand what someone else is saying, we simply cannot find the kinds of events and processes that according to the dual-process conception are necessary for meaningful discourse. *Second*, it does not fulfil its explanatory promise. Whatever kind of mental act, event or process we imagine to accompany our utterances, it does not seem to be any better at settling the meaning of what was said than the original utterances themselves. *Third*, the categorization of meaning and understanding as mental events, states or processes is mistaken. What we mean by what we say is determined by how we *use* the symbols we are employing, not by anything accompanying this use. Understanding what someone is saying or what a word means, on the other hand, is not some kind of mental state or occurrence accompanying the use of signs either. It rather consists in a variegated array of abilities, such as the ability to use the word in question correctly or to react to what someone else said in a manner that is appropriate in the light of its meaning.

3.1 Phenomenology

If words acquired their meaning through underlying mental acts or processes, it ought to be possible to mean whatever we wanted to mean by specific words. Hence, we should be able to mean that p while saying 'q'. There is undeniably a sense in which we can do this; namely by what Paul Grice calls 'conversational implicature'.[12] We can, for example, say 'What an interesting choice of wardrobe' and mean 'What an ugly dress'. But here we are not endowing the former sentence with the meaning ordinarily associated with the latter through an underlying mental act or process, we are rather trying to convey a belief of ours by a sentence that itself does not express this belief but whose use on the occasion suggests to our audience that we hold the belief in question.

However, we cannot endow a sentence or string of signs *ad hoc* with a specific meaning through a mental act of ours. We cannot say '"bububu" and mean "If it doesn't rain, I shall go for a walk"' (boxed remark after PI §35) nor can we mean 'toothache' by saying 'abracadabra' (cf. PI §665). And this is not because our audience will not understand us. For we can mean something by an expression our audience does not understand. We may, for example, use the expression 'comme il faut' in a situation in which it is not quite *comme il faut* to use it; say, when speaking to small children. Still we will not fail to mean something by it on that occasion.

We may, of course, succeed in meaning 'toothache' by 'abracadabra' by saying, 'By 'abracadabra' I mean toothache' (cf. ibid). But this sentence is a *definition* of 'abracadabra' (cf. ibid), which indeed allows us to subsequently mean toothache by our use of 'abracadabra', it is not, as Wittgenstein points out, 'a description of what goes on in me when I utter the word' (ibid.).

[12] Cf. Grice (1989, 24-40).

Contrary to what one might think when in the grip of the dual-process conception, understanding, Wittgenstein suggests, is not a mental process either. He illustrates this by the use of an example:

> A writes down a series of numbers; B watches him and tries to find a rule for the number series. If he succeeds, he exclaims: "Now I can go on!" – Now this ability, this understanding, is something that occurs in a moment. So let us have a look: what is it that occurs here? – A has written down the numbers 1, 5, 11, 19, 29; at this point B says he knows how to go on. What happened here? Various things may have happened; for example, while A was slowly writing down one number after another, B was busy trying out various algebraic formulae on the numbers which had been written down. After A had written down number 19, B tried the formula $a_n = n^2 + n - 1$; and the next number confirmed his supposition. (PI §151)
>
> Or again, B does not think of formulae. He watches… how A writes his numbers down, while all sorts of vague thoughts float through his head. Finally he asks himself, "What is the series of differences?" He finds: 4, 6, 8, 10, and says: "Now I can go on."
>
> Or he watches and says, "Yes, I know *that* series" – and continues it, just as he would have done if A had written down the series 1, 3, 5, 7, 9. – Or he says nothing at all and simply continues the series. Perhaps he had what may be called the feeling "That's easy!"

Wittgenstein is right to point out that understanding can occur 'in a moment' or, as he puts it on a different occasion, we can grasp something, such as the meaning of a word or the principle behind a series of numbers, 'at a stroke' (cf. PI §§139 and 191). This may make one suspect that there has to be some specific conscious occurrence that constitutes our understanding. However, the above examples show that it is difficult to find a common occurrence in consciousness in all cases in which we would like to say that someone understood something. In all of the four scenarios Wittgenstein considers B understands what series A is writing down. And in all of them there is a specific moment when understanding sets in. Yet, there does not seem to be one common conscious occurrence among all these cases, which could be said to be the understanding of the series. Looking at the first two scenarios, i.e. the scenarios where the formula '$a_n = n^2 + n - 1$' and the series of differences '4, 6, 8, 10' occur to B, one might conclude that understanding the series 1, 5, 11, 19 etc. does not require that one *specific* formula or rule occur to one but that *some* rule or formula, which correctly determines the steps of the series, occur to one. But this demand is already too strong, as the third and fourth scenarios show; for the

thoughts ‘that’s easy’ or ‘I know that series’ are clearly not thoughts about a principle or rule determining the steps of the series. Still these could have been the only thoughts that crossed B’s mind at the moment he understood what series A was writing down.

Furthermore, it seems all of the things Wittgenstein imagines to take place in B’s mind at the moment of understanding what series A is writing down could occur without B understanding the principle behind the series (cf. PI §152). The formula ‘$a_n = n^2 + n - 1$’, the series of differences ‘4, 6, 8, 10’, the thought ‘Yes, I know *that* series’ or the thought ‘That’s easy!’ could occur to B without him being able to continue the series correctly. We can, for example, imagine that the formula ‘$a_n = n^2 + n - 1$’ occurs to him, but that he dismisses or misapplies it (cf. PI §185). Accordingly Wittgenstein concludes: ‘“He understands” must have more to it than: the formula occurs to him’ (PI §152). For saying that B understands what series A is writing down implies that B is able to continue the series correctly.

It would be a mistake, however, to assume that since understanding is not a conscious mental process it is an unconscious one; an event or process that is, as Wittgenstein puts it, somehow ‘hidden behind those coarser, and therefore more readily visible, concomitant phenomena’ that are the conscious accompaniments of understanding (PI §153). For, first, if the thought of the formula ‘$a_n = n^2 + n - 1$’ cannot constitute the understanding of the principle of series A is writing down then how can any unconscious occurrence do so? If it is compatible with the formula occurring to B that B does not understand what series is being written down, then, it seems, it should be equally compatible with any other occurrence in B’s mind, regardless of whether it is conscious or unconscious. Second, there does not appear to be anything unconscious or ‘hidden’ about understanding. We are normally perfectly aware of having understood something. We say such things as ‘Now I understand’ or ‘Now I’ve got it’ because we understand something (cf. PI §153) and it

does not seem plausible that these statements report a speculative conclusion about some hidden phenomenon, which causally explains the 'coarser' and more 'readily visible' manifestations of understanding such as a formula occurring to one.

3.2 Explanatory futility

The second type of objections against the dual-process conception I want to consider try to show that the idea that meaningful speech consists in an overt activity of handling signs and an accompanying mental process is not only difficult to reconcile with the phenomenology of language use but also fails to serve its explanatory purpose, i.e. it does not tell us what is distinctive about the meaningful use of linguistic signs, as contrasted with the mere production of meaningless marks and noises. In other words, the dual-process conception embodies a philosophical theory with no explanatory value.

I will consider two of Wittgenstein's arguments that aim to show the explanatory futility of the dual-process conception of meaning and understanding. The *first* argument can be found at the beginning of the *Blue Book*. There Wittgenstein asks us to imagine replacing the inner process that is supposed to give our utterances meaning by an outer activity (cf. BB, 3ff, 33ff). If we think that, for example, the use of colour words has to be accompanied by mental images in order to be more than a mere production of noise, we should imagine that someone carries a colour chart with her, in which colour-samples are correlated with their names and which she consults whenever she speaks about colours (cf. BB, 3). If we are inclined to believe that the mere production of linguistic signs alone cannot determine their meaning, then we will probably not think that producing sounds plus handling a colour chart will do the trick. For colour charts are just another means of representing colours and can be variously applied, just as linguistic signs can. For example, it seems perfectly possible for someone to understand a colour chart, in which the names of colours are written next to their samples, in a way that she associates a colour sample not

with the name next to it, but with the name next to the sample directly underneath it (cf. PI §§86 and 163).

As soon as we replace the inner process or activity that is supposed to constitute our meaning and understanding by some outer activity, such as looking at colour samples, the idea that some accompanying process or activity should give our utterances their meaning becomes less appealing. If one thinks that one has to assume mental acts or processes in order to explain how 'mere dead signs' can come to carry intentional content one has fallen prey to a picture of the mind as some 'queer kind of medium' (BB, 3), which 'can bring about effects which no material mechanism could' (ibid.). The assumption of such a 'queer medium', such as the Cartesian soul, is a myth; a myth Wittgenstein tries to disabuse us of when he asks us to think of the assumed inner processes of meaning and understanding as being replaced by the outer activity of handling colour charts:

> [As] soon as you think of replacing the mental image by, say, a painted one, and as soon as the image thereby looses its occult character, it ceases to seem to impart any life to the sentences at all. (It was in fact just the occult character of the mental process which you needed for your purposes.) (BB, 5)

The *second* objection against the explanatory value of the dual-process conception I wish to discuss is the so-called 'rule-following paradox'. Ever since Saul Kripke's seminal *Wittgenstein on Rules and Private Language* (1982) this paradox and its discussion, in particular in *Philosophical Investigations,* §§138 to 242, and *Remarks on the Foundations of Mathematics* has been at the centre of much discussion, both among Wittgenstein scholars and among philosophers of language more generally. My main focus here will be on PI §§185 to 242, which I take to constitute the core of the rule-following discussion in PI and which I believe primarily aim to show (a) that the dual-process conception of meaning and understanding is mistaken, (b) that one can only mean something by one's use of linguistic signs against the background of an already existing linguistic practice, and (c) that one can only ascribe meaning to and understand the utterances of persons who

share our 'form of life' (cf. PI §241); where a shared form of life consists in shared habits, dispositions and more generally a shared way of doing things (cf. PI §241), as well as agreement about the definitions of the expressions one is employing and – more controversially – agreement in at least a substantial subset of one's beliefs and judgments (cf. §242). Many interpreters take it also to be Wittgenstein's aim to show that (d) language is essentially social, i.e. that there could only be a language where there is a community of speakers, and (e) to advance a sceptical challenge to the idea that there is such a thing as linguistic meaning at all.[13] Claim (d) and in particular claim (e), however, are quite controversial.[14] Here I will, naturally, focus on the first of these five points, i.e. the claim (a) that the dual-process conception is mistaken. Later I will also have occasion to discuss claims (b) and (e).

In order to appreciate how the dual-process conception is affected by the rule-following paradox it is best to attend briefly to what may be called the 'normative character' of linguistic meaning. The meaning with which we use a word seems to have a certain *normative* force. It tells us, among other things, how the word is *correctly* used, what we *commit* ourselves to by employing it within an assertion, and what it is *correct* or *incorrect* to infer from its use on a particular occasion. The meaning of the English expression 'dog', for example, appears to determine that we will only use it correctly if we apply it to dogs and not if we apply it to ducks; that when someone says 'The dog ate the sausages' she *ceteris paribus* commits herself to an event of a dog eating sausages having occurred; that if something may be called 'dog' then it may also be called 'animal'; and that when someone announces 'I will now walk the dog' she *ceteris paribus* commits herself to walk the dog in the near future. Thus, meaning something by one's use of a

[13] Claim (d) is advanced, for example in Kripke (1982), Wright (1980), McDowell (1984), Fogelin (1976) and Malcolm (1989). Kripke and Fogelin additionally endorse (e).
[14] Gordon Baker and Peter Hacker dispute (d) and (e) in Baker & Hacker (2009) and in Baker & Hacker (1984); (e) but not (d) is also prominently disputed by John McDowell (1984) and Norman Malcolm (1986).

word, means using it so as to determine all these consequences. Understanding what a word means on an occasion of its use, on the other hand, requires grasping these consequences, i.e. grasping how the word in question is correctly used, what the speaker commits herself to by using it, and what can and what cannot be inferred from its use on a particular occasion.

Puzzled by how the mere production of signs can make all these *normative determinations* about such things as use, inference and commitment one might conclude, in the spirit of the dual-process conception, that they are due to certain accompanying mental items, namely processes of meaning and understanding. It might seem that if we say something and mean it, all the consequences of what we are saying are thereby determined in a way in which only an act of meaning or a corresponding act of understanding could determine them. According to Wittgenstein, when we give an order and mean it, it might seem as though the act of 'meaning the order… in its own way already [takes] all those steps' (PI §188) that count as its correct execution. Meaning and understanding, according to this view, are mental items which are somehow capable of determining such things as how the words making up an utterance are correctly used, and what can and cannot be inferred from their use on a specific occasion. Accordingly, only if an act of speech is accompanied by some such mental act or process of meaning or understanding can it really be said to mean anything.

However, how processes of meaning and understanding can determine the meaning and thereby the normative implications of what is being said is rather unclear. In PI §185 Wittgenstein illustrates this problem with a well-known example. He sketches a scenario of someone who is taught to develop a variety of different series of integers and is given exercises and tests up to 1000. He is then told by his teacher to develop the series of even integers. He proceeds as he is supposed to, until he reaches 1000 after which he continues

with ‘1004, 1008, 1012, etc’. When the pupil is asked to continue to add two, he claims to be going on in the same way. What is it about the teacher’s order that determines that the pupil’s continuation of the series after 1000 does not conform to it? What justifies the teacher in claiming that the pupil does not obey his order? According to the dual-process conception it is the act or process of meaning with which he accompanied his order. But how can such an act of meaning determine all the infinitely many steps that count as executing the order ‘add 2!’? After all, the teacher could not possibly have explicitly thought of all these steps:

> So when you gave the order “+2”, you meant that he was to write 1002 after 1000 – and did you then also mean that he should write 1868 after 1866, and 100036 after 100034, and so on – an infinite number of such sentences? (PI §186)

Obviously the teacher could not have had all these steps explicitly in mind when he gave the order. On the other hand it seems equally obvious that he indeed meant his pupil to write ‘1002’ after ‘1000’ and ‘1868’ after ‘1866’ (cf. PI §§187, 693). As Wittgenstein points out, one would like to say that although the teacher could not possibly have thought of all the instances of the series he told his pupil to produce, he meant his order in a specific way and from this all the individual applications of the order follow (cf. PI §186). However, how can an act or process of meaning or understanding achieve this remarkable feat, which, according to the dual-process conception, cannot be achieved by the mere production of ‘dead’ signs? How can something that is in our minds at the moment we mean or understand a sign, determine its correct use, which is, after all, ‘extended in time’ (cf. PI §138)?

In the *Blue Book* and in the *Investigations* Wittgenstein discusses three types of answers a philosopher committed to the dual-process conception might give. *First,* she might claim that meaning and understanding are just very special mental phenomena, which in a ‘strange way’ (cf. PI §195) already contain the correct use of the words we are

employing, as well as all the inferential and practical consequences of their use. Wittgenstein compares this conception of meaning and understanding to the idea of an 'ideally rigid machine', which 'already contains its possible movements in some mysterious way' and cannot possibly break down (cf. PI §194). Since no biological goings-on, such as processes in our brain, could possibly have these remarkable features this approach will typically lead to mind-body dualism; namely to the Cartesian idea that mental phenomena have to occur in some 'queer kind of medium, the mind' (BB, 3), which is distinct from our body, since it has powers no biological medium has.

The problem with this approach to the question of how linguistic signs acquire their meaning is not merely its commitment to an ontologically dubious entity, i.e. a Cartesian soul, but also the fact that the *explanans* is more obscure than the *explanandum*. It replaces a puzzle, i.e. the question of how signs come to have meaning, by a mystery, i.e. the claim that they are given their meaning by processes that take place in some strange immaterial medium.

I take it that this approach exemplifies exactly the kind of metaphysical theorizing Wittgenstein repeatedly warns against in his later work. First, the role in our language of such perfectly ordinary expressions as 'meaning' and 'understanding' is misconstrued. In a somewhat scientistic spirit they are taken to refer to some 'inner' occurrences that are taken to explain the 'outer' phenomenon of meaningful speech; just as in the natural sciences the observable behaviour of macroscopic objects is sometimes explained by the unobservable behaviour of theoretical entities that is taken to be causally responsible for it; for example, when it is said that heat is molecular motion. Then it is recognized that the philosophical explanation of linguistic meaning in terms of underlying mental items requires that we ascribe to these items more exceptional qualities than the familiar causal powers commonly attributed to theoretical entities in the sciences. Finally, this problem is

accommodated by introducing an obscure entity such as the Cartesian soul, which belongs to an entirely new ontological realm and possesses powers which transcend everything encountered in the natural world. It is this kind of philosophical theorizing that Wittgenstein castigates when he says that '[when] we do philosophy, we are like savages, primitive people, who hear the way in which civilized people talk, put a false interpretation on it, and then draw the oddest conclusion from this' (PI §194).[15]

Second, a philosopher committed to the dual process conception might claim that we give the signs of language meaning by *interpreting* them. This suggestion, however, leads into a well-known regress. It seems that whatever interpretation we give a particular sign or set of signs it will not be better at determining the meaning of the sign or set of signs than the original sign or signs themselves. If the teacher and the pupil of PI §185 manage to agree on an interpretation for the order 'Add 2!', say 'Develop the series of even integers!' or 'Write down the series that is determined by the formula n = 2n!', they could still disagree about what these interpretations tell the pupil to do. The teacher might think that the pupil has to write down '1398' after '1396' while the pupil might believe that he ought to write down '1400'. Adding further interpretations will not help, for teacher and pupil might again agree on the interpretation while continuing to disagree about what it entails for the actions of the pupil, and so on *ad infinitum*. If the meaning of an order or a rule is determined by an interpretation, no order or rule can ever determine what one has to do in order to follow or obey it. This, as Wittgenstein points out, strikes us as paradoxical:

> This was our paradox: no course of action could be determined by a rule, because every course of action can be brought into accord with the rule. The answer was: if every course of action can be brought into accord with the rule, then it can also be brought into conflict with it. And so there would be neither accord nor conflict here. (PI §201)

[15] For a similar reading of this passage see Stroud (1996, 296ff).

It seems that for an order such as ‘add 2!’ or a rule such as ‘If a skier misses a gate her run will be disqualified’ we will always be able to find many conflicting interpretations. Someone might for example interpret ‘Add 2!’, as we would understand ‘Add 2 up to 1000, 4 up to 2000, 6 up to 3000, and so on!’ (PI §185) or she might interpret ‘gate’ as we would interpret ‘red gate’. However, if it can never be settled what a statement expressing an order or a rule tells us to do, then this statement fails to express an order or rule. Statements with a prescriptive force such as orders and rules divide the space of possible actions into those that accord with them and those that do not; if they fail to do so they loose their prescriptive force and thereby their meaning. And since every statement has certain prescriptive or normative implications, e.g. implications about what may and what may not be inferred from it, the problem encountered in the case of rules and orders is a perfectly general one pertaining to any meaningful use of language. Just as someone could always think of an apparently correct interpretation of a rule but still act in a way that shows that she misunderstood it, so someone could think of an apparently correct interpretation of some merely descriptive statement but still act in a way that shows that she misunderstood it; for example by drawing conclusions from it that are not licensed by what has been said.

A *third* answer to the question of how something in our mind can determine the correct use of the signs we are employing might be the ‘Platonist’ response that when we use a word or sentence with a particular meaning we are associating it with an abstract object, a concept or proposition respectively. This approach, which Wittgenstein associates with Frege (cf. BB, 4), suggests that ‘what must be added to the dead signs in order to make a live proposition is something immaterial, with properties different from all mere signs’ (ibid.). The abstract objects must have ‘properties different from all mere signs’ because

they have powers no 'mere sign' has: They can determine all the potentially infinite normative consequences the meaning of a sign has.

The problem with this account is not only its commitment to a dubious ontology but, more critically, that it is difficult to see how it can give an account of the *grasp* we need to have of the postulated abstract object that can avoid the regress already faced by the view that giving a sign meaning or understanding it consists in interpreting it.[16] How can something that is in my mind while I am giving or receiving the order 'Add 2!' determine that I am associating these words with the abstract rule to add two, rather than with the rule to add two until I have reached a million and three thereafter?

3.3 The metaphysics of meaning and understanding

In the light of the preceding considerations it should be obvious that the dual-process conception of the meaningful use of language cannot provide a satisfactory explanation of what it is to use a word or sentence with a particular meaning, or to understand someone else's use of it to have a particular meaning. Postulating mental processes or other occurrent conscious phenomena in order to explain what is distinctive about the meaningful use of language is explanatorily futile and phenomenologically inadequate.

At the root of the dual-process conception lie two category mistakes. *First*, when we give a sign a particular meaning we are not accompanying it with some kind of conscious mental process or event. We are rather *using* it in a specific way (cf. e.g. BB, 4; PI §43). Hence, when we say such things as 'By 'Napoleon' I meant Napoleon I not III' we are not reporting some mental event but what *use* we made of the sign 'Napoleon'. The same is true of predicative expressions such as, for example, 'is a gambler'. When we say such things as 'When I said that Napoleon was a gambler, I meant that Napoleon I had a habit of making risky political decisions, not that Napoleon III liked card games' we are again not

[16] For this point cf. Kripke (1982, 54).

reporting a mental occurrence that may or may not have accompanied our original utterance, but we are saying what *use* we made of the expression 'gambler' or what *role* 'gambler' played in our utterance. The use or role of a sign, however, is not something that exists apart from the sign when it is used and causes its employment. The use of a sign, however, is, as Wittgenstein puts it in the *Blue Book*, not anything 'co-existing' with it:

> The mistake we are liable to make could be expressed thus: We are looking for the use of a sign, but we look for it as though it were an object *co-existing* with the sign. (BB, 5)

Thus, the first category mistake at the bottom of the dual-process conception is its confusion of the use of a sign with an occurrent conscious phenomenon accompanying its use. The distinction between meaningful and meaningless marks and sounds, and the distinction between the utterances of a parrot and those of a competent speaker do not lie in the presence or absence of a conscious accompaniment of the production of signs, but in their role or use.

Second, the identification of understanding with a conscious mental occurrence is just as much a category mistake as the identification of meaning with such an occurrence. Linguistic understanding is a *potentiality* rather than an actuality. It is, according to Wittgenstein, a kind of ability:

> The grammar of the word "know" is evidently closely related to the grammar of the words "can", "is able to". But also closely related to that of the word "understand". (PI §150)
>
> [Understanding], the knowledge of the language, isn't a conscious state that accompanies the sentences of the language. Not even if one of its consequences is such a state. It's much more like the understanding or mastery of a calculus, something like the *ability* to multiply. (PG, 50)

Understanding or knowing the meaning of an *expression type* consists in a manifold set of abilities or capacities, such as *inter alia* the capacities to use the expression correctly, to draw the correct inferences from its use on a particular occasion and to respond to its use on a particular occasion appropriately; where 'responding to the use of an expression

appropriately' does not mean responding to it in a socially accepted manner but showing sensitivity to its meaning. One does not have to have all of these abilities in order to count as understanding an expression; one may, for example, count as understanding German without being able to speak it. However, one must have at least some of the abilities that are constitutive of understanding. For example, in order to count as understanding German one must be able under normal circumstances to report what a German speaker said, although the report itself need not be in German. Understanding *sign tokens*, i.e. words and sentences on particular occasions of their use, also consists in a variegated array of abilities, such as being able to report or paraphrase what has been said or respond to it in a way that is appropriate in light of its meaning.

Thinking of understanding as a mental phenomenon accompanying the use of language is a category mistake, because it is either a matter of confusing a potentiality, namely an ability, with an actuality, i.e. a mental process, or of committing the related mistake of thinking of a potentiality *as* an actuality.

Wittgenstein thinks that we make this latter mistake when we think of understanding as some kind of enduring mental state which underlies and causally explains the manifestation of understanding; a view which embodies a further form of the dual-process conception and which he addresses in PI §149:

> If one says that knowing the ABC is a state of the mind, one is thinking of a state of an apparatus of the mind (perhaps a state of the brain) by means of which we explain the *manifestations* of that knowledge. Such a state is called a disposition. But it is not unobjectionable to speak of a state of the mind here, inasmuch as there would then have to be two different criteria for this: finding out the structure of the apparatus, as distinct from its effects.

Two preliminary remarks are in order before commenting on this passage: First, Wittgenstein misuses the word 'disposition' in a way that obscures the very point he is trying to make when he calls a state of the 'apparatus of the mind' or the brain a 'disposition'. For a state of the brain (or the 'apparatus of the mind') is an actuality; a

disposition, on the other hand is a kind of power, i.e. a potentiality.[17] Second, this passage occurs in the context of a discussion of the concept of understanding, and it is quite clear that Wittgenstein's target is not only the idea that knowing is a state of the 'apparatus of the mind' but also the idea that understanding is. After all, linguistic understanding is a form of knowing, namely knowing the meaning of an expression or of what someone said. Accordingly, Wittgenstein says in the immediately following section that the 'grammars' of 'know' and 'understand' are 'closely related' (PI §150).

With these qualifications in mind, Wittgenstein's argument can be put as follows: Thinking of understanding as a state of the 'apparatus of the mind' or the brain which underlies its manifestations in the use of language and its comprehending reception means confusing an ability with its vehicle. It might be true that when we understand German our brain or our 'mental apparatus' is in a particular state[18], but this does not mean that it is identical with this state. If understanding German were identical with the underlying state of the brain that is correlated with it, then there would, at least potentially, be two criteria for whether someone understood German: first, what she says and does, and second, the state of her brain. This, however, is not the case. Even if we discovered, for example by means of neural imaging, that there was a brain state invariably correlated with the ability to understand German, this state, which we would presumably come to think of as the vehicle of this ability, would not be identical with the ability. For the only criterion for whether or not someone understood German would still be what she can do and not what state her brain is in. Hence understanding is not some kind of enduring state of the 'apparatus of the mind' or of the brain which underlies our use of language and its comprehending reception. Thinking otherwise, Wittgenstein appears to believe, is the result of a mistaken tendency to identify the powers of objects with states they are in:

[17] For this point see also Baker & Hacker (2005a, 373-4).
[18] Although Wittgenstein also says, on a different occasion, that we know 'next to nothing' about such states (cf. BB, 118).

> There are… various reasons which incline us to look at the fact of something being possible, someone being able to do something, etc., as the fact that he or it is in a particular state. Roughly speaking, this comes to saying that "A is in the state of being able to do something" is the form of representation we are most strongly tempted to adopt; or, as one could also put it, we are strongly inclined to use the metaphor of something being in a peculiar state for saying that something can behave in a particular way. And this way of representation, or this metaphor, is embodied in the expressions "He is capable of…", "He is able to multiply large numbers in his head", "He can play chess": in these sentences the verb is used in the *present tense*, suggesting that these phrases are descriptions of states which exist when we speak. (BB, 117)

Since understanding is an ability and therefore a potentiality rather than an actuality, it should not be identified with anything which accompanies speaking and listening, regardless of whether we think of this accompaniment as some kind of conscious mental process or as some kind of unconscious or neural state that causally explains the manifestations of understanding. This is true both for the understanding of sentence or word types, as well as for the understanding of sentence or word tokens. The fact that understanding often occurs at a specific time – normally at the time we hear an utterance, sometimes after we have thought about it for a while – is compatible with this contention; for although abilities are not actualities the onset or dawning of an ability is.

This last claim has been contested by Michael Dummett. He thinks that Wittgenstein is right to say that understanding of word and sentence types is a non-occurrent phenomenon but wrong in suggesting that understanding of utterance tokens is so too. He says:

> Grasping a sense, understood dispositionally, is evidently not a mental act, but a kind of ability: and this tallies with Wittgenstein's dictum that understanding, which he expressly compares to an ability, is not a mental process. But Frege allowed that, while thoughts are not mental contents, grasping a thought is a mental act, although one directed towards something external to the mind; and he must here be construing the notion of grasping a thought in its occurrent sense. Wittgenstein strove to dispel the idea that there is an occurrent sense of 'understand': but it is difficult to see how this can be successfully maintained. We cannot, for instance, simply reduce the conception of understanding an utterance to that of hearing it while possessing a dispositional understanding of the relevant words and constructions: for it is possible to be perplexed by a sentence on first hearing, through failure to take in its structure, and to achieve an understanding of it on reflection. (Dummett 1991, 274-5)

Dummett is right in claiming that understanding an utterance token is not the same thing as hearing it while possessing understanding of its constituent words and sentences. But that is not what Wittgenstein is claiming when he suggests that even the understanding of

utterance tokens is a non-occurrent phenomenon. As I suggested, understanding an utterance token consists in a variegated array of abilities, such as being able to report what has been said and being able to respond to it appropriately in action and judgement. It is perfectly possible to hear an utterance consisting of words whose meaning one knows without thereby acquiring these abilities, i.e. without understanding it. Furthermore, if Dummett is correct in claiming that there is an occurrent sense of the notion of grasping a thought (which I doubt), then 'grasping a thought' in this sense is not equivalent to 'understanding a thought' but to 'thinking a thought'. Thinking a thought is indeed an occurrent phenomenon but, although thinking a thought implies understanding it, it is not identical with the latter. Finally, when I have to figure out what someone meant by her words before understanding them, the process of figuring out what she meant is indeed an occurrent phenomenon but it is distinct from the understanding that is its upshot.[19]

4. Meaning, use and practice

What gives the signs of language their 'life', i.e. what makes them meaningful symbols rather than 'dead' sounds and marks, is, according to the later Wittgenstein, their use. This view is encapsulated in the following well-known remarks:

> But if we had to name anything which is the life of the sign, we should have to say it was its *use*. (BB, 4)
>
> For a *large* class of cases of the employment of the word "meaning" – though not for *all* – this word can be explained in this way: the meaning of a word is its use in the language. (PI §43)

I take it that these remarks, and others like them, are intended to suggest a positive conception of what it is that makes the sounds and marks we produce when using language meaningful; a conception that is intended to take the place of the family of views I have just discussed under the label of 'dual-process conception of meaning'. However, what

[19] For a criticism of Dummett along similar lines see Baker and Hacker (2005a, 381-82) and Rundle (2001, 109-10).

does it mean exactly to say that the meaning of a word is its use within the language? And how does this claim help us to overcome the dual-process conception?

One way we might understand the idea that the meaning of a word or sentence is its use is to see it as the claim that the meaning of words and sentences is determined by what people do with them. i.e. by what perlocutionary or illocutionary acts they perform with them. According to this conception a word or sentence acquires a meaning by being put to a specific use – where 'putting a word or sentence to a specific use' means using the word or sentence to achieve specific goals. In this sense of 'use' one might use a word such as 'Boche' in order to perform the illocutionary act of insulting someone or a sentence such as 'You do this so much better than I' in order to perform the perlocutionary act of getting someone to do something one feels little inclination to do oneself. I think it would be a mistake to interpret Wittgenstein as speaking about use in this sense. First, it seems that one can use words or sentences in order to achieve or try to achieve such a large variety of goals that it seems highly unlikely that the uses to which a word or sentence has been put or can be put, can give that word or sentence any clear meaning. Second, in the German original of the second remark quoted above, i.e. PI §43, Wittgenstein uses the word 'Gebrauch', as he usually does when speaking about the use and the meaning of linguistic expressions. The expression 'Gebrauch', however, is ill-suited to express the conception of use I have just characterized. If it had been Wittgenstein's intention to speak about this conception of use he would, I believe, rather have chosen a word such as 'benutzen' or the corresponding noun 'Benutzung'[20], which captures this sense of 'use' more adequately.[21]

[20] Wittgenstein usually uses the Austrian spelling 'Benützung'; for example in PI §43 ('employment' in the translation).

[21] The semantic differences between 'Gebrauch' (noun) or 'gebrauchen' (verb), on the one hand, and 'Benutzung' (noun) or 'benutzen' (verb), on the other, are subtle and difficult to render perspicuous. The task is further complicated by the fact that there are differences between the kinds of situation in which one would use the verbal expressions and the kinds of situation in which one would use the nominal expressions. At least a rough sense of the differences between the two pairs of expression might be provided by examples: Translating 'use' by 'gebrauchen' in 'Can I use your car to pick up granny?' would be wrong. Here the only

I believe the right way to think about Wittgenstein's conception of the use of an expression is to think of it as the distinctive *role* of the expression in all those situations in which it is or might be employed.[22] When he says that 'the meaning of a word is its use in the language' he should be understood as saying that a sound or mark acquires a specific meaning not through anything 'co-existing' with it but by playing a distinctive *role* within the 'language games' that characterize practice of speaking the language to which the sound or mark belongs. The meanings of the expressions 'slab', 'pillar', 'block' and 'beam' of the 'complete primitive language' of PI §2 can be understood in this way. To the degree that we can speak of meaning in the case of such an extremely primitive 'language' at all, the meanings of these expressions are determined by the distinctive role they play in the activities of the two builders A and B. The 'life' of 'block', 'pillar', 'beam' and 'slab' consists in them having a clear use or role in the intentional and cooperative activity of building something, a role or use meaningless or 'dead' sounds and marks, such as grunts or footprints, lack.

Having such a use or role in intentional human activities, however, does not appear to suffice to give sounds or marks meaning in any more full-blooded sense of this term. Clearing one's throat before giving an after-dinner speech or saying 'abracadabra' in the course of performing a magic trick clearly have a distinctive and established role in intentional human activities, but they do not constitute meaningful utterances. It is for this reason, I believe, that Wittgenstein says in the remark quoted above that 'the meaning of a word is its use *in the language*' and on another occasion that 'understanding a sentence means *understanding a language*' (BB, 5; emphasis added). In order for a sound or mark to carry meaning in the way the sounds and marks of human languages do, they have to not only have a distinctive role with respect to non-linguistic phenomena and behaviour, but

---

option seems to be 'benutzen'. In 'The use of the car to commute to work has significantly increased since the 1970s', on the other hand, we can also translate 'use' with 'Gebrauch'.

[22] For this point cf. Stroud (1996, 301-02).

also with respect to other linguistic behaviour. Unlike the sounds we produce when clearing our throat or saying ‘abracadabra’, words such as ‘dog’, ‘is’ or ‘and’ have not only a role within specific extra-linguistic contexts, such as an after-dinner speech or the performance of a magic trick, but make a distinctive contribution to all those different sentences in which they can occur. It is this distinctive role that words play in the generation of sentences, as well as the role that these sentences then play with respect to other sentences and within all the numerous activities that constitute our ‘form of life’ that gives them their meaning and distinguishes them from mere ‘dead’ sounds and marks.

Speaking and understanding a language is for the later Wittgenstein a matter of engaging in a *practice* (cf. PI §202). In order to engage in such a practice or to participate in the *custom* of speaking a language (cf. PI §199) one has ‘to have mastered a technique’ (PI §199). The ‘technique’ involved is the technique of using language correctly and of responding to someone else’s use of language in appropriate ways. In the case of the language of the builders of PI §2, the practice or custom the builders are engaged in is the practice of ordering specific types of stone by using the signs ‘beam’, ‘pillar’, ‘block’ and ‘slab’. The technique they have mastered is, on the one hand, the technique of requesting blocks, pillars, slabs, and beams by saying ‘block’, ‘pillar’, ‘slab’ and ‘beam’, and, on the other hand, the technique of delivering the right type of building stone upon an appropriately issued request.

In order for someone to engage in a practice or a custom there has to be such a thing as not engaging in it, as well as such a thing as acting in accordance with it or acting contrary to it. Similarly, in order for us to be able to speak about someone as having mastered a technique, there has to be such a thing as not mastering it, as well as such a

thing as carrying out the technique in question correctly or incorrectly.[23] The distinction between, on the one hand, someone not engaging in a practice and not having mastered a technique and, on the other, someone violating a practice or carrying out a technique incorrectly is, of course, rather vague. It seems fair to say that someone, who calls out 'Slab!' or 'Block!' completely randomly, is not participating in the practice the builders A and B of PI §2 are engaged in, and has not mastered the technique of ordering building-stones by the use of these expressions. Of someone who only uses 'beam', 'pillar', 'block' and 'slab' in the context of building something but who occasionally uses 'Pillar!' when she desires the kind of stone that is usually delivered upon 'Block!', on the other hand, we would probably rather say that she makes a mistake; in other words, that she is engaged in the same kind of practice A and B are engaged in, but that she occasionally violates it, that she occasionally carries out the technique of ordering a block incorrectly.

The claim that for every practice or technique there has to be such a thing as doing the things that constitute the practice or count as carrying out the technique correctly or incorrectly, amounts to the claim that for every practice or technique in this sense there have to be *rules* or *norms* that determine what counts as acting in accord with the practice and as carrying out the technique. Accordingly, saying that speaking a language is a practice we engage in by mastering a technique or set of techniques is tantamount to saying that it is an activity that is governed by *rules* or *norms*. Someone who is engaged in the practice of speaking a language by carrying out the technique of using words and sentences is *following the rules* governing the use of the words and sentences she is employing.

Saying that someone follows the rules for the language she is speaking implies not merely that there is a standard of correctness for the use of the words she is employing, but also that she is aware of this standard. Only because there is a standard of correctness for

[23] For this point cf. Stroud (1996, 302-3) and Backer & Hacker (2009, 144-45).

the use of the words ‘green’, ‘ideas’, ‘sleep’ and ‘furiously’ can we say of someone who says ‘Green ideas sleep furiously’ that she is using these expressions incorrectly. And only of someone who is aware of the standard for the correct use of, say, ‘green’ can we say that she understands this word and that she uses it with its ordinary meaning. Speaking a language is not merely a matter of producing signs in conformity with the rules of one’s language, one also has to know these rules.

I believe it is a central insight of Wittgenstein’s later work that the standard for the correct use of linguistic expressions is ultimately provided by how those expressions are in fact used. The norms or rules that govern the use of a language are, as it were, immanent to the practice of speaking the language and not external to it. Speaking a language is not like playing a board game with its rules explicitly laid down in a handbook. The rules for the use of the expressions of a natural language are not determined by what grammarians say about them but by how competent speakers use them and by which uses they treat as correct or incorrect.

Accordingly, grasp of the rules of one’s language does not consist in having formulations of those rules explicitly in mind, but in an ability to distinguish between correct and incorrect uses of a word in practice and in judgement. It consists in the ‘practical’ ability to apply words correctly on specific occasions and to correct misapplications when they occur, as well as the ‘theoretical’ ability to recognize correct applications of a word as correct and incorrect one’s as incorrect. A point Wittgenstein makes somewhat cryptically when he says:

> [T]here is a way of grasping a rule which is *not* an interpretation, but which, from case to case of application, is exhibited in what we call “following the rule” and “going against it”’ (PI §201).

Assuming that grasping the rules governing the use of the expressions of one’s language is a matter of knowing their interpretation leads straight to the vicious regress I have already

discussed. If grasping the rules of one's language were a matter of interpretation, then one would have to have an explicit formulation of the rules in mind when using a language, or at least be able to provide such an explicit formulation. The rules governing the expressions occurring in this formulation would then have to be grasped as well, which would require having a further formulation of rules in one's mind or being able to provide a further formulation of them and so on *ad infinitum*. It is for this reason that Wittgenstein suggests that knowledge of the rules of one's language is a kind of ability, namely the ability to distinguish between correct and incorrect applications of those rules; or, in other words, that knowledge of the rules of one's language is a form of knowledge-how rather than a form of knowledge-that.

Finally, the claim that language is a norm-governed practice implies that there has to be an established and regularly practiced way of using the sounds and marks making up natural languages before these sounds and marks can be said to have any meaning at all. Only against the background of certain established and highly complex regularities of conduct involving the production of the sounds and marks making up natural languages can they be said to mean anything. Ultimately saying that speaking a language is a practice or custom is for Wittgenstein of a piece with claiming that the use of language is only possible against the background of a complex set of established regularities of conduct:

> It is not possible that there should have been only one occasion on which only one person followed a rule. It is not possible that there should have been only one occasion on which a report was made, an order given or understood, and so on. – To follow a rule, to make a report, to give an order, to play a game of chess, are *customs* (usages, institutions). (PI §199)

In order for the 'complete primitive language' of the builders of PI §2 to count even as a primitive language the expressions making up this language, i.e. 'beam', pillar', 'block' and 'slab', have to have been used repeatedly in certain regular ways. Only because these

expressions are involved in certain regularities of conduct of ordering and passing stones in the course of building something, can they be said to have any meaning at all.

It is important to notice that for the Wittgenstein of the *Investigations* both the conventional meaning of words, sentences or utterance *types* as well as what we mean by the use of *tokens* of them on a particular occasion is determined by the role or use of these words, sentences or utterances within a linguistic practice.[24] Even when, on a particular occasion, we use an expression, whose conventional meaning is ambiguous, it is, according to Wittgenstein, not any mental accompaniment that determines what we mean by it, but the use we are making of the expression on that occasion, or the role it plays within the linguistic activities we are engaged in. The role or use of an utterance, in turn, is determined by the 'wider context' in which it is made (cf. PI §686) as well as by the dispositions of the speaker (cf. PI §684). Whether I meant Napoleon I or Napoleon III by a particular utterance of 'Napoleon', or whether I meant 'financial institution' or 'bank of the river' by a particular utterance of 'bank' is thus determined by the circumstances in which my utterance occurred, as well as by how I would respond to questions about what I meant. Since subsequent linguistic performances belong to what Wittgenstein calls the 'wider context' of an utterance, a speaker's statement about what she meant by a particular past utterance can be partly constitutive of what it is purporting to report. At least, this is what Wittgenstein seems to suggest in PI §682, where he discusses the utterance of 'It'll stop soon' by someone who has toothache and is at the same time annoyed by the sound of piano tuning in an adjacent room:

[24] I am here addressing, roughly, the difference between what H. P. Grice calls the 'conventional meaning' and the 'occasion meaning' of utterances (1989, 118-19). Grice puts the difference between conventional meaning and occasion meaning in terms of two different ways of characterizing the meaning of utterances. We speak about conventional meaning in statements of the form "X (utterance-type) means '…'" (ibid, 118); about occasion meaning in statements of the form "U (utterer) meant *that*…" (ibid.).

> "You said, 'It'll stop soon'. – Were you thinking of the noise or of your pain?" If he answers, "I was thinking of the piano tuning" – is he stating that the link existed, or is he making it by means of these words? – Can't I say *both*? If what he said was true, didn't the link exist – and is he not for all that making one which did not exist?

What the speaker in this example means or what he refers to by his use of 'It' is not only reported but partly determined by his subsequent remark that he was thinking of the piano tuning rather than the toothache. This does not mean that Wittgenstein is a kind of anti-realist about occasion meaning. He does not say that the truth of a report about what one meant is not answerable to anything outside of that report itself. He merely points out that subsequent reports are one of the many elements in the context of an utterance which determine its meaning. To be true, however, the report has to cohere with other elements of that context, such as statements and actions preceding the utterance or the dispositions of the speaker at the time of the utterance (cf. PI §684).

The claim that speaking a language is a matter of engaging in a practice not only provides a plausible picture of what gives the signs of our language their 'life' or meaning, which avoids the problems of the dual-process conception, it also explains a range of phenomena that are particularly troubling for the latter view. First, it explains why it is impossible to do such things as to mean 'toothache' by saying 'abracadabra' (cf. PI §665) or to say '"bububu" and mean "If it doesn't rain, I shall go for a walk"' (boxed remark after PI §35). Signs are given their meaning by having an established role within a language. Accordingly we cannot mean anything by using signs that do not have such a role. The only way we can give sounds or marks such as 'bububu' a particular meaning is by establishing a specific role for them within our language by connecting them with signs that already have such a role through a public act of definition (cf. PI §665).

Second, the claim that speaking a language is a matter of engaging in a practice explains what I above called the 'normative character' of linguistic meaning. As we have

seen, the use of words is governed by certain norms or rules which are instituted by the practice of using those words. Unlike the dual-process conception assumes, these norms cannot be instituted by what goes on in speakers' minds when they use language, but they can be instituted through shared practices or customs. Unlike many philosophers assume, the institution of norms of conduct through shared practices or customs is an utterly non-mysterious phenomenon. As Wittgenstein points out we can easily imagine people playing a game with a ball according to rules, which they 'make up as [they] go along' (cf. PI §83). In other words, they play a game according to rules which are not explicitly stated at any point prior to or during the game, but which are nevertheless implicit in their actions and in what actions they treat as correct or incorrect. Similarly the rules of natural languages are implicit or immanent in the practice of speaking those languages. And it is these rules that tells us such things as how a word is *correctly* used, what we *commit* ourselves to by employing it and what it is *correct* and what *incorrect* to infer from its use on a particular occasion. Engaging in a linguistic practice allows us to use words so as to determine all these consequences.

4.1 Meaning scepticism

I have presented Wittgenstein's famous rule-following paradox as one of his reasons for rejecting the dual-process conception of linguistic meaning. I have also pointed out that Wittgenstein's remarks on rules and following rules, in particular in PI §§185 to 242 and the *Remarks on the Foundations of Mathematics*, chapter VI, are intended to convey a variety of further insights; namely, that one can only mean something by one's use of a linguistic sign against the background of an already existing linguistic practice, and that one can only ascribe meaning to and understand the utterances of persons who share our 'form of life' (cf. PI §241). Finally, I said that many interpreters ascribe to Wittgenstein the

further aims of showing that language is essentially social and of advancing a sceptical challenge to the very idea of linguistic meaning.

The dispute over whether or not language is *essentially* social and whether or not Wittgenstein thought it was is, I believe, not very important and is certainly not pertinent to the issues of this chapter. Ordinary human languages are obviously social, what their expressions mean depends on their use within the linguistic community. And it is with the nature of ordinary human languages that Wittgenstein is mostly concerned. Any language that was never used in conversation and was only spoken by one solitary speaker would presumably look very different from ordinary human languages.[25] The claim that Wittgenstein rightly thought such a language to be impossible, however, strikes me as contentious at best.[26] On the other hand, the claim that Wittgenstein advocated a general scepticism concerning the notion of linguistic meaning, which has been famously advanced in Saul Kripke's *Wittgenstein on Rules and Private Language* (Kripke 1982) strikes me as interesting and important. For it presents a clear challenge to my interpretative claim that according to the later Wittgenstein the meaning of linguistic expressions is their use.

According to Kripke the scenario of the deviant pupil in PI §185 shows not only, as I have argued above, that no *interpretation* one may have in mind can give an expression of a rule or any other kind of statement a particular meaning, but suggests the much stronger conclusion that there is nothing that could give linguistic expressions any determinate meaning. As he puts it, Wittgenstein's rule-following paradox, i.e. the paradox Wittgenstein expresses when he says: 'This was our paradox: no course of action could be determined by a rule, because every course of action can be brought into accord with the

[25] This point has been well made by Cora Diamond in Diamond, 1989.

[26] It is, for example, difficult to reconcile with the following remark from PI §243: 'A human being can encourage himself, give himself orders, obey, blame and punish himself; he can ask himself a question and anwer it. So one could imagine human beings who spoke only in monologue, who accompanied their activities by talking to themselves. – An explorer who watched them and listened to their talk might succeed in translating their language into ours. (This would enable him to predict these people's actions correctly, for he also hears them making resolutions and decisions.)'

rule' (PI §201), points towards the radical conclusion that '[there] can be no fact as to what I mean by… any… word at any time' (Kripke 1982, 21).

Kripke argues for this radical conclusion with the help of an example, which is inspired by Wittgenstein's example of the deviant pupil. He introduces the mathematical function of 'quaddition', which yields the sum of any two numbers x and y, if and only if x and y are both less than the highest numbers for which the function has been actually computed, otherwise it yields 5. It is imagined that for '68 plus 57' only the second condition is fulfilled (cf. ibid, 9). Kripke now imagines a sceptic who claims that whenever we used 'plus' in the past, we actually meant quaddition and that therefore we have to answer not with the sum but with the 'quum' to such questions as 'What is the value of 68 plus 57?' or any other question apparently asking for the sum of two numbers at least one of which is higher than any we have previously computed.

The question now is what *reason* we can give the sceptic that we actually mean addition by 'plus' rather than quaddition; or in Kripke's terminology what *fact* about us we can cite to support our claim that we mean addition. It seems that nothing that goes on in our minds when we use 'plus' can be cited as a reason for the claim that we meant addition rather than quaddition. For, as we saw, whatever goes on in our minds appears to be compatible with all sorts of uses of the word 'plus', among them uses which are clearly incompatible with its meaning addition. Furthermore, Kripke claims, facts about our dispositions with respect to the use of 'plus', or facts about our past use of that expression cannot determine its meaning either. First, since by stipulation past uses of 'plus' or '+' did not involve the numerals '57' and '68' they do not provide us with a precedent for the answer to the question 'What is 57 plus 68?'. Consequently Kripke concludes there is no fact about our past use of 'plus' that tells us whether to answer this question by giving the sum or the 'quum' of 57 and 68. Second, the fact that we are disposed to give the sum

rather than the quum, and that we may even have been disposed to do so in the past had the question arisen, does not tell us that giving the sum is the *right* answer (cf. ibid, 23ff). The relationship between the question 'What is 57 plus 68?' and its answer ('125') is *normative* (cf. ibid, 37); only one answer, namely '125', is *correct* all others are *incorrect*. If by 'plus' we simply meant that mathematical function that yields the answers we are actually disposed to give to such questions as 'What is 57 plus 68?', then it seems that the idea that some answers are correct and others incorrect, and that we sometimes make an arithmetical mistake when answering such a question makes no sense anymore. It seems that whatever answer we are disposed to give is, by definition, the correct answer. As Wittgenstein puts it: '[Whatever] is going to seem correct to me is correct' (PI §258); and that, as he rightly concludes, 'only means that here we can't talk about 'correct'' (ibid.).

Since neither facts about what goes on in my mind when I use 'plus', nor facts about my past use of 'plus', nor facts about my dispositions with respect to 'plus' can tell me whether by 'plus' I mean quaddition or addition, Kripke concludes that 'there is no fact about me that distinguishes between my meaning plus and my meaning quus' (1982, 21). And since problems like those illustrated by the discussion of 'plus' and 'quus' can be created for all linguistic expressions, Kripke ascribes to Wittgenstein the radically sceptical conclusion that '[there] can be no fact as to what I mean by 'plus', or any other word at any time' (ibid). The 'entire idea of meaning', as Kripke puts it, 'vanishes into thin air' (ibid, 22). It is this sceptical conclusion that according to Kripke Wittgenstein expresses when he says:

> This was our paradox: no course of action could be determined by a rule, because every course of action can be brought into accord with the rule. The answer was: if every course of action can be brought into accord with the rule, then it can also be brought into conflict with it. And so there would be neither accord nor conflict here. (PI §201)

The sceptical interpretation of this remark, which Kripke provides, is obviously incompatible with the claims I made above, that linguistic expressions acquire a meaning by being used in specific ways and that how someone means an expression on a specific occasion depends on the circumstances of the utterance and the dispositions of the speaker. These claims ascribe a specific kind of *realist* conception of linguistic meaning to Wittgenstein and not a *sceptical* one. Accordingly, as I explained above, I do not interpret PI §201 as giving expression to a general form of *meaning scepticism,* but merely as criticising the idea that following a rule is a matter of interpreting it. As has often been pointed out,[27] this interpretation is much easier to align with how Wittgenstein continues this section than Kripke's interpretation. For immediately after stating the paradox, Wittgenstein appears to say that it rests on a mistake, which is overcome by abandoning the idea that following a rule requires interpreting it:

> That there is a misunderstanding here is shown by the mere fact that in this chain of reasoning we place one interpretation behind another, as if each one contended us at least for a moment, until we thought of yet another lying behind it. For what we thereby show is that there is a way of grasping a rule which is *not* an *interpretation.*… (PI §201).

The core problem with Kripke's interpretation is that he appears to assume that there could only be facts of meaning, i.e. such facts as an expression's or an utterance's having a meaning, if these facts could be identified in non-semantic and non-normative terms. He derives powerful arguments from Wittgenstein's writings, which show that it is impossible to explain what someone means by her use of a linguistic expression by merely citing such 'natural', i.e. non-semantic and non-normative, facts as past uses of the expression and the dispositions of the speaker or the community of speakers. Kripke is right to say that it is possible that while in the past people answered such questions as 'What is x plus y?' by giving what we would intuitively call their sum or followed such orders as 'Slab!' by

[27] See e.g. McDowell (1984, 331); Schroeder (2006, 199).

delivering a slab, it would be a mistake for them to give the sum or deliver a slab tomorrow. It could indeed be the case that people actually meant quaddition by 'plus' or understood the order 'Slab!' as 'Slub!', where the word 'slub' applies to slabs before some specified time *t* and to blocks thereafter. Furthermore, certain responses to questions of the type 'What is x plus y?' which we would normally classify as mistaken are obviously compatible with the dispositions of the speaker. After all, we are assuming that the speaker actually gives them (cf. Kripke 1982, 23). What is not possible is for people to *mean* addition by 'plus' or to be *following the rule* for addition in their use of 'plus' and it's not being correct to answer to 'What is 57 plus 68?' with '125' or it's being correct to answer with '5'.[28] There are many ways in which we can describe a person's past and present linguistic behaviour which are compatible with her using words with all sorts of different meanings or following all sorts of different rules. Yet, as soon as we describe her use of words in semantic or normative vocabulary, i.e. as soon as we say such things as 'By 'plus' she means *plus*' or 'She uses 'Slab!' according to the rule: Call 'Slab!' if you require a slab', this is not the case anymore. In other words, descriptions of linguistic behaviour in normative or semantic terms are not reducible to descriptions in non-normative or non-semantic terms. We cannot describe the linguistic activities of a person or a group of persons in entirely 'natural' terms and infer from this description what this person or group mean by their words. And since facts are the objects of true sentences we can therefore conclude, in the light of Kripke's arguments, that normative or semantic facts are irreducible to natural, i.e. non-normative and non-semantic, facts.

Insofar as Kripke's arguments show this irreducibility of intentional and normative facts to natural facts they have captured a genuinely Wittgensteinian thought. However, his inference from this observation to the conclusion that there are no facts about what

[28] For this point see Stroud (1996, 309).

someone means by her use of linguistic expressions, or about the rules she is following in her use of them, seems both philosophically problematic and mistaken as an interpretation of Wittgenstein. It is philosophically problematic in that it simply assumes that non-normative or non-semantic facts are somehow the only genuine facts there are; that statements such as ‘‘Plus’ means addition’ or ‘She is following the rule for addition in her use of ‘plus’’ never report any genuine facts unless they can be reduced to statements containing no normative or semantic vocabulary.[29] The assumption that the only real facts are natural facts, however, expresses a commitment to a very strong form of philosophical naturalism. A commitment one does not have to share and that needs to be supported by arguments which Kripke does not provide.

Contrary to what Kripke appears to assume, Wittgenstein does not believe that the irreducibility of semantic statements to non-semantic and non-normative statements shows that the former do not express genuine facts. As he puts it ‘[following] according to the rule is FUNDAMENTAL to our language-game’ (RFM, 330). Describing people as following certain rules of language and other rules of conduct is an irreducible and indispensable form of description. We cannot describe what someone is saying or what rules she is following in her linguistic behaviour without using semantic or normative vocabulary. We cannot translate statements such as ‘‘Grün’ means *green*’, ‘She is following the rule: Deliver a slab when the builder shouts ‘Slab!’’ or ‘By ‘plus’ she means addition’ into statements free of normative and semantic vocabulary. But for Wittgenstein this is no reason to conclude that statements such as these do not describe genuine facts. He rather suggests that one should resist the temptation to reduce talk in normative or semantic terms to talk in non-normative and non-semantic vocabulary. In the *Remarks on the Foundations of Mathematics* he puts this as follows:

[29] For this point see Stroud (1996, 311-13).

> To what extent can the function of language be described? If someone is not master of a language, I may bring him to a mastery of it by training. Someone who is master of it, I may remind of the kind of training, or I may describe it; for a particular purpose; thus already using a technique of the language.
>
> To what extent can the function of a rule be described? Someone who is master of none, I can only train…
>
> The difficult thing here is not, to dig down to the ground; no, it is to recognize the ground that lies before us as the ground. (RFM, 333)

The problem with Kripke's interpretation is that it fails 'recognize the ground that lies before us as the ground'. He thinks that if there are no facts about the dispositions and states of a person or about the use she and others within her linguistic community have previously made of the expressions she is employing on a particular occasion which can be stated in non-normative and non-semantic terms, then there could be no fact about her meaning anything specific by her use of those expressions on that occasion. However, when Wittgenstein advises us 'to recognize the ground that lies before us as the ground' and – on another occasion – not to dig below 'bedrock' (cf. PI §217), he suggests that we should recognize descriptions of human behaviour in normative and semantic terms as basic and not even in principle replaceable by descriptions not containing vocabulary of this sort.

It is important to recognize, though, that the claim that semantic and normative talk is irreducible to non-semantic and non-normative talk does not imply that when we ascribe a specific meaning to an expression or utterance we are referring to something that exists besides its rule-governed use. It just means that a description of the use of an expression in non-semantic and non-normative terms is not a specification of its meaning.

Closely related to the idea that semantic facts are irreducible is the idea Wittgenstein expresses in the above quote by saying that the 'function of language' or the 'function of a rule' cannot be described to anyone who does not already know a language or a rule. Wittgenstein is not making the banal point that only someone who speaks a language will be able to understand such a description. He is rather pointing out that when we specify the

meaning of an expression we can only do so by employing other expressions or combinations expressions, which have the same or at least a closely related meaning; and that the person who understands this specification of meaning already has to know the meaning of these other expressions. Statements such as ‘‘Plus’ means addition’ or ‘‘Und’ means *and*’ will only succeed in specifying the meaning of ‘plus’ and ‘und’ for someone who already has the concept of addition and the concept of conjunction. In other words, these statements can only be understood by someone who knows the meaning of ‘addition’ and ‘and’. Knowing the meaning of ‘addition’, however, means, among other things, being able to use a word that has the same role as ‘plus’ and ‘+’ have within the English language or within mathematical notation; and knowing the meaning of ‘and’ means being able to use it. Accordingly no one can specify the meaning of an expression for us, which cannot be specified in terms we do not already know how to use. If we are to learn the meaning of such an expression and acquire the concept for which it stands, we have to be *trained* in its use.

One thing that Wittgenstein’s example of the deviant pupil and Kripke’s example of ‘quus’ bring out quite clearly is that there is nothing to this training that makes it inevitable that we will always use the expression we are trained to use in the same way as our trainer or other members of our linguistic community use it. It is always possible that at some point we start using an expression according to some bizarre quus-like rule. If that were to happen, it would indeed turn out that we used an expression with a different meaning than our linguistic peers were likely to expect. If it happened on a massive scale the possibility of communication would break down.

However, the possibility of such a break-down of communication does not mean that statements about what certain expressions mean within a language or about what someone means by her use of certain expressions are always mistaken; that they appear to report a

fact, which does not obtain. It merely shows that their truth is contingent and that it is conceivable that they turn out to be false. Furthermore it shows that communication, understanding and the ascription of a specific meaning to an expression or to a specific utterance is only possible against the background of what Wittgenstein calls a shared 'form of life' (cf. PI §241). We can only ascribe a specific meaning to the utterances of a person who is disposed to use the expressions making up her utterance in the same way we use them, or would use them if we spoke her language. In order to be able to say of someone that she means addition by 'plus' she must use 'plus' in roughly the same kinds of situation we would use it, she must agree with us on such statements about the meaning of 'plus' as ''Plus' means addition' and she must agree with us on the answers to at least a substantial subset of questions of the kind 'What is x plus y?', especially if these questions ask for the sum of numbers which are easily computable (cf. PI §242). It is such an agreement in our 'form of life' (cf. PI §241), which according to Wittgenstein provides 'the scaffolding from which our language operates' (PI §240).

5. Conclusion

In the first chapter I have tried to show that there are two crucial assumptions in the philosophy of mind underlying the picture theory of the proposition of the *Tractatus*. First, the early Wittgenstein assumed that our thoughts consisted of signs in a mental symbolism; that they were, as he put it in an often-quoted letter to Russell, constituted by 'psychical constituents that have the same sort of relation to reality as words' (NB, 130-31). Second, he assumed that 'propositional signs' were 'projected' onto the states of affairs they represented by acts of 'thinking the sense of the proposition' (cf. TLP 3.11). These acts of thinking were taken to establish a 'pictorial relationship' (cf. TLP 2.1513-2.1514) between propositional signs and states of affairs by projecting the ultimate simples of language, i.e.

names, onto the ultimate simples of reality, i.e. the simple objects that make up the 'substance of the world' (cf. TLP 2.021).

In the present chapter I have tried to present the later Wittgenstein's criticism of these two assumptions. While I have dealt with the first assumption, i.e. the idea that thinking consists in the operation of mental signs in a 'language of thought', roughly in the form it presented itself in the writings of the early Wittgenstein, I have discussed the second assumption in what might be conceived as a slightly different guise, namely in form of the idea Wittgenstein discusses in the *Blue Book* that mental acts, states or processes of meaning and understanding are necessary to give the signs of language their 'life' or meaning. Finally, I tried to elaborate and defend the later Wittgenstein's claim that the meaning of a sign is its use or role in the language.

However, from a Tractarian point of view the criticism of what I called the 'dual-process conception' of meaning as well as the idea that the meaning of linguistic signs is their use, which I suggested replaces this conception, might appear unsatisfactory. One might object that the point of the postulation of mental acts of meaning in the *Tractatus* is not to explain how 'mere dead signs' acquire a 'life', but to connect the names making up elementary sentences with the objects that *are* their meaning. After all, according to the *Tractatus*, a 'name means an object. The object is its meaning' (TLP 3.203). From the perspective of this remark the claim that the meaning of linguistic expressions, including names, is their use seems simply wrong-headed. Clearly the object a name stands for cannot be the use of the name. Thus, there appears to be a fundamental difference between the conception of meaning operative in TLP 3.203 and the conception of meaning advocated in the *Investigations*. And indeed, the later Wittgenstein points out that to use 'meaning' as he himself does in TLP 3.203 is a mistake, since it confuses the meaning of a name with its bearer: 'It is important to note that it is a solecism to use the word "meaning"

to signify the thing that 'corresponds' to a word. That is to confound the meaning of a name with the *bearer* of the name' (PI §40; italics in the original).

However, rejecting the idea that mental acts of thinking connect the names occurring in elementary sentences with the objects that are their meaning completely undercuts the whole tractarian account of how language relates to reality. For how, if the meaning of simple names is *not* identical with the objects they stand for, can what a true proposition says still be *identical* with the fact that makes it true? How can propositions 'reach right out to' reality (cf. TLP 2.1511)? These are the sorts of questions that will be the topic of the next chapter.

III. Intentionality

> Like everything metaphysical the harmony between thought and reality is to be found in the grammar of the language. (PG, 162)

1. The rejection of the 'pictorial relationship'

The beauty of the account of intentionality provided by the picture theory of the *Tractatus* lies in its explanation of how a proposition or a thought 'reaches right out' to reality (cf. TLP 2.1511). The picture theory elegantly combines the idea that every proposition with a sense is *bipolar*, i.e. that it *can* be true *or* false and is therefore independent of the reality it represents, with a strong form of direct realism, namely the idea that what a proposition says just *is* what is the case, if it is true. According to the picture theory a proposition 'represents a possible situation in logical space' (TLP 2.202), which is it sense – at least according to some passages of the *Tractatus* (cf. TLP 2.221, 4.031). If the 'situation' or 'state of affairs' it represents is not merely possible but actual, or, in other words, a fact, then the proposition is true. So what a true proposition represents, i.e. its sense, simply *is* the fact that makes it true. On the other hand, since it is one of the cardinal insights of the *Tractatus* that a proposition does not represent the state of affairs that is its sense in the same way a name stands for the object that is, according to the *Tractatus*, its meaning, the sense of a proposition is independent of its truth. As Wittgenstein puts it, '[in] a proposition a situation is, as it were, assembled by way of experiment' (TLP 4.031, my translation). Only if the objects that are the meanings of the simple names making up a proposition are 'assembled' in the same way as the names are 'assembled' in the proposition, is the sense of the proposition the very fact that makes it true; if they are not so assembled what the proposition says is a merely possible fact. Finally, as I have tried to show in the first chapter, this analysis of propositions as 'logical pictures' extends to the mental phenomenon of thinking and, as Wittgenstein's invocation of the picture theory in

his criticism of Russell's theory of *belief* and *desire* suggests, presumably also to other intentional mental phenomena.

Just as Sellars claims in 'Being and Being Known' (BBK, 41-2)[1], the picture theory allows the early Wittgenstein to steer between the Scylla of the idea advocated by Russell in *The Analysis of Mind*[2] that when we think or believe something we are related to a content, which is a 'present [occurrence] in the believer' (Russell 1921, 233) and only externally related to the 'objective' in the world which makes our belief true, and the Charybdis faced by such philosophers as the 'new American realists', i.e. the idea that there have to be subsistent complex entities and states of affairs which are the content of our false thoughts and beliefs. Unlike Russell and other adherents of the first view, the early Wittgenstein can explain how we can desire the very thing that would fulfil our desire and believe exactly that which would make our belief true. Unlike the new realists and other adherents of the second view, Wittgenstein did not have to postulate subsisting but non-existent states of affairs and complex objects corresponding to our false beliefs.

Still the picture theory is not entirely free of the kind of metaphysical assumptions that characterize the views of 'extreme realists' (BBK, 42) such as Holt, Montague and Perry and their collaborators, since Wittgenstein did postulate necessarily existing simple objects which 'make up the substance of the world' (TLP 2.021) and correspond ineffably in their 'logical form' to the simple elements of fully analysed propositions, i.e. names. The postulation of these sempiternal objects constituted a metaphysical leap of faith; as is made evident by the fact that Wittgenstein never came forth with even a single example of such an object.

He presumably thought that he did not have to be able to give examples of the postulated simple objects because he had valid *transcendental* reasons for assuming their

[1] Cf. Chapter I.
[2] Russell 1921.

existence. He clearly thought that the picture theory provided the only possible explanation of how language and mind come to represent the world. It required the existence of simple sempiternal objects corresponding to the simple elements of fully analysed propositions. Therefore, there had to be such objects.

In particular the idea that the linguistic picture, i.e. the fully analysed proposition, 'includes the pictorial relationship' (TLP 2.1513) which correlates 'the picture's elements with things' (TLP 2.1514) seems to necessitate the existence of simple objects corresponding to the simple elements of the picture, i.e. the names. Since these 'correlations of the picture's elements with things' (ibid) are, as internal relational properties of the picture, essential to the picture, and since they would not exist if there were not relata 'in the world' which they related to the elements of the picture, the simple objects have to exist if picturing is to possible at all.

I have argued in the first chapter that the early Wittgenstein thought of this 'pictorial relationship' as being brought about by acts of 'thinking the sense of the proposition' (cf. TLP 3.11). I have also pointed out that this means that his conception of this relationship is in tension with a further commitment of the *Tractatus,* namely the idea that 'thinking is a kind of language' and a thought 'a kind of proposition' (NB, 82) consisting of 'psychical constituents that have the same sort of relation to reality as words' (ibid, 131). If whenever we employ a proposition we have to accompany it by an act of thinking its sense and this act of thinking the sense consists of elements that 'have the same sort of relation to reality as words' then it too has to be accompanied by an act of thinking *its* sense, and so on *ad infinitum*. Thus, the combination of the ideas that (i) the pictorial relationship between proposition and state of affairs is established by acts of thought and that (ii) acts of thought are kinds of propositions which stand in the same kind of relation to what they depict as the propositions of public languages leads to a vicious regress.

Furthermore, the account of the pictorial relationship Wittgenstein gives in the *Tractatus* is clearly an instance of the 'dual-process conception of meaning' and therefore incompatible with his later rejection of this conception. As I have tried to show in the previous chapter, according to the Wittgenstein of the *Investigations* no mental accompaniment can give a sign its meaning or sense. Whatever mental act, state or process we imagine to accompany the use of some linguistic expression, its occurrence is compatible with the expression in question having all sorts of different meanings.

Finally, the idea that the pictorial relationship is part of the picture (cf. 2.1513) and connects the names with the elements in reality that are their meaning (cf. TLP 3.203) is incompatible with Wittgenstein's later view that the meaning of a sign is not an 'object *co-existing* with the sign' (BB, 5, italics in the original) but its use (cf. ibid., 4). Hence, whereas in the *Tractatus* Wittgenstein claims that a 'name means an object. The object is its meaning'; he says in the *Investigations* that '[it] is important to note that it is a solecism to use the word "meaning" to signify the thing that 'corresponds' to a word. That is to confound the meaning of a name with the *bearer* of the name' (PI §40; italics in the original). And only a few paragraphs later he famously concludes: 'For a *large* class of cases of the employment of the word "meaning" – though not for *all* – this word can be explained in this way: the meaning of a word is its use in the language' (PI §43).[3]

[3] Wittgenstein's talk about the meaning of names may occasion dissent, since one might think, as for example J. S. Mill did, that names do not have a meaning. P. M. S. Hacker and G. P. Baker for example, who rarely disagree with Wittgenstein, claim that he made a mistake when he talked about the meaning of names (cf. Baker and Hacker 2005a, 244-49). I think, however, that regardless of what one's views on the idea that names have meaning are, one still ought to grant the distinction Wittgenstein draws in PI §41 when he distinguishes between the meaning of a name and its bearer; for the main purpose of this distinction is to highlight the difference between the referent of a name and the properties of a name in virtue of which it has its referent and makes its characteristic contribution to the sentences in which it occurs. I will not go into the question of whether or not names have meaning, but I take it that the distinction between the referent of a name and the properties of a name in virtue of which it fulfils its typical role ought to be recognized even by those who do not think that names have a meaning. Accordingly, when I talk about the meaning of a name and say that it consists in its use I only make the claim that it is in virtue of its use that serves its usual function.

However, it is not only the idea that the pictorial relation is established through acts of 'thinking the sense of the proposition' and the correlated idea that it connects names with the objects that are their meanings that turn out to be mistaken in the light of Wittgenstein's later views, but also the idea that the picture 'includes the pictorial relationship' (TLP 2.1513). In *Philosophical Grammar* Wittgenstein brings this mistake out by distinguishing between the 'method of projection' and the 'projection lines' (cf. PG, 213). He suggests that when one imagines a worker constructing an artefact from a blueprint one might call 'the way in which the workman turns such a drawing into an artefact "the method of projection"' (ibid). One might think of the method of projection as mediating between the drawing and the object that is to be made from it and one may be inclined to compare it to 'projection lines' which reach from one object to another, i.e. from picture to artefact. Comparing the 'method of projection' with 'projection lines', however, can be misleading:

> This comparison conceals the fact that the picture *plus* the projection lines leaves open various methods of application; it makes it look as if what is depicted, even if it does not exist in fact, is determined by the picture and the projection lines in an ethereal manner; every bit as determined, that is to say, as if it did exist. (It is 'determined give or take a yes or no.') In that case what we may call 'picture' is the blueprint plus the method of its application. And we now imagine the method as something which is attached to the blueprint whether or not it is used. (PG, 213)

What Wittgenstein is saying here is that when one thinks of the 'method of projection' in terms of the 'lines of projection' one is easily tempted to think of a picture as consisting of a depiction plus a method of projection and of the method of projection as lines of projection which are, as it were, attached to the picture. If one now thinks of a picture as including the lines of projection one may be inclined to think that the picture in some sense includes at least the possibility of the reality it depicts; as Wittgenstein puts it, it determines what it depicts 'give or take a yes or no'. The projection lines connect the elements of the picture with the elements of the reality it depicts. The only thing that is

then left open is the question of whether these elements of reality are arranged as the picture depicts them.

Wittgenstein is here explicitly sketching doctrines of the *Tractatus*.[4] As I have discussed in the first chapter, the *Tractatus* states that the 'method of projection' for linguistic pictures, i.e. propositions, is 'to think the sense of the proposition' (cf. TLP 3.11). These acts of thinking establish the 'pictorial relationship' between the proposition or picture and the state of affairs it represents, which is part of the picture (cf. TLP 2.1513) and which ensures that the picture 'reaches right out' to reality. The claims that (i) the picture 'includes the pictorial relation, which makes it into a picture' (ibid) and which 'consists of the correlations of the picture's elements with things' (TLP 2.1514) and that (ii) the simple elements of the picture share a 'logical form' with the simple elements of the reality it depicts imply that the picture 'contains the possibility of the situation that it represents' (TLP 2.203). All that a linguistic picture, i.e. a proposition, leaves open is whether the elements in reality that are correlated with the names making up the fully analysed proposition are arranged as they are depicted to be. Hence: 'A proposition must restrict reality to two alternatives: yes or no' (TLP 4.023).

In *Philosophical Grammar* Wittgenstein suggests that this line of thought rests on a confusion of the 'method of projection' with the 'projection lines' (cf. PG, 213-14).[5] When we imagine lines of projection between the blueprint used by the workman and the artefact he constructs from it, we merely illustrate in our imagination a particular method of

[4] A reference to the *Tractatus* occurs on the page preceding the quoted passage; cf. PG, 212.
[5] In the English version of *Philosophical Grammar* the fact that Wittgenstein takes himself to have *confused* the method of projection with the lines of projection is obscured by an infelicity in Anthony Kenny's translation. The relevant German sentence is: 'Um aber diese Fälle Gleich zu machen – denn *dazu* reizt es uns – dient die Verwechslung zwischen Projektionsstrahlen, die das Bild mit dem Gegenstand verbinden, und der Projektionsmethode' (PG, 213). Kenny translates this as: 'What the comparison between the method of projection and the projection lines connecting the picture with the object does is to make all these cases alike—because *that* is what attracts us' (ibid). The problem is that he translates 'Verwechslung' which means 'confusion' with 'comparison', which would in German be 'Vergleich'. An amended translation should read: 'What the *confusion* of the method of projection with the projection lines connecting the picture with the object does…'

projection. One may, of course, accompany every picture of an object or state of affairs with such an illustration of its method of projection or with some other kind of description of this method in the imagination or in some publicly accessible medium of representation, such as a linguistic description or a drawing. And one may also claim that this illustration or description belongs to the picture. But this does not mean that the picture then 'reaches right out' to reality and determines what it depicts 'give or take a yes or no'. For a picture plus a description of its method of projection or a picture plus a depiction of this method – for example in form of a further picture depicting the first picture and the object it depicts connected by projection lines – is just a more complex representation or picture and no more closely related to what it depicts than the picture without the representation of its method of projection. In the *Investigations* Wittgenstein puts this as follows:

> [What] if not just the picture of [a] cube, but also the method of projection, comes before our mind? – How am I to imagine this? – Perhaps I see before me a schema showing the method of projection: say, a picture of two cubes connected by lines of projection. – But does this really get me any further? Can't I now imagine different applications of this schema too? (PI §141)

Obviously one *can* imagine different applications of the schema consisting of two representations of a cube connected by lines. One does not even have to understand this schema as illustrating the method of projection for a depiction of a cube. One might, for example, simply take it to represent a single, more complex object and not two cubes connected by lines. As Wittgenstein puts it: '[The] picture *plus* the projection lines leaves open various methods of application' (PG, 213). Hence, it is an illusion to think that a picture plus lines of projection 'reaches right out to reality' and determines what it depicts 'give or take a yes or no'. Ultimately, it is the same kind of illusion as the one that is at the centre of the rule-following discussion; namely the idea that when a picture, a proposition or a word is accompanied by an interpretation or act of meaning all its possible applications are 'in some *unique* way predetermined, anticipated' (PI §188).

The ‘method of projection’ on the other hand may be said to connect a picture with what it depicts or, more generally, a representation with what it represents. After all, in the scenario of the workman who constructs an artefact from a blueprint Wittgenstein characterizes it as ‘the way in which the workman turns… a drawing into an artefact’ (PG, 213). Similarly we might think of the method of projection as the way someone turns an order into the action that is its execution or the way someone draws a picture from a description. However, understood this way, the method of projection does not belong to the picture or representation. It consists in what we do with the picture, in how we translate it into actions. These actions are not part of the representation, for they do not precede its use in a concrete situation. As Wittgenstein puts it: ‘[If] the method of projection is a bridge, it is a bridge which isn’t built until the application is made’ (ibid; cf. also PG, 160).

Once one has clearly distinguished between the ‘method of projection’ and the ‘lines of projection’ the idea that a representation, whether it is pictorial or linguistic, ‘includes the pictorial relationship’ becomes unappealing. Whatever a representation includes is itself a representation. Hence it cannot include its own relation to the object or state of affairs it represents. As Wittgenstein rightly points out in *Philosophical Grammar,* the ‘pictorial relationship’ of the *Tractatus* is really a mongrel between the method of projection, which may, in a sense, connect a representation with what it represents, but which does not belong to the representation, and the lines of projection, which are themselves means of representation and which one may therefore count as belonging to a representation, but which do not connect a representation with what it represents. Once Wittgenstein recognized that his conception of the pictorial relationship in the *Tractatus* rested on a confusion of the ‘method of projection’ with the ‘projection lines’ he gave up the idea that there is a ‘pictorial relationship’ which is part of our representations – as I have argued in the first chapter, an internal relational property of all our representations –

and which connects them with what they represent: 'You may say: I count the projection lines as part of the picture—but not the method of projection' (ibid).

Abandoning the Tractarian conception of the pictorial relationship between picture and pictur*ed*, in particular between proposition and state of affairs, means abandoning an important element in the transcendental considerations supporting the metaphysics of the *Tractatus*. Whereas the distinction between a name and its bearer in PI §40 undermines the claim made in TLP 3.203 that the meaning of a name is the object it stands for, the distinction between the 'method of projection' and the 'projection lines' in *Philosophical Grammar* undercuts the idea that there has to be an object corresponding to every name occurring in a proposition with a sense. If a proposition does not 'include the pictorial relationship' connecting the names of the fully analysed proposition with objects in states of affairs, then there do not have to be such objects serving as relata of the pictorial relationship. Obviously the Tractarian conception of the pictorial relationship between propositions and states of affairs is not the only reason for the particular metaphysical assumptions of the *Tractatus*. For example, it does not follow from this conception that the postulated objects have to be simple and indestructible, nor does it follow that they have to share the 'logical form' of the names that stand for them. Still, once the idea of a pictorial relationship as an internal relational property of every proposition with a sense is abandoned, the assumption that there are objects corresponding to every name occurring in the fully analysed form of such a proposition becomes much less tempting.

Furthermore, rejecting the idea that there is an object corresponding to every name fits much better with our use of 'name' in ordinary language. Proper names of ordinary language can make their characteristic contribution to meaningful propositions although there might not be anyone or anything bearing that name. Names of fictional characters

such as 'Odysseus' or 'Sherlock Holmes', for example, occur in meaningful sentences although they do not have a bearer. Furthermore, names that have a referent do not cease to fulfil their function if their referent ceases to exist.[6] As Wittgenstein puts it in PI §40:

> When Mr N.N. dies, one says that the bearer of the name dies, not that the meaning dies. And it would be nonsensical to say this, for if the name ceased to have meaning, it would make no sense to say "Mr N.N. is dead".

This last observation, i.e. the observation that ordinary names continue to have meaning when their bearer ceases to exist, is, of course, one of the main motivations for the assumption of Russell and the early Wittgenstein that ordinary names are not really names and disappear in fully analysed propositions.[7] However, once the Tractarian conception of the pictorial relationship and the idea that the meaning of a name is its bearer are rejected, there is not much less reason left to make this assumption.

Finally, even when a name that we thought had a referent turns out to have none, the sentences in which it occurs do not cease to have a meaning, nor is the sentence asserting the non-existence of the referent in question nonsensical. When we say 'Moses did not exist' we may thereby mean a variety of different things, for example that 'the Israelites did not have a single leader when they came out of Egypt' or that 'there wasn't anyone who accomplished all that the bible relates of Moses' (cf. PI §79), but we are not making a nonsensical statement; for we are not denying the existence of an entity whose existence is a necessary precondition of the meaning of the subject expression.

What is true of names is equally true of other expressions. It is quite generally not a precondition of the sense of a linguistic expression that there is some object or property 'in the world' which corresponds to that expression. Definite descriptions can be meaningful

[6] These two points are perfectly compatible with the, quite possibly sound, idea that the *ability* to use names for fictional entities as well as the *ability* to name objects that have ceased to exist are dependent on the *ability* to use names for objects with which one is in perceptual contact.
[7] In PI §39 Wittgenstein brings out this motivation very succinctly.

although they fail to refer and predicative phrases or expressions can be meaningful although nothing exemplifies the property they stand for. Thus, we can speak about 'The present king of France' although there is no such person, and we can speak about something's being a winged horse although there is no horse with wings.

It might seem that despite the difficulties highlighted so far there are at least some expressions which are related to elements in the extra-linguistic order as to their meanings. These are expressions which can be defined ostensively. One might think that when we define an expression such as 'red' ostensively by pointing at a red piece of cloth and saying 'This is red' we are connecting an item in the linguistic order, namely the word 'red', with an element of reality which is its meaning, namely the colour red as it is exemplified in the piece of cloth we are pointing at.

This idea might be reinforced by the idea that while definitions of words in terms of other words, such as 'A bachelor is an unmarried man' or 'An unmarried man is called 'bachelor'', only succeed in giving the *definiendum* a meaning if the words making up the *definiens* already have a meaning, ostensive definitions can give words a meaning without relying on the meaning of other words. They thereby play a particularly fundamental role. They, as it were, infuse language with meaning in the first place, whereas other types of definition merely serve to associate a meaning already available to the person to whom the definition is given with a word not yet in her vocabulary. At the beginning of the *Blue Book* Wittgenstein presents this idea by distinguishing ostensive definitions from 'verbal definitions', i.e. definitions of words in terms of other words:

> The verbal definition, as it takes us from one verbal expression to another, in a sense gets us no further. In the ostensive definition however we seem to make a much more real step towards learning the meaning. (BB, 1)

The view expressed here is prominent in Russell's work and can also be found in the *Tractatus*. According to the *Tractatus* every sign that is not a simple name, i.e. every sign that can be defined by other signs, acquires its meaning through the signs defining it. Defined signs stand in the pictorial relation to what they mean through the mediation of the signs that define them. As Wittgenstein puts it in the *Tractatus*: 'Every sign that has a definition signifies *via* the signs that serve to define it; and the definitions point the way' (TLP 3.261). These chains of definition terminate in a definiens that only contains simple names, which are indefinable: 'Names *cannot* be anatomized by means of definitions' (ibid). Albeit indefinable, the meaning of names can be explained by what Wittgenstein calls 'elucidations':

> The meaning of primitive signs can be explained by means of elucidations. Elucidations are propositions that contain the primitive signs. So they can only be understood if the meanings of those signs are already known. (TLP 3.263)

What Wittgenstein appears to have in mind in this explanation of 'elucidation' is something closely akin to an ostensive definition. It makes much sense to think of these elucidations as statements of the form 'This is A', which, just like ostensive definitions, are made in the presence of the object signified by 'A'.[8] For one might think of these statement as true descriptions of an object which at the same time serve to point out the object that is the meaning of 'A', i.e. the meaning of the primitive sign that is being explained. If one is given such an elucidation one has, in a sense, to already know the meaning of the primitive sign since it is the object referred to by the indexical 'this' and since one has to somehow recognise that the speaker is trying to say something true; but one can also, in a sense, learn the meaning of the primitive sign with the help of these elucidations. If one knows that the speaker is trying to say something true and one knows what object she is referring

[8] For this point see Hacker 2001, 176.

to by means of the indexical ‘this’, then one recognises the object as the meaning of the primitive sign.

The later Wittgenstein takes the idea of ostensive definitions as connecting signs with elements of reality that are their meaning to rest on a misunderstanding of the role of such sentences as ‘This is red’ when they are used to define the word ‘red’. They rest, in particular, on a misconstrual of the role played by the item to which the ostensive reference is made. When we point to a red piece of cloth and say ‘This is red’ or ‘This is called ‘red’’ we are not connecting an item in the linguistic order, i.e. the word ‘red’, with an item in the real order that is its meaning, i.e. the colour of the piece of cloth pointed at; we are rather using the piece of cloth as a ‘means of representation’, namely as a *sample* of red (cf. PI §50). The role of such a sample is significantly different from the role of the object referred to by the subject term in declarative sentences such as ‘The car is backfiring’ or ‘The cat is on the mat’. Wittgenstein argues, it is rather like the role of ‘‘the’’ in ‘Pronounce the word ‘the’’ (cf. PI §16). Here we will count the ‘‘the’’ as part of language rather than as an item within the non-linguistic order. He suggests that we should do the same with the items referred to in ostensive definitions, such as the red piece of cloth in our example: ‘It is most natural, and causes least confusion, if we count the samples as tools of language’ (ibid).

Counting samples as ‘tools of the language’ or as ‘means or representation’ (cf. PI §50) causes ‘least confusion’ because it highlights the essential similarity between ostensive definitions and verbal definitions. Just like verbal definitions, ostensive definitions provide a rule for the use of a word. When we say ‘A bachelor is an unmarried man’ we are not attributing the property of being unmarried to one or all bachelors, we are rather stating the rule that whoever we may call ‘unmarried man’ we may also call ‘bachelor’. Similarly, the ostensive definition of red by pointing to a red piece of cloth and

saying ‘This is red’ does not attribute the property of redness to the piece of cloth pointed at (cf. BB, 2) but states the rule that we may call ‘red’ whatever has the colour of the sample. Basically both verbal and ostensive definitions are nothing more than substitution rules (cf. PG, 202). The verbal definition ‘A bachelor is an unmarried man’ tells us that we may substitute ‘bachelor’ for ‘unmarried man’ and the ostensive definition ‘This → is red’[9] or ‘This → colour is red’ tells us that we may substitute ‘red’ for the phrase ‘this colour’ accompanied by a deictic gesture at the sample.

Ultimately the idea that ostensive definitions provide a link between the linguistic and the natural order is informed by the same kind of confusion as the Tractarian idea of a pictorial relationship and cannot be used in its support. It is the result of not properly distinguishing between a representation of the use of an expression and its use; just as, according to *Philosophical Grammar*, the Tractarian conception of the pictorial relationship between a picture and the state of affairs it depicts rests on confusing a representation of how a picture is to be projected with the projection of a picture, i.e. on a confusion of the ‘lines of projection’ with the ‘method of projection’. As I have said, Wittgenstein believes that there is a sense in which one may think of the ‘lines of projection’, which we might, for example, conceive as imagined lines connecting the picture with what it depicts, as belonging to the picture. But to that extent they do not connect the picture with the pictured but merely provide a representation of how to project the picture. They are a ‘means of representation’. A connection between the picture and reality can only be found in what someone does with the picture; for example, in the actions of the worker, whom Wittgenstein imagines to build an artefact from a blueprint. These actions embody a ‘method of projecting’ the blueprint onto reality but they are not part of the picture, nor is the ‘method of projection’ they manifest.

[9] I use the arrow ‘→’ as a representation of a deictic gesture.

Similarly, there is a sense in which sentences such as 'This → is red' *give* us the meaning of the word that is being defined by relating it to the item pointed at. But when so used these sentences do not relate the word 'red' to its meaning, but rather serve as a rule for its use, i.e. as a representation of how 'red' is to be used. If, on the other hand, we use 'This → is red' to attribute the property of redness to the object ostensively referred to, we are speaking about an object outside language but we are not providing a definition of 'red'. Hence, one might either use 'This → is red' as an ostensive definition in which case the sentence is not used to relate an item in the linguistic order to an item in the non-linguistic order that is its meaning, or one might use it to speak about an item in the non-linguistic order in which case it ceases to be an ostensive definition. Thinking of ostensive definitions as relating an item in the linguistic order to an object that is its meaning is the result of confusing the use of sentences such as 'This → is red' as ostensive definitions with the use of the same type of sentence to attribute a property to the object ostensively referred to; just as thinking of the pictorial relationship as an internal relational property of every picture is the result of confusing the projection of a picture on a specific occasion of its use with the representation of the way a picture is, can or should be projected.

At the beginning of the *Blue Book* Wittgenstein points out the danger of confusing the use of sentences such as 'This → is red' as ostensive definitions with their use as a declarative sentence attributing a property to an object of ostensive reference. In the course of discussing the ostensive definition of the word 'tove' by pointing to a pencil and saying 'This is tove' he states:

> Instead of "this is tove" I could here have said "this is called 'tove'". I point this out to remove, once and for all, the idea that the words of the ostensive definition predicate something of the defined; the confusion between the sentence "this is red", attributing the colour red to something, and the ostensive definition "this is called 'red'" (BB, 2)

His reason for this remark may be that he was himself subject to this confusion when writing the *Tractatus*. In conversation with members of the Vienna Circle he is recorded as saying that at the time of the *Tractatus* the nature of ostensive definitions was unclear to him and that '[at] that time I thought that there was 'a connexion between language and reality'' (WVC, 209-10). The 'elucidations' of TLP 3.263 seem indeed to play the double role of attributing a property to an object of ostensive reference and of ostensively defining the meaning of a word. It is this mongrel role that elucidations are supposed to play that explains the strangely circular character Wittgenstein attributes to them, when he says that they, on the one hand, explain the meaning of primitive terms and, on the other, presuppose knowledge of this meaning.

Once the confusions at the root of the Tractarian conception of the 'pictorial relationship' and the correlated notion of 'elucidation' are uncovered, all that is left of the idea of sempiternal simple objects, whose existence is a necessary condition of sense, is the idea that there are some expressions, whose meaning presupposes the existence of samples by means of which it can be explained:

> A paradigm that is used in conjunction with a name in a language game – that would be an example of something which corresponds to a name and without which it would have no meaning. (PI §55)

Such a 'paradigm' or 'sample', however, shares few of the features characteristic of the simple objects of the *Tractatus*. It clearly does not exist necessarily, nor is there any clear sense in which it can be said to share the 'logical form' of the expression it is used to define. It does not make much sense to say of a red piece of cloth used to ostensively define 'red' that it has a 'logical form'; and it is wrong to say of any particular piece of cloth that it exists necessarily. The role of a sample of red can be played by all sorts of objects. So even if it is true that there has to be an object that can serve as a sample of red

if ‘red’ is to retain its meaning, this does not imply that there is a particular object that has to exist in order for ‘red’ to keep its meaning. All that Wittgenstein is saying when he is saying that a paradigm ‘would be an example of something which corresponds to a name and without which it would have no meaning’ is that it is a typical feature of some types of expression that their meaning can be explained by reference to a sample.

The most important difference between simples and samples, however, is that samples are ‘means of representation’ or ‘tools of language’. It is this difference that leads Wittgenstein to conclude:

> What looks as if it *had* to exist is part of the language. It is a paradigm in our game; something with which comparisons are made. (PI §50)

Since the samples or ‘paradigms’ figuring in ostensive definitions are ‘part of the language’, language is not beholden to a particular metaphysical structure of the world; it is rather, as Wittgenstein puts it, ‘autonomous’. What looked in the *Tractatus* like a relationship between language and reality, i.e. the pictorial relationship between simple names and simple objects, is rather a relation in language, namely a relation between samples and words:

> The connection between ‘language and reality’ is made by definitions of words, and these belong to grammar, so that language remains self-contained and autonomous. (PG, 97)

2. A non-relational account of intentionality

At the beginning of this chapter I presented Wilfrid Sellars’s claim that it is a virtue of the picture theory of the *Tractatus* that it steers between the Scylla of a position that holds that intentional phenomena acquire their significance by being related to contents, which have, as Sellars puts it in allusion to Descartes, merely ‘being-for-mind’ or ‘objective being’ (BBK, 41) and are only externally related to objects and events in the world, and the

Charybdis of a view that sees intentional phenomena as directly related to items in the real order, but at the cost of having to postulate subsistent states of affairs and complex objects in order to account for the possibility of false thoughts and empty names. I associated the former view with Russell's *The Analysis of Mind* and the latter view with the American 'new realists'.

According to Sellars the core problem of these two views is the shared assumption that 'intellectual acts differ *not* in their intrinsic character as acts, but by virtue of being related to different relata' (BBK, 41). The *Tractatus* on the other hand, Sellars suggests, belongs to a tradition which holds that 'acts of the intellect differ intrinsically *qua* acts in a way which systematically corresponds to what they are about' (ibid, 43).

There is more than a grain of truth in this way of situating the Tractarian account of intentionality within a broader field of possible views. For, first, it was clearly a central concern of Wittgenstein's in the *Tractatus* to avoid both the problem encountered by the kind of view for which I took Russell's *The Analysis of Mind* to be representative, namely an inability to explain how the contents of intentional phenomena are related to the items in the world they are supposed to be about, and the problem faced by the new realists, which consisted in an inability to give a plausible account of false thoughts. Accordingly, the picture theory was supposed to explain both how language can 'reach right out' to reality (TLP 2.1511) and to provide a plausible answer to the question of how propositions and thoughts can be bipolar. Second, the early Wittgenstein clearly thought that the right way to answer these challenges was by supposing that, as Sellars puts it, 'acts of the intellect differ intrinsically *qua* acts in a way which systematically corresponds to what they are about', namely by assuming that all meaningful propositions share a 'logical form' with the states of affairs they are said to be 'about'.

However, my discussion of the pictorial relationship should have made clear that, contrary to what Sellars suggests, the Wittgenstein of the *Tractatus* did not completely reject the view that 'intellectual acts differ *not* in their intrinsic character as acts, but by virtue of being related to different relata' (BBK, 41). He certainly thought that different propositions and thoughts differ intrinsically, for they must be made up of different simple constituents or their simple constituents must at least be arranged differently. But he also thought that being related to different relata in the world was an internal property of every thought and proposition. For, although he thought that propositions or thoughts do not stand for states of affairs in the same way as names stand for objects, he also thought that every name in a fully analysed proposition or thought stood in an internal relation to an object that is its meaning. He thus shared, at least in a qualified sense, the presupposition of Russell and the 'new realists' that different thoughts or propositions differed in virtue of being related to different relata which are their content. He only fully abandoned this assumption when he abandoned the picture theory and the idea that thoughts and propositions stand in an internal pictorial relation to states of affairs in the world that are their sense.

I therefore think the later Wittgenstein's rejection of the idea of a pictorial relationship between propositions and thoughts, on the one hand, and states of affairs or facts, on the other, should be seen as a rejection of the whole idea that propositions, thoughts and other intentional acts or states stand in a *relation* to their contents. In particular, Wittgenstein's move from the picture theory of meaning, according to which the sense of a proposition is the state of affairs onto which it is projected and the meaning of a name is the object it stands for, to the later conception of meaning as use, should be seen as a move from what Sellars calls a 'relational' account of meaning to a 'non-relational' account (cf. e.g., SM, ix). I think it is corollary of the idea that the meaning of linguistic expressions is wholly

determined by their use in 'language games' that meanings are not kinds of things to which expressions are related. As Wittgenstein points out in the *Blue Book* the use of a sign is not an 'object *co-existing* with the sign' (BB, 5, italics in the original). It is this insight, I take it, that Sellars tries to capture when he calls the kind of view of meaning as use, which he shares with Wittgenstein, 'non-relational'.

Wittgenstein's contention that ostensive definitions remain within language as well as his more general claim that the 'connection between 'language and reality' is made by definitions of words… so that language remains self-contained and autonomous' (PG, 97) are a consequence of the non-relational character of his account of meaning. If the meaning of a linguistic expression is fully determined by its rule-governed role in language games, then the purpose of statements specifying or explaining this meaning cannot be to connect the expression with an abstract or concrete item, which is its meaning, but to specify its use or role in language games. Since we typically think of these meaning-specifying or semantic statements, such as 'This → is red', ''Rot' means red' or ''Red' stands for redness', as establishing semantic relations between language and world, this implies that there is a sense in which language indeed 'remains self-contained and autonomous'.

Sellars has developed his own 'non-relational' account of meaning by focusing on this aspect of semantic statements. I think a discussion of Sellars's views will help to substantiate and illuminate the kind of view of linguistic meaning Wittgenstein suggests should replace the picture theory.

2.1 Sellars on meaning

Like Wittgenstein, Sellars thinks that the meaning of sentential as well as sub-sentential expressions is fully determined by the role they play in 'language games'. According to Sellars, the meaning of an expression is determined by the role it plays within a rule-governed linguistic practice or, as he usually likes to put it, its role within 'pattern-

governed behaviour'.[10] Unlike Wittgenstein, however, who repeatedly stresses the diversity of different language games (cf. e.g. PI §§12, 23, 24 and 27), Sellars tends to focus on what might be called the 'cognitive role' of language rather than its role in human interaction.[11] Accordingly he usually focuses on three kinds of linguistic performances corresponding to the categories of perception, inference and action. The *first* kind of performances or 'moves in the language game' are 'language entry transitions', that is linguistic responses to non-linguistic stimuli; the *second* are 'intra-linguistic moves, i.e. transitions from one utterance to another, and the third are 'language departure transitions', utterances which are followed by non-linguistic performances such as drawing a gun (cf. MFC, 87-88). All sentential and sub-sentential expressions with the same norm-governed role within the network of language entry transitions, intra-linguistic moves and language departure transitions within their respective languages have the same meaning.

Sellars brings out the 'non-relational' character of his account of linguistic meaning by suggesting analyses of semantic statements, such as '…means---' and '… stands for---', according to which they do not *relate* an expression to some abstract or concrete entity, property or state of affairs that is its meaning but rather provide a classification – as Sellars

[10] For Sellars's notion of 'pattern-governed behaviour' see SRLG, especially pp. 324-27, and MFC, especially pp. 86-87. The somewhat awkward phrase 'pattern-governed behaviour' has for Sellars a technical sense and contrasts with what he calls 'rule obeying behaviour'. Roughly speaking, pattern-governed behaviour is a kind of behaviour which is governed by rules, although the subject displaying it does not consciously entertain these rules. Rule obeying behaviour, on the other hand, is guided by rules consciously entertained by the subject displaying it. In what follows I will ignore the difference between pattern-governed and rule obeying behaviour and will talk interchangeably about he role of linguistic expressions within pattern-governed behaviour, language games and linguistic practices.

[11] Sellars highlights his primary concern with the cognitive role of language use at the beginning of his 'Meaning as Functional classification' (cf. MFC, 81-85). In lectures delivered at the University of Notre Dame and published on the internet he puts the difference between his and Wittgenstein's conception of meaning as follows: '[Wittgenstein] uses that [i.e. the claim that the meaning of an expressions is its use, S. B.] as a kind of slogan and I would be prepared to use it as a kind of slogan – the fundamental difference is that in the *Investigations* for Wittgenstein, when Wittgenstein's [*sic*] speaks of the meaning of the expression as its use, he has in mind a whole range of uses of kinds that are quite unlike like [*sic*] those I concentrate on or stress' (WSNDL, 123; cf. also AAE, 197-98). Although Sellars took this difference between his and Wittgenstein's philosophy of language to be important, it is irrelevant in the present context. I merely mention it in order to avoid the impression that Sellars's somewhat skeletal account of linguistic practices is the result of an oversight.

likes to put it a 'functional classification'[12] – of the role the expression plays within the linguistic practices of the speakers of the language to which it belongs.

What does such a functional classification look like? I will begin, just as Sellars usually does, by concentrating on the analysis of what he calls 'meaning statements' (cf. MFC passim.), i.e. statements of the form '… means---'. Sellars suggests that we should interpret the term occurring in the object position of these statements as *exhibiting* an expression which plays in our language a normative-functional role that is equivalent to the role that the expression that is being classified, i.e. the expression occurring in the subject position, plays in its language. This, however, does not mean that the correct analysis of, for example,

(a) 'Rot' means *red*,

is:

(a_1) 'Rot' plays in German the same normative-functional role as 'red' plays in English.[13]

[12] Sellars's use of the word 'functional' may raise doubts as to whether his conception of language is really compatible with Wittgenstein's. In contemporary philosophy of mind the word 'functionalism' is usually used to refer to a view which is quite alien to Wittgenstein. According to this view the content of a mental state is fully determined by its *causal role* within the behaviour of a person. Content bearing mental phenomena are individuated in terms of their role within a causally but not normatively defined structure. When extended to language, this form of functionalism appears to be incompatible with Wittgenstein's insistence on the essentially normative or rule-governed character of linguistic behaviour. Although, as I will show in the next chapter, Sellars indeed endorses a form of functionalism in the philosophy of mind, his talk about the 'functional role' of linguistic expressions should not be understood in the purely causal sense characteristic of contemporary functionalism, for he repeatedly stresses the essentially normative or rule governed character of our linguistic practices. When Sellars speaks of the functional role of a linguistic expression he speaks about the role of that expression within a linguistic practice that can *only* be fully understood in terms of the norms or rules that govern it. In order to make this difference to contemporary functionalism explicit I will occasionally speak about the 'normative-functional role' of an expression.

[13] See MFC, 95; cf. also deVries 2005, 29.

The problem with (a_1) is that it does not really *give* us the meaning of the German word 'rot', it merely points out an analogy between 'rot' and the English expression 'red'. If (a_1) was the correct analysis of (a), (a) would translate into French as ''Rot' (en Allemand) veut dire 'red''. The correct translation, however, seems to be ''Rot' (en Allemand) veut dire *rouge*'. Sellars, therefore, gives an analysis of (a) which tries to reconcile the fact that the grammatical object of (a) is, contrary to its grammatical subject, not a straightforwardly mentioned expression with the idea that it is still, in a sense, meta-linguistic.

He suggests interpreting 'means' as a special form of the copula, which – taken together with the object expression of a meaning statement (in the above case 'red') – forms a one-place predicate, which functionally classifies the expression occurring in the subject position (in the above case 'rot'). He interprets the object term as an *inter-* and *meta-linguistic illustrating sortal* which applies to any expression in any language whose functional role in its language is closely akin to the role of the expression occurring in the object position within the language of the meaning statement. The grammatical object is a meta-linguistic expression because it applies to linguistic expressions. It is a sortal because linguistic expressions are countable. It is inter-linguistic because it applies not merely to expressions of some specific language, for example German or English, but to any expression in any language that has in its language the same role as the object expression within the language of the meaning statement. In this respect the object term '*red*' in (a) is similar in character to the sortal 'lieutenant', which applies to any soldier in any army who has the rank of a lieutenant. Unlike 'lieutenant', however, the grammatical object of a meaning statement is an *illustrating* sortal, since it consists of an expression that, when it is not used as a meta-language expression, has the same role as the expressions it is used to classify.[14] Thus, the object expression not merely classifies the quoted expression in the

[14] Obviously, the expression 'lieutenant', being a word rather than a human being, cannot possibly 'play the role' of a lieutenant.

subject position as playing a particular normative-functional role, it also illustrates this role to everyone, who understands the meaning statement, by being itself an instance of an expression type, which plays the same role when not used meta-linguistically. In order to indicate this special role of the object expression, Sellars introduces a new form of quotes, *dot-quotes,* which indicate that the expression they flank has the role of such an inter-linguistic sortal (cf. MFC, 95). Dot-quotes characterize a quoted expression only with respect to its normative-functional role and abstract completely from the specific 'sign design' of the quoted expression.

The subject term of a meaning statement is interpreted as a meta-linguistic *distributive singular term* (DST, henceforth) (cf. MFC, 94-95). Sellars's example for a DST is the term 'the lion' as it functions in sentences such as,

(b) The lion is tawny.

Here 'the lion' does not refer to a particular lion but to lions in general. In MFC Sellars suggests that when 'the lion' functions as a DST, (b) is '*roughly* equivalent in meaning to 'All lions are tawny'' (MFC, 94). He gives his reason for the qualification *'roughly'* in his paper 'Abtstract Entities' (AE henceforth), where he claims that universally quantified statements of the form 'All Ks are *f*', which are the analysis of statements of the form 'The K is *f*' when 'The K' functions as a DST, express 'non-accidental truths' about Ks (cf. AE, 167-8). According to Sellars, statements of the form 'The K is *f*' or 'All Ks are *f*' express a 'non-accidental' truth about Ks when '*being f* is either one of the criteria for *being a K* or is implied by the latter on inductive grounds' (AE, 167, italics in the original).

I think the best way to interpret this remark is as claiming that statements of the form 'The K is *f*', in which 'The K' functions as a DST, are equivalent to statements of the form

'Necessarily all Ks are *f*', where the necessity in question is either conceptual, namely when *being f* is a criterion for *being a K,* or it is the necessity of a natural law, namely when 'The K is *f*' or 'All Ks are *f*' has been accepted on inductive grounds.[15] Since (b) is the kind of statement which one might come to accept on inductive grounds and since Sellars usually claims that inductive arguments lead to the acceptance of laws of nature, it presumably expresses for Sellars a law of nature.[16] The importance of the difference between interpreting 'The K is *f*' as 'Necessarily all Ks are *f*' and interpreting it as 'All Ks are *f*', where the latter statement is understood as merely expressing an 'accidental truth' about Ks, in the context of a discussion of meaning statements is that (i) statements of the form 'The K is *f*' are not automatically true when there are no Ks and (ii) meaning statements, if true, appear to be necessarily true.[17]

---

[15] More explicitly, for Sellars 'All Ks are *f*' expresses a necessary truth, if we are entitled to infer directly, i.e. without an additional premise, 'X is f' from 'X is a K'; either because the criteria for the application '…is f' are among the criteria for the application of '…is a K' or because we can give an inductive argument to the effect that one may directly infer 'X is f' from 'X is a K'. (Roughly speaking, inductive arguments are for Sellars practical arguments to the effect that one is entitled to draw certain inferences, cf. his 'Induction as Vindication' (IV)). I have suggested above that in the first case it is a conceptual necessity that all Ks are f whereas in the latter case it is a natural necessity or, in other words, a law of nature. It is important to note, however, that within Sellars's system it is potentially misleading to distinguish between conceptual and natural necessities, as according to him natural necessities are a variety of conceptual necessities. For Sellars *all* statements of the form 'Necessarily all Ks are F' are true if and only if the rules of our language allow us to directly infer 'X is f' from 'X is a K', or, as Sellars likes to put it, if and only if the inference from 'X is a K' to 'X is f' is 'materially valid' (cf. IM passim.). Since all material inferences play a role in determining the meaning of their relata even causal necessities are due to the meaning of the terms in which they are expressed, i.e. are a variety of conceptual necessities. Still there is clearly a difference between, on the one hand, inferences from 'X is a K' to 'X is f' which are licensed by the fact that being f is among the criteria for being a K and, on the other hand, inferences of the same form which are licensed by an inductive argument. Since the former kind of inferences seem to roughly coincide with what is normally called a conceptually valid inference and since Sellars usually claims that inductive arguments lead to the acceptance of laws of nature, I have chosen to say that the universally quantified statement expressing the acceptance of the inferences in question says in the former case that it is conceptually necessary that all Ks are f and in the latter case that it is naturally necessary that all Ks are f.

[16] This is, of course, inconsistent with using 'law of nature' as equivalent with 'causal law', which Sellars does quite frequently. However, I do not think that this counts against my interpretation as I think it is impossible to find any consistency in Sellars's use of 'law of nature' and equivalent expressions.

[17] If my interpretation is correct, it is puzzling that Sellars does not say in MFC that (b) is equivalent to 'Necessarily all lions are tawny' but rather claims that it is '*roughly* equivalent in meaning to 'All lions are tawny' (MFC, 94). I believe his reason for this is that he presupposed his views on necessity when writing MFC but did not wish to discuss them. Sellars's writings quite frequently contain these sorts of unexplained presuppositions of results he takes himself to have established elsewhere; almost always with the result of unnecessarily obscuring his views. Furthermore, one might wonder why Sellars does not simply suggest that (b) should be interpreted as 'Lions are tawny' or 'Lions in general are tawny' which seem to be much more adequate natural language paraphrases of (b) and its ultimate analysis 'Necessarily all lions are tawny' than 'All lions are tawny'. To this second question Sellars does provide an answer in a footnote. He suggests that

Unlike ‘the lion’, DSTs occurring as grammatical subjects of meaning statements are meta-linguistic. In other words, they refer to linguistic rather than non-linguistic objects, namely to all possible as well as actual correct uses of the expressions between the quotes within the language to which they belong.

Let us return to the above meaning statement:

(a) ‘Rot’ means *red*.

This statement expresses a functional classification of the German word ‘rot’ in English as meta-language. ‘‘Rot’’ is a DST which refers to all correct uses of ‘rot’ in the German language, while ‘*red*’ is an illustrating sortal which refers to a functional linguistic kind by *exhibiting* one instance of this kind. ‘Means’ is a special form of the copula which combines with a meta-linguistic sortal to form a one-place meta-linguistic predicate. Statement (a) says that all correct uses of ‘rot’ in German belong to the normative-functional kind referred to by ‘*red*’. Finally, since the subject expression is a DST (a) expresses a necessary truth; and since it is clearly not the kind of statement which is accepted on inductive grounds it must be, according to Sellars, true because being a ‘*red*’ is a ‘criterion’ for being a ‘Rot’. In other words (a) expresses a conceptual truth.

In its final analysis (a) is equivalent to the following statement (a_2), in which ‘means’ is replaced by the copula ‘are’, the status of the grammatical subject as a DST is made explicit by the use of plurals, and the object expression is flanked by dot-quotes indicating its specific meta- and inter-linguistic character:

---

there are occurrences of ‘the lion’ where this expression has roughly the same meaning as ‘lions’ but does not function in the same way as in (b), i.e. it does not function as a DST. His example of such an occurrence of ‘the lion’ is its use in ‘The lion once roamed the western plains’ (MFC, 94), which, though equivalent to ‘Lions once roamed the western plain’, is clearly not equivalent to ‘Lions in general once roamed the western plains’.

(a_2) ‘Rot’s are •red•s.[18]

A similar analysis is applied to meaning statements specifying the meaning of whole sentences, for example,

(c) ‘Der Ball ist rund’ means *the ball is round*.

According to Sellars (c) is to be analysed in exactly the same way as (a). Hence the final analysis of (c) reads as follows:

(c_1) ‘Der Ball ist rund’s are •the ball is round•s.

Statement (c) illustrates the importance of interpreting meaning statements as expressing a ‘non-accidental’ truth about the expression quoted in the subject position. If statements such as (c), which specify the meaning of whole sentences, simply said that all actual uses of a certain sentence type belonged to a certain normative functional kind, then they would be automatically true for all sentences that have never been used – at least according to a fairly widespread interpretation of universally quantified statements of this type. If we, for example, assume that there has never been a non-meta-linguistic use of the sentence ‘Didcot is north of Castrop-Rauxel’ then the sentence:

(d) ‘Didcot is north of Castrop-Rauxel’ means *Perth is south of Mumbay,*

[18] cf. MFC, 95.

would be true under the assumption that ‘‘Didcot is north of Castrop-Rauxel’’ in (d) only referred to actual uses of the quoted sentence. However, since (d) is obviously false and since there are infinitely many more sentences that have never been used but whose meaning we can specify with statements of the form ‘…means---’, the subject expression of a meaning statement cannot merely refer to actual uses of the expression within the quotes.

Sellars calls his account of meaning ‘non-relational’ because statements such as (a_2) and (c_1) and therefore the statements (a) and (c) of which they are the analysis do not relate the expressions whose meaning they specify to some abstract or concrete item that is their meaning, but rather classify them as playing a certain normative-functional role. This is made explicit by the replacement of ‘means’ by the copula ‘are’ in (a_2) and (c_1). As Sellars puts it: ‘According to this analysis, *meaning is not a relation* for the very simple reason that ‘means’ is a *specialized form of the copula*’ (MFC, 95, italics in the original). Put differently, meaning is not a relation because in their final analysis meaning statements such as (a) and (c) have the same logical form as such straightforwardly non-relational statements as ‘Weimeraners are dogs’ or ‘Bavarians are Germans’.[19]

Of course, Sellars’s analysis of meaning statements as classificatory rather than relational cannot be anything more than a first step towards a non-relational account of meaning. One might, for example, object that often, when we appear to specify the meaning of an expression, we do not classify the expression but relate it to an abstract object. The following three artificial statements can be seen as representing what we do when we do this kind of thing:

[19] For this point cf. deVries 2005, 30.

(e) ‘Dreieckig’ stands for triangularity.

(f) ‘Hund’ stands for the concept *dog*.

(g) ‘Der Ball ist rund’ stands for the proposition that the ball is round.

Statement (e) appears to relate the German predicative expression ‘dreieckig’ (‘triangular’ in English) to an abstract object that is its meaning, namely the universal or property *triangularity*, (f) seems to relate the expression ‘Hund’ to a different kind of abstract entity, namely the concept *dog* that is its sense, and finally (g) appears to relate the sentence ‘Der Ball ist rund’ to a third kind of abstract object, namely the proposition *that the ball is round*. Thus, ‘stands for’ and ‘says’, as they are used in (e), (f) and (g), seem to express a *relation* between a linguistic expression and an abstract entity. And since (e), (f) and (g) seem to fulfil the same function as meaning statements, such as (a) and (c), this appears to make Sellars’s claim that the latter are non-relational appear rather suspect.

Sellars counters this objection by suggesting that one should interpret ‘triangularity’, ‘the concept *dog*’ and ‘that the ball is round’ not as names of abstract objects but as inter-linguistic DSTs, which refer to all expressions in all languages which play in their language the role ‘triangular’, ‘dog’ and ‘the ball is round’ play in English. ‘Stands for’ and ‘says’ are again read as special forms of the copula and ‘‘dreieckig’’, ‘‘Hund’’ and ‘‘Der Ball ist rund’’ are again understood as DSTs referring to all correct uses of the expressions between the quotes. Thus, Sellars suggests to analyse (e), (f) and (g) as,

(e_1) The ‘dreieckig’ is the •triangular•,

(f_1) The ‘Hund’ is the •dog•,

and,

(g_1) The 'Der Ball ist rund' is the •the ball is round•.

Statements (e_1), (f_1) and (g_1) consist of the copula flanked on both sides by meta-linguistic DSTs. The meta-linguistic DST in the grammatical subject position refers to all correct uses of 'dreieckig', 'Hund' and 'Der Ball ist Rund' in the German language. The DST in the object position is, in addition to being meta-linguistic, also inter-linguistic and refers to all uses of expressions in all languages that play in their language the role 'triangular', 'dog' and 'the ball is round' play in English. The logical form of (e_1), (f_1) and (g_1) is the same as that of such sentences as,

(h) The pub is the poor man's club,[20]

which also consists of a copula flanked by two DSTs.

It seems quite natural to say that (h) is equivalent to:

(h_1) Pubs are poor men's clubs.

Hence, it seems equally natural to say that (e_1), (f_1) and (g_1) are equivalent to:

(e_2) 'Dreieckig's are •triangualar•s,

(f_2) 'Hund's are •dog•s

and,

[20] For this example see MFC, 96-97.

(g_2) 'Der Ball ist rund's are •the ball is round•s.

Thus, according to Sellars, the ultimate analysis of statements such as (e), (f) and (g) is the same as that of meaning statements such as (a) and (c). And this means that they too do not relate the expressions whose meaning they specify to an abstract object that *is* their meaning, but rather classify their functional role by means of an inter-linguistic illustrating sortal. For an expression to stand for the universal triangularity is to be a •triangular•, to stand for the concept *dog* is to be a •dog• and to stand for the proposition *that the ball is round* is to be a •the ball is round•.

Sellars suggests using the expressions 'sense', 'concept', 'universal' and 'proposition' so that all expressions stand for senses but only some stand for concepts, universals or propositions. An expression stands for a concept and a universal, rather than merely a sense, if it is a predicate; and it stands for a proposition if it is a sentential rather than a sub-sentential expression.[21] Furthermore, he claims that concepts and propositions are a variety of senses. Concepts are the senses of predicates and propositions are the senses of sentences. Finally, when one speaks about senses, concepts and propositions, one does not commit oneself to their being abstract entities that are senses, concepts or propositions. All one commits oneself to is there being expressions that stand for the senses, concepts and propositions one is speaking about. And to stand for a sense, concept or proposition is nothing more than to play a certain role.

[21] Sellars acknowledges that this way of distinguishing between expressions that stand for concepts and expressions that do not is legislative rather than descriptive of our actual use of the expression 'concept', since it implies that individual terms and logical connectives do not stand for concepts (cf. AE, 192). Thus, according to Sellars, it is wrong to say that 'not' stands for the concept of negation and that 'the silliest member of the Drones' Club' stands for the individual concept *the silliest member of the Drones' Club*. His reason for this regimentation is that he wants to bring his use of 'concept' in line with Frege's: 'Since, as I see it, Frege's distinction between concepts and objects was of decisive importance in the history of ontology, I am strongly inclined to follow his lead and limit the term "concept" to predicative senses' (AE, ibid.).

Finally, Sellars suggests that even statements specifying the reference or denotation of a linguistic expression do not relate the expression to an item in the world but rather to another expression. According to Sellars, even in sentences such as

(i) 'The Big Apple' denotes (or refers to) New York,

or

(j) 'Rational animal' denotes (or refers to) featherless bipeds,

the expressions on the right-hand side, i.e. 'New York' and 'featherless biped', do not play their ordinary object language role but function as meta-linguistic expressions. Sellars claims that the difference between, on the one hand,

(k) 'New York' means New York

and

(l) 'Federloser Zweifüßer' means featherless biped,

and (i) and (j), on the other, does not lie in a difference between the expressions on the right-hand side, i.e. 'New York' and 'featherless biped', which function as meta-linguistic expressions in both types of statement, but in a difference between 'means' and 'refers to' or 'denotes' (cf. SM, 85). Accordingly, Sellars suggests an analysis of reference specifying

statements such as (i) and (j) that makes explicit their meta-linguistic status while doing justice to their apparently extensional character, i.e. to the fact that unlike meaning statements they do not assert a semantic equivalence or sameness of meaning between the expression whose reference is specified and the expression used to specify this reference.[22] He introduces a variable 'S' (short for 'sense') which takes as its substituends common nouns formed by dot-quoting. Thus, •rational animal•, •New York• and •or• are all Ss. He then introduces the notion of material equivalence, which is expressed in statements of the form:

(m) S_1 is materially equivalent to S_2.

Examples of substitution instances of this schema are,

(n) •Rational animal• is materially equivalent to •featherless biped•,

and,

(o) •The Big Apple• is materially equivalent to •New York•.

Statement (n) is true if and only if the following biconditional is true:

(p) For all *x*: *x* is a rational animal iff *x* is a featherless biped.

[22] For the following discussion see SM, 83-87.

In other words (n) is true just in case all rational animals are featherless bipeds. Statement (o), on the other hand, is true if and only the following biconditional is true:

(q) For all *f*: The Big Apple is *f* iff New York is *f*.

Put differently, (o) is true just in case whatever is true of the Big Apple is also true of New York; or, according to Leibniz law, just in case the Big Apple is identical with New York. Generalizing from these examples we may say that material equivalence is a relation, which holds between predicative expressions if and only if they are true of the same objects and holds between referring expressions if and only if exactly the same predicative expressions can be combined with them to form true sentences.

With the help of these materials Sellars now analyses (i) as,

(i_1) For some S, 'The Big Apple' stands for S, and S is materially equivalent to •New York•.

Statement (i_1) is true because 'The Big Apple's are •The Big Apple•s and •The Big Apple•s are materially equivalent to •New Yorks•s. Statement (j) is analysed along the same lines:

(j_1) For some S, 'rational animal' stands for S, and S is materially equivalent to •featherless biped•.

This is true because 'rational animal's are •rational animal•s and •rational animal•s are materially equivalent to •featherless biped•s.

The quantification over abstract entities, namely senses, in (i_1) and (j_1) is only apparent, for – as I already pointed out – the substituends of Ss are 'common nouns formed by dot quoting' (SM, 84). Sellars is only committed to functionally classifiable expressions of languages.

2.2 Sentences and facts

It has to be admitted that Sellars's somewhat awkward technical analyses of statements of the form '…means---', '…stands for---' and '…refers to---' are not exactly in the spirit of Wittgenstein's later approach to philosophy. However, I think they lend substantial support to Wittgenstein's *prima facie* puzzling claims that he ceased to think that there was a 'connexion between language and reality' (WVC, 209) and that the 'connection between 'language and reality' is made by definitions of words… so that language remains self-contained and autonomous' (PG, 97). I think these remarks are not supposed to suggest that we do not use language in order to speak about phenomena outside of language, which would be absurd, but to convey the idea that, contrary to what Wittgenstein had assumed in the *Tractatus*, linguistic expressions are not related to items in the concrete world as to their meanings, nor are they so related to any other items outside language, such as concepts and propositions in Frege's 'third realm of ideas'. That is the sense in which language 'remains self-contained and autonomous'.

Sellars's analysis of semantic discourse supports this view by, first, extending the insight contained in Wittgenstein's idea that ostensive definitions remain within language to other semantic statements. Sellars's insight that the expressions on the right-hand side of statements of the form '…means---', '…stands for---' and '…refers to---' have a meta-linguistic role is of a piece with Wittgenstein insight that the items pointed at in ostensive explanations of words such as 'This → is red' figure as samples: Just as 'This →' in the ostensive explanation 'This → is red' is not used to refer to an item outside language but to

a ‘means of representation’ so ‘*red*’ in ‘‘Rot’ means *red*’ does not serve its usual role as a predicate applying to red things but is a meta-linguistic sortal applying to linguistic expressions, i.e. to ‘means of representation’.

Second, the later Wittgenstein repeatedly stresses the danger of postulating ‘shadowy’ entities (cf. BB, 36) in order to explain what false and negative sentences are about. In PI §448, for example, he says:

> “If I say I did *not* dream last night, still I must know where to look for a dream; that is, ‘I dreamt’, applied to this actual situation, may be false, but mustn’t be nonsense.” – Does that mean, then, that you did, after all, feel something, as it were the hint of a dream, which made you aware of the place which a dream would have occupied?
>
> Again, if I say “I have no pain in my arm”, does that mean that I have a shadow of a sensation of pain, which, as it were, indicates the place where a pain could have been?

In the *Blue Book* he raises a similar issue by speaking about the objects, in the sense of the contents (cf. BB, 38), of false thoughts as shadows of facts and suggests that we typically call these shadows ‘propositions’ or ‘senses’:

> [We] are inclined… to think that as the object of our thought isn’t [a] fact it is a shadow of [a] fact. There are different names for this shadow, e.g. “proposition”, “sense of the sentence”. (BB, 32)

And a few pages later Wittgenstein introduces a further motivation for introducing these shadows of facts:

> [We] say “Surely two sentences of different languages can have the same sense”; and we argue, “therefore the sense is not the same as the sentence”, and ask the question “What is the sense?” And we make of ‘it’ a shadowy being, one of the many which we create when we give meaning to substantives to which no material objects correspond. (BB, 36)

The states of affairs of the *Tractatus* are such ‘shadowy beings’. As we have seen in the first chapter, they are possible arrangements of necessarily existing simple elements of reality which only differ from facts in not being actual. In *Philosophical Grammar* Wittgenstein speaks about the temptation to think of an event that someone is expecting as something that already ‘stood in readiness before the door of reality and were then to make

its appearance in reality' (PG, 137), just like a man, whom we are expecting to come into the room, might already stand before the door. I take it that the early Wittgenstein succumbed to this kind of thinking when he thought of states of affairs as possible arrangements of actual objects, which share the 'logical form' of the sentences that speak about them. Their 'shadowy being' consists in the fact that, on the one hand, they do not obtain, but, on the other, they are possible arrangements of actually existing elements and are the sense of actually existing sentences or – in the Tractarian terminology – 'propositions'.

The virtue of Sellars's analysis of statements of the form '…stands for---' consists in the fact that he can make perfectly good sense of the idea of a sentence standing for something that is not the case and of two sentences in two different languages as standing for the same sense or proposition, without committing himself to there being any 'shadowy beings' these sentences stand for. If I say 'I did not dream last night' what I say does indeed stand for the proposition that I did not dream, but that merely means that my utterance is correctly classified by the inter-linguistic sortal •I did not dream last night•. In other words my utterance stands for or expresses the proposition that I did not dream last night because it is a •I did not dream last night•. Similarly it is true to say that the German sentence 'Ich habe heute nacht nicht geträumt' and the English sentence 'I did not dream last night' have the same sense or stand for the same proposition. However, this is not because there is a shadowy being, a sense or a proposition, to which both sentences stand in the relation of 'expressing' or 'standing for'. It is rather because they play the same normative-functional role within their languages. In Sellars's terminology they both are •I did not dream last night•s.

Similarly, Sellars's analysis of '…stands for---' allows him to interpret statements such as ''Triangular' stands for the universal triangularity' or ''Triangularity' stands for the

concept of triangularity (or the concept *triangular*)' in a way that does not commit him to the existence of 'shadowy' entities, such as concepts or universals, which exist independently of the linguistic expressions that stand for them. 'Triangular' stands for the concept of triangularity or the universal triangularity because it is a •triangular•; and 'dreieckig' and 'triangular' stand for the same concept or the same universal because they are both •triangular•s.

Sellars's analysis of the grammatical objects of statements of the form '…stands for---' as meta- and inter-linguistic distributive singular terms rests on a general nominalistic thesis about the interpretation of abstract singular terms. In his 'Meaning as Functional Classification' he sums this thesis up with the following words:

> In general, I suggest that so-called nominalizing devices which, when added to expressions, form corresponding abstract singular terms, thus '-ity', '-hood', '-ness', '-tion', 'that...', etc., are to be construed as quoting contexts which (a) form metalinguistic functional sortals, and (b) turn them into distributive singular terms. (MFC, 97)

According to this view 'triangularity', 'the concept *triangular*', 'the concept of triangularity' and 'the universal triangularity' are all meta-linguistic DSTs referring to all uses of •triangular•s, i.e. to all rule-governed uses of signs, which function as 'triangular' functions in the English language. In other words, 'triangularity', 'the concept *triangular*', 'the concept of triangularity' and 'the universal triangularity' are all a •the•triangular•• and therefore equivalent to ••triangular•s•. Likewise, 'that the class of hypocritical politicians is not empty' and 'the proposition that the class of hypocritical politicians is not empty' are meta-linguistic DSTs referring to all correct uses of sentential expressions which function in their language as the expression 'the class of hypocritical politicians is not empty' functions in English. Hence, 'that the class of hypocritical politicians is not empty' and 'the proposition that the class of hypocritical politicians is not empty' are both •the•the

class of hypocritical politicians is not empty••s and thus equivalent to ••the class of hypocritical politicians is not empty•s•.

The general thesis that all abstract singular terms are meta- and inter-linguistic DSTs is obviously highly contentious. It will only be sustainable if it can be shown that *all* statements, in which, for example, 'triangularity', 'that the class of hypocritical politicians is not empty' and 'the proposition that the class of hypocritical politicians is not empty' function as abstract singular terms, can be paraphrased by meta-linguistic statements in which these expressions are replaced by their dot-quoted counterparts •the•triangular•• and •the•the class of hypocritical politicians is not empty••. Sellars provides such paraphrases for a fairly large number of statements.[23] A general discussion of these reconstructions is beyond the scope of this chapter but it is also irrelevant within the context of a discussion of Sellars's semantic views. For, even if one remains sceptical about the thesis that all abstract singular terms are meta- and inter-linguistic DSTs, one may still accept the idea that abstract singular terms occurring as grammatical objects within statements of the form '…stands for---' are meta-linguistic DSTs.

Accepting this idea and accepting Sellars's analysis of statements of the form '…stands for---' has the virtue of substantiating Wittgenstein's claim that 'language remains self-contained and autonomous'. Sellars's analysis of the grammatical objects of these sorts of meaning-specifying statements shows that their truth does not only not commit us to the existence of specific concrete items, such as the objects of the *Tractatus*, but that it also does not commit us to the existence of specific abstract objects such as propositions, universals or concepts.

Finally, Sellars's analysis of statements of the form '…means---' and '…stands for---' and his introduction of dot-quotes gives him the tools to explain the *internal relation*

[23] See e.g. his AE and NAO.

between sentences and facts, which the early Wittgenstein stressed in the *Tractatus*, in a way that is in line with Wittgenstein's later claim that there is no 'connexion between language and reality'. According to Sellars, Tractarian 'ladder-language' (cf. TC, 209) statements specifying the 'pictorial relation' between a sentence and the fact it depicts, i.e. statements of the form,

(r) *S* pictures the fact that p,

should be interpreted as,

(r_1) *S* means (or stands for the proposition that) p and it is a fact that p.[24]

For example, the statement,

(s) 'Grass ist grün' pictures the fact that grass is green,

should be interpreted as,

(s_1) 'Grass ist grün' means (or stands for the proposition that) grass is green and it is a fact that grass is green.

To say that it is a fact that grass is green or – more generally – that it is a fact that p, Sellars suggests, is nothing more than to say that it is true that grass is green or that it is true that p (cf. NS, 246; TC, 209-10). Hence (r_1) is synonymous with,

[24] For this suggestion cf. TC, 209 and NS, 245.

(r_2) *S* means (or stands for the proposition that) p and it is true that p.

And (s_1) is synonymous with,

(s_2) 'Grass ist grün' means (or stands for the proposition that) grass is green and it is true that grass is green.

If we apply to (r_2) and (s_2) Sellars's analysis of meaning statements and his analysis of expression of the form 'that p' as dot-quoted DSTs we arrive at the true forms of (r) and (s), which are,

(r_3) 'S's are •p•s and •p•s are true,

and,

(s_3) 'Grass ist grün's are •grass is green•s and •grass is green•s are true.

Thus, statements, which appear to specify the 'pictorial relationship' between sentences and facts, functionally classify the quoted sentential expression in the subject position and then assert that all sentences of that normative-functional kind are true.

To say that a sentence is true, however, is for Sellars not the same thing as saying that it stands in a 'pictorial relationship' or a relationship of correspondence to facts 'in the world'. Sellars takes the meaning of the truth predicate to be fully determined by the two inferences or 'inter-linguistic moves' from 'It is true that p' to 'p' and from 'p' to 'It is true

that p'. In more contemporary terms, he holds the 'deflationary' view that the meaning of the truth predicate is fully determined by the biconditional:

(t) It is true that p if and only if p.[25]

If we apply Sellars's analysis of that-clauses to (t), i.e. his analysis of expressions of the form 'that p', we arrive at the following statement,

(t_1) •p•s are true if and only if p.

All that statements of the form '*S* pictures the fact that p' or its non-tractarian equivalent '*S* stands for the fact that p' tell us is that *S* means p and that p. Statement,

(s) 'Grass ist grün' pictures the fact that grass is green,

for example, tells us that the German sentential expression 'Grass ist grün' means *grass is green* and that grass is green.

[25] For Sellars 'deflationist' account of truth see TC, especially p. 206. He later changed his account of truth to the view that for a sentence or a proposition to be true is for it to be 'semantically assertible', where 'this means not *capable* of being asserted (which it must be to be a proposition at all) but correctly assertible; assertible, that is, in accordance with the relevant semantical rules, and on the basis of such additional, though unspecified, information as these rules may require… 'True', then, means *semantically* assertible ('S-assertible') and the varieties of truth correspond to the relevant varieties of semantical rule' (SM, 101; Sellars appears to use 'proposition' in this passage in its Tractarian sense, i.e. as roughly equivalent to 'sentence'; for Sellars's change of mind cf. the note at the end of TC in SPR, TC, 224; for the most extensive elaboration of his later views cf. SM, chapter IV). I find this later characterization of the meaning of 'true' as 'S-assertible' not very illuminating. First, Sellars does not explain the difference between sentences that are merely assertible and those which are correctly assertible. Second, he does not clarify what the 'additional, though unspecified, information' is that semantical rules 'may require'. Third, it is far from clear whether the difference between merely assertible and correctly assertible (or 'S-assertible') sentences can be elaborated without recourse to the notion of truth, i.e. without specifying the correctly assertible and not *merely* assertible sentences as those sentences that are true. This, however, would mean that the definition of truth as S-assertibility is circular. For reasons of charity I will, therefore, here rely on Sellars's earlier 'deflationary' account of truth. In any event, since his later account of truth is not a version of the 'correspondence theory' of truth either, his change of mind does not significantly affect the present argument.

The necessary character of statements of the form '*S* pictures the fact that p' or '*S* stands for the fact that p', which led the early Wittgenstein to assume that they express an *internal relation* between language and world, is explained by the fact that the first conjunct in statements of the form

(r_3) 'S's are •p•s and •p•s are true,

which are their final analysis, are conceptual truths. However, as we have seen, these conceptual truths do not express a relation between language and world, for the simple reason that they do not express a relation. This means that, contrary to what the early Wittgenstein thought, statements of the form '*S* pictures the fact that p' or '*S* stands for the fact that p' are ultimately not relational at all. Their final analysis is a conjunction of two atomic sentences, none of which is relational.

Still, in his writings on the *Tractatus* Sellars usually claims that there is a relation between sentences and facts (cf. e.g. TC, 211; NS, 245). In his most important essay on the *Tractatus* 'Naming and Saying', for example, he speaks about the relation between sentences and facts as the 'translatability relation' (cf. NS, 245-6). His reason for still speaking about a *relation* between sentences and facts appears to be that his analysis of statements of the form '…means---' and '…stands for---', although it ultimately shows that these statements are not really relational, still relies on a relation of synonymy or 'translatability' between the expression quoted in the subject position and the dot-quoted expression occurring in the object position. It is only because 'Grass is green' stands in the 'translatability relation' to 'Grass ist grün' that we can use '•grass is green•' to functionally classify 'Grass ist grün'. Thus, although statements of the form '…stands for---' and '…means---' do not assert a relation of translatability or synonymy between the expression

in the subject position and the expression occurring in the object position, as can be seen from the fact that only the former but not the latter expression is translated when the whole statement is translated into another language, they presuppose that there is such a relation. It is this synonymy relation, which is a relation within language, namely a relation between two linguistic expressions, which, according to Sellars, the early Wittgenstein mistook for a relation between language and world. Sellars, therefore, concludes:

> The crucial ineffability in the *Tractatus* concerns the relation between statements and facts. Is there such a relation? […] The answer seems to me to be the following. There is a meaning relation between statements and *facts*, but both terms are in the linguistic order. (NS, 245)

Just as, according to the Wittgenstein of the *Investigations*, all that remains of the Tractarian idea of a necessary relation between simple names and simple objects is a necessary relation between ostensively defined expressions and samples, which are part of language, so, according to Sellars, all that remains of the Tractarian internal relation between true sentences and facts is the 'meaning' or 'translatability relation between an expression which may or may not be in our language and one which is' (NS, 245-6). This relation is presupposed by statements of the form '*S* means *p*' or '*S* stands for the proposition that p', which are part of the analysis of statements of the form '*S* stands for (or pictures) the fact that p'. Since the translatability relation is, just as the relation between ostensively defined expression and samples, a relation between a linguistic item whose meaning is being specified and another linguistic item specifying this meaning, we can view Sellars's analysis of 'S stands for (or pictures) the fact that p' as substantiating Wittgenstein's claim in *Philosophical Grammar* that the 'connection between 'language and reality' is made by definitions of words' (PG, 97).

2.3 The intentionality of the mind

Sellars understands the intentionality of mental phenomena in terms of his account of linguistic meaning. To have a thought that p is to have a thought that plays the same normative-functional role in the 'game of reasoning' (cf. SRLG, 324) as the utterance of 'p' in 'language games'. He suggests that ultimately dot-quoted expressions, in particular when they are sentential expressions, do not only apply to all uses of expressions that play the same normative-functional role within 'language games' as uses of the expressions within the dot-quotes, but also to all mental events that play an equivalent role in 'games of reasoning' (cf. SM, 87-88). Thus, the utterance of 'p' and the act of thought that p both stand for the proposition that p because they are both •p•s. Similarly a thought that p stands for or represents the fact that p because it is a •p• and it is true that p.

Intentional mental *states*, such as beliefs and desires, are understood in terms of Sellars's account of acts of thought. Roughly speaking, to have the belief that p is to be disposed to think or judge that p in situations in which the question of whether or not p arises (cf. LTC, 75-77). A desire that p shall be the case consists, again roughly speaking, in the disposition to enjoy thinking that p shall be the case in the relevant circumstances and the disposition to act upon that thought if the circumstances permit it (cf. TA, 117-126).

A detailed discussion of Sellars's account of acts of thought will be provided in the next chapter. I have here merely indicated his account of intentional mental phenomena in order to show how his account of meaning fits into a more general view of the intentionality of linguistic as well as mental phenomena.

I shall now bring this chapter to a conclusion by discussing the later Wittgenstein's account of the relation between intentional phenomena, such as utterances, thoughts, beliefs and desires, and the objects and events in the world they are directed at.

3. Thoughts and objects

One of the great strengths of the picture theory of the *Tractatus* is what at the beginning of this chapter I called its 'direct realism'. The content or sense of a true thought or sentence simply is the state of affairs *in the world* that makes it true. And even the content of a false thought or sentence is a state of affairs that would obtain if the thought or sentence was true. For even the false thought or sentence is made up of simple elements, i.e. names, whose meanings are elements of reality, i.e. objects. Thus, if I assert or think truly that Caesar crossed the Rubicon then it is the fact or state of affairs obtaining in the world and containing the conqueror of Gaul and a river in northern Italy that is the content of my thought or utterance. And if I say or think falsely that Caesar killed Mark Antony then the content of my thought still contains Caesar and Mark Antony. So it is still of Caesar and Mark Antony, the actual men, that I mistakenly think or claim that the former killed the latter.

The 'pictorial relationship' serves the purpose of explaining this intimate connection of thoughts and sentences with the objects in the world they are about. Since thoughts and sentences are made up of simple of names, which stand in an internal relation to the objects in the world that are their meaning, and since their sense is a function of the meanings of their constituents, we can say that their sense or content, i.e. what they say or stand for, is identical with what is the case if they are true.

In the first chapter I tried to bring out the importance of this direct realism by presenting Wittgenstein's criticism of a theory of belief and desire, which Bertrand Russell developed in his *The Analysis of Mind* (Russell 1921). The aspect of Russell's view that Wittgenstein is objecting to comes out most clearly in Russell's treatment of the relation between an expectation and its verification (cf. Russell 1921, 269-71). According to Russell an expectation, like all propositional attitudes, consists of an attitude, a content and

an objective (cf. ibid, 233). Both attitude and content are 'present occurrences in the believer' (ibid.). The objective, which Russell variously calls 'fact' (ibid, 232), 'event' (ibid, 233) or 'occurrence' (ibid, 270), on the other hand, is an item in the world which stands in an 'external and causal' (ibid.) relation to our expectation. Our expectations are verified or fulfilled if their objective causes us to experience a 'feeling of expectedness' (ibid.).

As I have tried to show, Wittgenstein suggests in his *Philosophical Remarks* (cf. especially chapter III) as well as in his Cambridge lectures of Lent term 1930 (LWL, 1-14; cf. especially, 9-10) that this view leads to a dilemma: Either we assume that whatever occurrence gives us a 'feeling of expectedness' is the fulfilment of our expectation. This, however, seems implausible. There is no reason to think that it is impossible that something that we did not expect still gives us a feeling of expectedness when it happens. Or we assume that if we expect that p will happen and experience a feeling of expectedness, we also have to recognize the feeling of expectedness as the right sort of feeling for the expectation that p will happen, rather than, say, for the expectation that q will happen. However, recognizing something seems itself a phenomenon very similar to the verification or fulfilment of an expectation. Hence, according to Russell's view it will itself involve a feeling of recognition or expectedness. This feeling of recognition will again have to be recognized as the feeling of recognition for the feeling of expectedness for the expectation that p will happen, rather than the feeling of recognition for the feeling of expectedness for the expectation that q will happen, and so on *ad infinitum.*

Within the Tractarian framework, which Wittgenstein still at least partially endorsed when he first criticised the theory of belief and desire of *The Analysis of Mind*,[26] this dilemma is avoided by the assumption that intentional acts and states stand in an internal

[26] For this point see Hacker 1996, 30; for Wittgenstein's continued endorsement of the idea that sentences and other bearers of content are pictures at this time of his life see also PR, 57, 63 and LWL, 1 and 4.

'pictorial relationship' to the states of affairs in the world that are their senses. The existence of this relationship between representations, on the one hand, and states of affairs, on the other, implies, first, that the facts, which Russell calls the 'objectives' of representations, are identical with their content and, second, that the relationship between representations and their objectives is not 'causal and external'. According to the Tractarian view, what we expect, in the sense of the content of our expectation, is identical with what we expect, in the sense of the event in the world that has to take place for our expectation to be fulfilled. And the relationship between an expectation and its content or fulfilment is internal and necessary. This is why the *Tractatus* can account for the fact that when we expect some specific event E to happen it is really only E that we expect. We do not expect something that feels or looks like E but E itself. Similarly if we think, believe or say that Caesar crossed the Rubicon what we think, believe or say is identical with the actual historical event involving Caesar and the Rubicon.[27]

Wittgenstein never abandoned his opposition to Russell's idea that the relation between intentional phenomena, such as thoughts, utterances, beliefs, desires and expectations, and their objectives in the world is causal. In the *Investigations* the core of his criticism is repeated in the following remarks – albeit without mentioning Russell by name:

> Saying "I'd like an apple" does not mean: I believe an apple will quell my feeling of non-satisfaction. *This* utterance is an expression not of a wish but of non-satisfaction. (PI §440)
>
> Could a justification of an action as the execution of an order run like this: "You said 'Bring me a yellow flower', whereupon this flower gave me a feeling of satisfaction; that's why I've brought it"? Wouldn't one have to reply: "But I didn't tell you to bring me a flower that would give you that sort of feeling in response to my words!"? (PI §460)

[27] Russell takes the relation between a belief about the past, such as the belief that Caesar crossed the Rubicon, and its objective to be basically the same kind of relation as the relation between an expectation and its objective. What according to Russell connects my belief that Caesar crossed the Rubicon with its objective is that I experience a feeling of expectedness when confronted with evidence for the fact that Caesar crossed the Rubicon (cf. Russell 1921, 270-71).

The target of these remarks is clearly the kind of causal account of the relationship between intentional phenomena and their objectives advocated by Russell. If I desire an apple, the objective of my desire is to eat a particular kind of fruit and not whatever event causes me to cease to desire an apple – such as a punch in the stomach, as Wittgenstein suggested in his initial criticism of Russell in the *Philosophical Remarks* (cf. PR, 63). Similarly if I order someone to bring me a yellow flower, I do not order him to bring me whatever floral item gives him a particular feeling after having received the order.

In the *Philosophical Investigations*, just as in his initial criticism of Russell's view in *Philosophical Remarks* and his lectures of 1930, Wittgenstein makes clear that a causal theory of belief and desire fails to account for the *internal* and *necessary* character of the relation between intentional acts and states and their objectives. This comes out quite clearly in the following remarks:

> "Thinking must be something unique." When we say, *mean*, that such-and-such is the case, then, with what we mean, we do not stop anywhere short of the fact, but mean: *such-and-such – is – thus-and-so.* (PI §95)
>
> "A thought – what a strange thing!" – But it does not strike us as strange when we are thinking. A thought does not strike us as mysterious while we are thinking, but only when we say, as it were retrospectively, "How was that possible?" How was it possible for a thought to deal with *this very* object? It seems to us as if we had captured reality with the thought. (PI §428)
>
> A wish seems already to know what will or would satisfy it; a sentence, a thought, to know what makes it true – even when there is nothing there! Whence this determining of what is not yet there? This despotic demand? ("The hardness of the logical must"). (PI §437)

What Wittgenstein is talking about when he is talking about us not stopping 'anywhere short of the fact' in what we say, the 'hardness of the logical must', the 'determining of what is not yet there' and the ability of thought to 'capture reality' is the same phenomenon as the one he tries to capture in the *Tractatus* when he speaks of pictures as reaching 'right out' to reality: the ability of intentional phenomena, such as thoughts, beliefs and utterances, to determine exactly at what objects, events or states of affairs in the world they are directed.

Wittgenstein is clearly right about this phenomenon. Thought, utterances, beliefs, desires and so forth clearly stand in an internal, necessary relation to the objects and events in the world at which they are directed. If I think, say or believe that Caesar crossed the Rubicon, my thought, utterance or belief does not only necessarily stand for the proposition that Caesar crossed the Rubicon it is also necessarily about a particular historical event taking place at a particular place and time and involving a particular man and a particular river. If my thought, utterance or belief was not directed at this particular event involving this particular man and this particular river, it would not be the thought, utterance or belief that Caesar crossed the Rubicon; just as it would not be the thought, utterance or belief it is, if it did not stand for the proposition or content that Caesar crossed the Rubicon.

However, one might wonder, how the later Wittgenstein can account for this fact after rejecting the idea that there is a 'pictorial relationship' between our representations, such as our thoughts and utterances, and their 'objectives' in Russell's sense of this word. After all, it is the 'pictorial relationship' that explains in the *Tractatus* how our representations come to be internally related to the objects, events and states of affairs in the world at which they are directed. It might therefore appear doubtful whether the later Wittgenstein can explain the internal character of the relationship between representations and their objectives and whether he can avoid the kinds of problems faced by Russell's theory.

It may seem as though Sellars's analysis of sentences of the form '*S* stands for the fact that p' provides the resources for explaining the phenomenon Wittgenstein tries to capture in PI §95, when he says that '[when] we say…that such-and-such is the case, then… we do not stop anywhere short of the fact', without relying on the discredited idea of a 'pictorial relationship'. For this analysis explains how a true thought or sentence that p comes to

stand necessarily for the fact that p. As we have seen, Sellars believes that to say that a thought or sentence stands for the fact that p is to say that it is (i) necessarily a •p• and (ii) that •p•s are true. For example, the thought that Caesar crossed the Rubicon or the sentence 'Caesar crossed the Rubicon' stand necessarily for the fact that Caesar crossed the Rubicon because (i) they are both necessarily •Caesar crossed the Rubicon•s and (ii) •Caesar crossed the Rubicon•s are true. What we say or think when we say or think truly that p is that p and that p is a fact. Hence what we think, i.e. that p, does not 'stop anywhere short of the fact' because that p is a fact and what we think *is* that p.

However, Sellars's notion of a fact is different from the notion of fact in play in PI §95. PI §95 occurs in the context of a discussion of the *Tractatus*[28] and it seems that 'fact' is here used in the sense in which it used in the *Tractatus* as well as in Russell's work. For the early Wittgenstein facts are complexes of objects and in *The Analysis of Mind* Russell identifies facts with the 'objectives' of propositional attitudes existing in extra-mental reality. Sellars, on the other hand, employs Frege's notion of a fact as a true proposition.[29] For Sellars a sentence or thought stands for a fact if it stands for a true proposition. And to stand for a true proposition is for him not a matter of standing in a relation to some item outside of language and thought but to play a certain role and to be true. Facts belong to what he calls the 'logical order' or 'space of reasons' rather than to the 'natural order'.[30]

There is reason to think that the later Wittgenstein agrees with Sellars's Fregean conception of facts as true propositions. Clearly neither Wittgenstein nor Sellars share Frege's platonistic conception of what propositions are, but they agree with Frege that facts are true propositions. Wittgenstein repudiates his earlier conception of facts as complexes

[28] For this see in particular the two immediately following sections PI §§96-97.
[29] Using 'thought' in the sense of 'proposition' Frege puts this as follows: 'What is a fact? A fact is a thought that is true' (Frege 1918-19, 74)
[30] This is also why Sellars says in the passage from 'Naming and Saying' I have quoted above that '[there] is a meaning relation between statements and *facts*, but both terms are in the linguistic order' (NS, 245). Sellars usually treats the phrase 'linguistic order' as equivalent to such phrases as 'logical order', 'intentional order' or 'space of reasons'.

of objects in a number of places. In the *Blue Book*, for example, he claims that this conception is the result of a 'confusion' (cf. BB, 31). The 'confusion', or rather mistake, he seems to have in mind is a tendency to reify 'objects of thought', when we mean by this phrase the content of what we are thinking rather than the objects we are thinking about (cf. BB, 31 & 38). In other words, it is the result of a tendency to think of propositions, i.e. the contents of our thoughts, as objects existing in the world. And in *Philosophical Grammar* he claims that it is 'misleading' to say of the fact that a certain circle is red that it is a complex of a circle and redness and that Frege had already told him so (PG, 200).[31] I think the following remark from *Philosophical Grammar* contains a fairly explicit rejection of the tractarian conception of facts as complexes of objects in favour of a conception of facts as true propositions:

> Complex is not like fact. For I can, e.g., say of a complex that it moves from one place to another, but not of a fact. But that this complex is now situated here is a fact. (PG, 199)

What Wittgenstein is saying here is that a fact is not a complex item in the world which can move about in space but that it is what a true expression of the form 'that p' stands for – for example the expression 'that this complex is now situated here'. This indicates an endorsement of the idea that a fact is a true proposition, because expressions of the form 'that p' are typically thought to stand for propositions.

If this interpretation is right, it confirms my suggestion that Sellars's explanation of what it means for a sentence or a thought to 'stand for' a fact is in the spirit of Wittgenstein's later views. However, it also means that Sellars's analysis of expressions of the form '*S* stands for the fact that p' cannot help to illuminate the 'internal' relation between thoughts and sentences and their 'objectives' in Russell's sense, i.e. the objects and events in the world at which they are directed. If facts are true propositions they are

[31] Frege made the point in a letter to Wittgenstein from June 28th 1919 (Frege 1989, 19-20). An English translation of the relevant section of Frege's letter can be found in Künne 2009, 35.

contents, and the 'relationship' between intentional phenomena, such as thoughts and sentences, and their contents is not really a relationship at all. For a thought or a sentence to stand for a fact is for it to play a certain role and to be true.

The question now is: How can the later Wittgenstein explain the necessary or 'internal' character of the relation between intentional acts and states and their 'objectives' after he has rejected the 'pictorial relationship'? How can he explain the fact that when we think that Caesar crossed the Rubicon we are really thinking of Caesar, the actual man, and the Rubicon, the actual river, when he rejects the idea that this thought is made up of elements whose 'meanings' are elements of reality?

Wittgenstein's remark that there is no 'connexion between language and reality' (WVC, 209-10) and his claim that the 'connection between 'language and reality' is made by definitions of words… so that language remains self-contained and autonomous' (PG, 97) might lead one to assume that he simply has no account of the necessary relation between representations and the objects they are directed at. However, this would be a mistake. Wittgenstein's point about the 'autonomy of grammar' and his criticism of the idea of a 'connexion between language and reality' is not that there is no such relation but rather that what he took to be a 'pictorial' relation between language and mind, on the one hand, and items in the world, on the other, is really a relation within language. This comes out quite clearly in the following remarks from *Philosophical Grammar* and *Philosophical Investigations:*

> "The sentence determines in advance what will make it true." Certainly, the sentence "p" determines that p must be the case in order to make it true; and that means:
> (the sentence p) = (the sentence that the fact p makes true). And the statement that the wish for it to be the case that p is satisfied by the event p, merely enunciates a rule for signs:
> (the wish for it to be the case that p) = (the wish that is satisfied by the event p).[32] (PG, 161-62)

[32] I have slightly changed Kenny's translation. I have chosen to translate 'Satz' with 'sentence' rather than with 'proposition' in order to make clear that Wittgenstein speaks about sentences rather than their contents.

> The agreement, the harmony, between thought and reality consists in this: that if I say falsely that something is *red*, then all the same, it is *red* that it isn't. And in this: that if I want to explain the word "red" to someone, in the sentence "That is not red", I do so pointing to something that *is* red. (PI §429)
>
> But now one might ask: what does his coming look like? – The door opens, someone walks in, and so on. – What does my expecting him to come look like? – I walk up and down the room, look at the clock now and then, and so on. – But the one sequence of events has not the slightest similarity to the other! So how can one use the same words in describing them? – But then perhaps I say, as I walk up and down: "I expect he'll come in." – Now there is a similarity here. But of what kind?!
> It is in language that an expectation and its fulfilment make contact. (PI §§444-45)
>
> "An order orders its own execution." So it knows its execution before it is even there? – But that was a grammatical proposition, and it says: if an order runs "Do such-and-such", then *doing such-and-such* is called "executing the order". (PI §458)

The essential point of all these remarks is the following: The 'internal relation' between intentional phenomena, such as utterances (PI §458, PG, 161),[33] thoughts (PI §429) and propositional attitudes (PI §§444-45, PG, 161-62) and the objects and events they are directed at is explained by the grammatical truth that we attribute or express the former with the same words with which we describe and refer to the latter. For example, if Jack expects that Jill will come into the room, he will express his expectation with such phrases 'Jill will come in', 'I expect Jill to come in' or 'I expect that Jill will come in'; and we will attribute this expectation to him with such statements as 'Jack expects that Jill will come' or 'Jack expects Jill to come in'. If Jill comes into the room and thereby fulfils Jack's expectation, we, as well as Jack himself, can employ the same words to speak about the event that fulfils his expectation, namely 'Jill came in', 'Jill is coming in', 'The event of Jill coming into the room occurred', etc.

What is true of expectations and their fulfilment is equally true of all other intentional acts and states. If Jack claims that Caesar crossed the Rubicon we can use the words, he uses to make that claim, and which we use to attribute the claim to him, to identify the historical event at which his claim is directed and which is responsible for its truth, i.e.

[33] I take it that when Wittgenstein talks about sentences in the first of the remarks quoted above (PG, 161) he should be interpreted as speaking about utterances of sentences. This comes out in his remark 'The sentence determines in advance what will make it true'. Sentences do not have a location in time, utterances of sentences do.

Caesar's crossing of the Rubicon. In other words, if Jack claims 'Caesar crossed the Rubicon' we can attribute this claim to him with such phrases as 'Jack asserted that Caesar crossed the Rubicon' or 'Jack claimed that Caesar crossed the Rubicon' and we can talk about the historical event at which Jack's assertion is directed and whose occurrence is responsible for its truth with such phrases as 'Caesar crossed the Rubicon', 'Caesar's crossing of the Rubicon' etc. Similarly if Jack thinks, believes, hopes, wishes or fears that Jill will come for dinner, the words we use to attribute this thought, belief, hope, wish or fear to him and which he himself might use to express it are the same words we use to specify the event at which his thought, belief, hope, wish and fear are directed.

We typically attribute intentional acts and states to people by combining an expression referring to a person with an expression specifying what type of act or state we are attributing to her (i.e. assertion, thought, belief, expectation etc.) and with a so-called 'content-clause', i.e. a that-clause specifying the content of the intentional act or state attributed. The resulting sentences have the general form 'A Vs that p', where 'A' stands for the person to whom the intentional act or state is attributed, 'Vs' for the type of intentional act or state attributed, and 'that p' for the content of that act or state. If the person to whom the intentional act or state is attributed speaks our language, she can use the sentential expression occurring in the content clause or a grammatical transformation thereof to express the attributed intentional act or state; e.g. if we attribute the belief that p to her, she can use 'p' to express her belief.

When Wittgenstein says in PI § 445 that '[it] is in language that an expectation and its fulfilment make contact' he is saying that the connection between an expectation and its objective, i.e. the object or event at which the expectation is directed, is explained by the fact that if the expectation is correctly expressed by 'p' and correctly attributed by a statement of the form 'A expects that p' then we can use 'p', or at least some of the

expressions occurring in 'p', to speak about its objective. Likewise for all other intentional acts or states: if A Vs that p, it is always correct to use 'p', or some of the expressions occurring in it,[34] to describe the objective of A's Ving.

Furthermore, this phenomenon is not accidental. It is a conceptual truth – in Wittgenstein's terminology a 'grammatical truth' – that if A Vs that p, then we may use 'p' to speak about the objective of her Ving. This is the point of his remark in *Philosophical Grammar* that statements such as 'The sentence determines in advance what will make it true' and '[The] wish for it to be the case that p is satisfied by the event p' merely express the rules for signs: '(the sentence p) = (the sentence that the fact p makes true)' and '(the wish for it to be the case that p) = (the wish that is satisfied by the event p' (cf. PG, 161-62). The context of these remarks makes clear that the target of their criticism is the Tractarian 'pictorial relationship' and Wittgenstein's swift transition from talk about facts to talk about events indicates that he is here using 'fact' in its Tractarian sense, i.e. as referring to complex items in the world that are the objectives of true sentences. Hence, what he is saying in this passage is that the internal relation between sentences in use and wishes, on the one hand, and their objectives, on the other, is not explained by invoking an internal 'pictorial relationship' between mind and world but rather by the humdrum linguistic rule that whatever we may call 'the sentence p' or 'the utterance p' we may also call 'the sentence [or utterance, S. B.] that the [Tractarian, S. B.] fact p makes true' and what we may call 'the wish for it to be to be the case that p', we may also call 'the wish that is satisfied by the event p'.

In the *Investigations* Wittgenstein makes the same point when he says that the statement 'An order orders its own execution' is a grammatical proposition, which says that 'if an order runs "Do such-and-such", then *doing such-and-such* is called "executing

[34] I shall in future drop this qualification, but the reader should keep it in mind.

the order"' (PI §458). The internal relationship between an order and its execution is not explained by assuming that there is a 'pictorial relationship' between the signs produced in a speech act and an action but rather by pointing out that it is a grammatical truth that if the order says 'Do A!' then doing A is what is called 'executing the order to do A'. Ultimately, all that remains of the Tractarian necessary 'pictorial relation' between thought and language, on the one hand, and the world on the other, is the necessity of a linguistic rule that tells us that if A Vs that p then we may use the expressions we use to specify the content of A's Ving to also specify the objective of her Ving.

This observation not only explains how we can expect the very thing that fulfils our expectation and have thoughts and beliefs about the very states of affairs and events in the world that make our thoughts and beliefs true, it also explains how even false and negative representations can 'reach right out to reality' and can be 'laid against reality like a measure' (cf. TLP 2.1511-2.1512). Obviously, if Jack thinks that Caesar did not cross the Volga and Jill thinks that Caesar killed Mark Antony then there are no events at which Jack's and Jill's thoughts are directed. After all, Caesar neither crossed the Volga nor killed Mark Antony. Accordingly, we cannot simply use the *sentences* occurring in the content-clauses of 'Jack thinks that Caesar did not cross the Volga' and 'Jill thinks that Caesar killed Mark Antony' to speak about events that took place in the real world and whose occurrence is responsible for the truth of Jack's negative thought and the falsity of Jill's positive thought. Still we can use the expressions 'Caesar', 'Volga' and 'Mark Antony', which occur in these content-clauses, to speak about actual objects, namely two Roman generals and a river in Russia. And it is because we use 'Caesar' to refer to Caesar, 'Mark Antony' to refer to Mark Antony and 'Volga' to refer to the Volga that Jack's negative thought and Jill's false thought are still directed at Caesar, Mark Antony and the Volga. It is this idea, I take it, that Wittgenstein expresses rather cryptically in PI §429

when he say that the 'agreement, the harmony, between thought and reality' consists, on the one hand, in the idea 'that if I say falsely that something is *red*, then all the same, it is *red* that it isn't' and, on the other hand, in the idea 'that if I want to explain the word "red" to someone, in the sentence "That is not red", I do so pointing to something that *is* red'.

3.1 Crane's objection

In a forthcoming paper Tim Crane raises an instructive objection to Wittgenstein's account of the relation between intentional acts and states and their objectives. He sketches the following scenario: Suppose you are expecting the postman to deliver your mail on Tuesday. Unbeknownst to you, the local parson, Mr. Smith, supplements his income by working as a postman. On Tuesday you are surprised to find him in your yard with a mailbag. It seems that your expectation that the postman will deliver the mail on Tuesday has been fulfilled; after all, Mr. Smith *is* the postman. But it has been fulfilled in a way you find surprising, since you did not expect that Mr. Smith would deliver the mail. Crane puts it this way: 'An expectation… can be fulfilled even if it is fulfilled in ways you do not expect' (Crane forthcoming, 20). He suggests that Wittgenstein cannot accommodate this fact:

> What this shows is that you can describe what actually fulfils your expectation – what I shall call the *object* of your expectation – in a way that is independent of the expectation itself. Wittgenstein's point, by contrast, is that you can only describe the object of the expectation in the way it is specified in the description of the expectation itself. Yet this does not seem to be true: you did not expect Mr Smith the parson to deliver your mail, yet this is what fulfils your expectation that the postman deliver your mail. (ibid.)

Crane claims that this point is not restricted to expectations but also applies to thoughts, propositions, beliefs, desires and other intentional acts and states (cf. ibid, 22-4). For example, I might believe that Cicero was a great public speaker without believing that Tully was a great orator, although Tully and his exceptional oratory skills are the objectives of my belief.

Crane's observation is accurate. Our expectations can be fulfilled in ways that we find surprising and might not even recognize as their fulfilment. However, he is wrong in claiming that Wittgenstein cannot accommodate this fact. In PI §441 Wittgenstein's quotes a line from Goethe's *Hermann and Dorothea*, which shows his awareness of the phenomenon Crane is highlighting: 'Denn die Wünsche verhüllen uns selbst das Gewünschte' ('For our wishes themselves conceal from us that which we wish for').[35] The context of this quote within Goethe's poem shows its appropriateness to Crane's concern. It occurs within a passage in which the local pastor is pleading with Hermann's father to consent to Hermann's marriage to Dorothea, who falls *prima facie* short of the father's expectations:

> Be not amazed and perplext, that thus all at once there appeareth
> That which you long have desired: your wish is not realized, truly,
> Under the very same form which your fancy presented it to you;
> For our wishes themselves conceal from us that which we wish for:
> Gifts come down from above, retaining their own form and aspect.
> (Goethe 1857, 89)

Wittgenstein's allusion to this passage shows that he was aware of the phenomenon Crane's objection exploits. Crane's mistake is to ascribe to Wittgenstein the view that one 'can *only* describe the object of the expectation in the way it is specified in the description of the expectation itself' (Crane forthcoming, 20; my italics). This is not what Wittgenstein is saying. He is merely saying that when A Vs that p we *can* describe the object of A's Ving by using 'p', not that we *have* to do so. And this seems to be true. If Mr. Smith was not working as a postman to supplement his income, his delivering the mail on Tuesday would not be a case of the postman delivering the mail and would therefore not be the fulfilment of the expectation that the postman will deliver the mail on Tuesday. Similarly if

[35] P. M. S. Hacker in his commentary on this passage points out that this is a line from Goethe (cf. Hacker 1996, 101-02). In the translation of the *Investigations* by Anscombe, Hacker and Schulte the quote is rendered as: 'For wishes themselves are a veil between us and the thing wished for'. I have chosen to use the translation from Goethe 1857 (the translator is not named in this edition) since it is from this translation of Goethe's poem that I will quote below.

Hermann's marriage to Dorothea could not be described as p, if Hermann's father's desire was that p, then it would not be the fulfilment of the father's desire. The point of the pastor's words is that Hermann's marriage would be a matter of p taking place and therefore would be a fulfilment of the father's desire, although it might not appear that way on first sight.

4. Conclusion

At the outset of this chapter I quoted the following words from *Philosophical Grammar*:

> Like everything metaphysical the harmony between thought and reality is to be found in the grammar of the language. (PG, 162)

We are now in a position to fully appreciate this remark. In the *Tractatus* Wittgenstein developed a complex metaphysics of symbolism to account for the relation between thought and reality: Thoughts and fully analysed sentences are made up of simple elements that stand in an internal relation to elements of reality, which (i) share their 'logical form' and (ii) are their meaning. This internal relation between the simple elements of thoughts and fully analysed sentences and the simple elements of reality establishes a 'pictorial relationship' between thoughts and sentences, on the one hand, and states of affairs that are their senses, on the other. This metaphysical link between thought and reality explains how we can think exactly what is the case, if our thought is true, and how even false thoughts can 'reach right out' to reality.

When Wittgenstein returned to the topic of intentionality in his later work, he realized that the idea of such a metaphysical link between thought and reality was radically mistaken. Like Wilfrid Sellars he came to reject the idea that intentional acts and states stand in a relation to their content. He never put forth anything amounting to a full account of how mental acts and states, such as thoughts and propositional attitudes, come by their

content. But his endorsement of what Sellars calls a non-relational account of intentionality comes out quite clearly in his criticism of the 'pictorial relationship' and in his account of linguistic meaning: The meaning of a linguistic expression is not anything in extra-linguistic reality onto which it is projected but its role in language games. What may look like relations between linguistic expressions and items in the world that are their meaning, i.e. the relation between ostensively defined expressions and the objects by means of which they are defined and the 'relation' between true sentences and facts, are really relations within language. The former is a relation between a linguistic expression and a sample, which is a 'means of representation' and is 'part of language'; and the latter is, if Sellars is right, not really a relation at all. According to Sellars's analysis all there is to the 'pictorial relation' between sentences and facts is a relation of translatability or synonymy between two sentential expressions, which is presupposed by statements of the form 'S stands for the fact that p'.

Finally the necessary relation between intentional acts and states and their objectives or 'intentional objects', as they are sometimes called,[36] is due to the grammatical fact that if the act or state is correctly attributed by statements of the form 'A Vs that p' and can be correctly expressed by 'p' then it is also correct to use 'p', or at least some of the expression occurring in 'p', to speak about the objective of A's Ving.

All that remains of the internal 'pictorial relation' and the idea of a shared 'logical form' of thought and language, on the one hand, and extra-mental reality, on the other, is a set of grammatical relations. That is why Wittgenstein says that '[like] everything metaphysical the harmony between thought and reality is to be found in the grammar of the language'.

[36] Cf. e.g. Crane (2001, 13-18).

IV. Sellars on Thinking

1. Introduction

I have suggested in the preceding chapter that Wittgenstein takes the *internal relation* between intentional acts and states and their objectives, i.e. the objects and events at which they are directed, to be explained by the fact that we may use the same expressions we use to attribute or express the former to speak about the latter. I have also suggested that the content of these acts and states, i.e. the property in virtue of which they are directed at their objectives, is not something to which they are related. In other words, Wittgenstein's conception of intentionality is – in Sellars's terms – 'non-relational'.

However, I have not explained how the later Wittgenstein thought mental acts and states *acquired* their content. In the second chapter I have presented Wittgenstein's – clearly only schematic – account of how acts of speech acquire their content; namely by playing a particular role in 'language games'. But I have not yet addressed the question of how, according to the later Wittgenstein, intentional *mental* phenomena, such as thought, belief and desire, come by their content. In particular, I have not addressed the question of whether we can find anything amounting to a positive account of thinking in Wittgenstein's later work, which can be seen as replacing the mistaken conception underlying the semantics of the *Tractatus*.

The reason for this omission is that Wittgenstein does not provide such an account. He is famous for his concern for the diversity of uses of language. It seems to me that this type of concern is even greater in his philosophy of mind. A constant theme running through his

remarks on the mind is the enormous diversity of mental phenomena. Here Wittgenstein is particularly concerned to 'teach us differences'.[1]

In Wittgenstein's thinking about intentional mental phenomena this concern for the diversity of the mental is reflected in two particularly salient features. *First*, he stresses that the content of intentional mental events and states depends to a very large extent on their context. In *Philosophical Grammar* he makes this point by discussing the case of someone imagining another person:

> Suppose I say something like: 'What I see in my mind isn't just a picture which is like N (and perhaps like others too). No, I know that it is him, that he is the person it portrays.' I might then ask: *when* do I know that and what does knowing it amount to? There's no need for anything to take place during the imagining that could be called 'knowing' in this way. Something of that sort may happen after the imagining; I may go on from the picture to the name, or perhaps say that I imagined N, even though at the time of the imagining there wasn't anything, except perhaps a kind of similarity, to characterize the image as N's. Or again there might be something preceding the image that made the connection with N. And so the interpretation isn't something that accompanies the image; what gives the image its interpretation is the *path* on which it lies. (PG, 147)

Whether my image is an image of N or of someone else who merely looks like N depends on the context in which it occurs or the 'path on which it lies', for example on what I do before or after imagining N; or, as Wittgenstein puts it on another occasion, on 'certain transitions [I] make or would make' (LW I §308).[2] In PI §581 Wittgenstein repeats this point by discussing expectation:

> An expectation is embedded in a situation, from which it arises. The expectation of an explosion may, for example, arise from a situation in which an explosion *is to be expected*.

Wittgenstein is not saying that we can only expect an explosion when such an event is likely to occur, but rather suggests that what makes my state of mind at a particular time an expectation of an explosion is not anything residing exclusively in my consciousness at that time but the context in which it is situated. One such content-determining aspect of the

[1] M O'C. Drury relates (in Fann 1967, 69) that Wittgenstein once thought of using as a motto for PI a quotation from *King Lear*: 'I'll teach you differences'.
[2] This point is also highlighted by Malcolm Budd (1989, 137) and William Child (2011, 115-119).

context might be, but need not be, the fact that an explosion is likely to occur; other aspects might be what I am doing or saying or what I would do or say given specific circumstances.

The *second* salient feature of Wittgenstein's thinking on intentional mental acts and states is his concern with the multiplicity of the mental phenomena referred to by a single expression. In PI §577, for example, he says:

> We say "I'm expecting him" when we believe that he'll come, though his coming does not *occupy* our thought… But we also say "I'm expecting him" when it is supposed to mean: I'm eagerly awaiting him. We could imagine a language in which different verbs were consistently used in these cases. And similarly, more than one verb where we speak of 'believing', 'hoping', and so on. The concepts of such a language would perhaps be more suitable for understanding psychology than are the concepts of our language.

In *Zettel* he makes essentially the same claim, now focussing on the expression 'thinking':

> Remember that our language might possess a variety of different words: one for 'thinking out loud'; one for thinking as one talks to oneself in the imagination; one for a pause during which something or other floats before the mind, after which, however, we are able to give a confident answer. One word for a thought expressed in a sentence; one for a lightning thought which I may later 'clothe in words'; one for wordless thinking as one works. (Z §122)

The intention of these remarks, and others like them, is clearly to 'teach us differences', i.e. to make us aware of the multiplicity of the phenomena we refer to by means of vocabulary for intentional acts and states, such as 'thinking, 'believing', 'expecting' and 'hoping'.

I believe the diversity of intentional mental phenomena, which comes out both in their context dependency and in the fact that we typically refer to a variety of quite different phenomena by a single expression, is Wittgenstein's main reason for not giving any positive account of thoughts, beliefs and desires. His reason is not some sort of philosophical 'quietism' but rather the conviction that this diversity means such an account cannot be given and attempts to do so will only lead to a distorted picture of the grammar of these concepts. This conviction comes out quite clearly in the following remark on thinking:

> We are not at all *prepared* for the task of describing the use of e.g. the word "to think" (And why should we be? What is such a description useful for?)
>
> And the naïve idea that one forms of it does not correspond to reality at all. We expect a smooth contour and what we get to see is ragged. Here it might really be said that we have constructed a false picture. (Z §111)

Despite these discouraging remarks I will now turn to an attempt to provide an analysis of our concept of thought, namely Wilfrid Sellars's account of thinking.

Sellars did not share Wittgenstein's scepticism about our ability to describe the use of 'to think'. He was not entirely oblivious to the diversity of the phenomena we call 'thought'. For example, he suggests that thinking out loud (cf. e.g. MEV, 282), talking to oneself in the imagination (cf. e.g. SK, 304-05), as well as mental occurrences without a conscious or publicly observable vehicle (cf. e.g. SK, 321) are all forms of thinking. Still he thought he could provide an analysis of our concept of thinking as well as of a number of related concepts pertaining to intentional mental act and states.

Sellars's account of thinking stands at the centre of his analysis of intentional mental phenomena more generally. All the other intentional mental acts and states he discusses are understood in terms of this account. Volitions and intentions are practical thoughts, i.e. thoughts with an action-directed content, and beliefs and desires are dispositions to have specific kinds of thoughts in specific types of circumstances.[3] Hence, for Sellars understanding intentional mental phenomena in general presupposes understanding the nature of thinking. Accordingly I will here focus exclusively on the latter.

Before discussing Sellars's account of thinking I will try to present the philosophical setting in which it is located. I will conclude the chapter by discussing whether Sellars's account, which can be understood as claiming that there is such a thing as a 'language of

[3] For Sellars's account of intention and desire see his TA and IILO, for his account of volition see his VR.

thought', can avoid the later Wittgenstein's objections to the idea of a language of thought, as he seems to have presupposed it in the *Tractatus*.

2. The myth of the given

Sellars is most famous for his attack on the so-called 'myth of the given' which is at the centre of his seminal essay 'Empiricism and the Philosophy of Mind' (EPM, henceforth). The mistaken assumption, which according to Sellars lies at the root of much of the empiricism of his time, and which he attacks under the label of the 'myth of the given', consists in the idea that, *firstly*, the senses provide us with an awareness of facts, an awareness that something is thus-and-so, which is independent of and prior to the use of language and symbolic representation; and that, *secondly,* this awareness would give us a point of reference outside of our conceptual framework, providing us with a firm and immutable foundation for all our empirical knowledge.

Sellars rejects the myth on the basis of four insights central to his philosophy. *First*, perceptual awareness of the kind that can provide *reasons* for further empirical claims involves the application of concepts to one's perceptual environment. *Second*, the possession of a concept depends on the ability to use a word expressing that concept. *Third*, in order to be able to use a word with meaning, and thereby grasp the concept that is expressed by that word, one has not only to be able to use that word correctly when presented with its conditions of application, but one should also be able to recognize at least some of the *material* inferential consequences of its application; an ability that is exercised when one moves from the use of the word in question to the use of other words in inferentially related sentences in the 'language game'. *Fourth*, and most controversially, according to Sellars one only counts as being able to use a word if one is not only able to

employ it in accordance with the rules that govern its use, but is also aware of one's own capacity to do so.[4]

Only if a subject meets these four conditions can her reactions to her environment considered to be *rational* responses possessing *propositional content*. Only if the subject is able to use language and make and assess inferences does she become a concept using creature, whose responses to the environment belong into what Sellars famously calls 'the space of reasons' (EPM, 169). By entering the space of reasons one's responses to one's environment acquire a *normative status* that distinguishes them from any merely 'natural' reactions to the environment such as stimulus response reactions. For only the former kind of responses but not the latter are subject to *norms of rationality*. Only they can justify and be justified and can be assessed for correctness in the light of other, inferentially related claims. Sellars famously concludes his discussion of the myth of the given in EPM with a remark that speaks about the need of dispositional and occurrent states of knowledge to be placed in the space of reasons. A remark that should be seen as applying not only to occurrent and dispositional states of knowledge but to any dispositional or occurrent state of a person that possesses propositional content.[5] He claims:

> The essential point is that in characterizing an episode or a state as that of *knowing*, we are not giving an empirical description of that episode or state; we are placing it in the logical space of reasons, of justifying and being able to justify what one says. (EPM, 169)

[4] How controversial this fourth criterion turns out to be depends, obviously, on what one believes Sellars takes to be implied by it. In his account of perception, as presented in EPM, this fourth criterion is operative in his demand that in order for a perceptual judgment to express perceptual *knowledge* it not only has to be accompanied by knowledge of further, inferentially related claims but also by an awareness that one's act of judging is an exercise of a reliable differential disposition to react with observational concepts to features of one's environment. I believe Sellars takes this demand to be an application to the use of concepts in perceptual judgements of his more general demand, that in order for someone to count as a speaker and thinker at all she not only has to be able to make a sufficiently complex use of object language expressions and have a variety object-level thoughts, but also has to be able to use meta-language and have meta-level thoughts, in which she can criticize and correct her object level use of concepts. This demand, I believe, Sellars takes to imply that in order for someone's object language statement to express knowledge, the speaker also has to know that she is able to use the words she is employing. Whether this general demand and its application to the use of concepts in perceptual judgements are tenable will depend on how Sellars understands our knowledge of our own reliable differential dispositions and the use of concepts on the meta-level. Unfortunately, I do not think that Sellars has ever arrived at a clear view on these issues.

[5] For this point cf. Brandom (1997), 160.

In particular the third and fourth of the above requirements, i.e. the requirement that one should be able to recognize at least some of the *material inferential consequences* of the application of a concept to one's environment and the requirement that one be able to assess one's own thoughts and utterances by speaking and thinking about them, lead Sellars to a non-foundationalist, anti-empiricist account of empirical knowledge. According to Sellars, in order to be able to make a perceptual judgement, one also has to be able to make at least some of the judgements that follow from it as well as be able to make meta-judgements about one's first-level perceptual judgement. And in order for one's perceptual judgement to express *perceptual knowledge* one also has to know at least some of these inferential consequences and meta-level claims. Putting this point in terms of the 'observation reports' and 'empirical propositions' that give expression to perceptual knowledge and the non-perceptual empirical knowledge with which it is inferentially related, Sellars concludes that 'if there is a logical dimension in which other empirical propositions rest on observation reports, there is another logical dimension in which the latter rest on the former' (EPM, 170).

3. Concept empiricism

A central element of Sellars's rejection of the myth of the given and a clear upshot of his views as presented so far, is his rejection of what he variously calls 'concept empiricism' (ITSA, 307, 311) or the 'abstractive theory of concept formation' (SM, 19ff., 27f.; CSCA, 8.12.1961). According to concept empiricism the linguistic neophyte grasps the meaning of an expression of the language she is acquiring by simply associating certain sounds or sign designs with items in her experience. She grasps the meaning of the colour word 'green', for example, by noticing that her parents use the word 'green' in connection with green items in their environment. In his 'Is There a Synthetic A Priori?' (ITSA, henceforth) Sellars caricatures concept empiricism with the following words:

> A mind is about to learn the meaning of the word 'red'. The abstract entity in question is lurking in the manifold of sense. But so are many others. This one stands out clearly. Here! and here! No, that can't be it! Aha! a splendid specimen. By the Methods of Mill! That must be what Mother calls 'red'! (ITSA, 309)

This is the same account of language acquisition as the one provided by the 'Augustinian' picture of language, which Wittgenstein attacks in the opening sections of the *Philosophical Investigations* (roughly PI §§1-64). In EPM Sellars explicitly refers to the Augustinian picture and Wittgenstein's criticism of it (cf. EPM, 161-62); and his caricature of concept empiricism in the above quote from ITSA is strongly reminiscent of the famous passage from Augustine's *Confessions* quoted at the beginning of the *Investigations*:

> When grown-ups named some object, and at the same time turned towards it, I perceived this, and I grasped that the thing was signified by the sound they uttered, since they meant to point *it* out. This, however, I gathered from their gestures, the natural language of all peoples, the language that by means of facial expression and the play of eyes, of the movement of the limbs and the tone of voice, indicates the affections of the soul when it desires, or clings to, or rejects, or recoils from, something. In this way, little by little, I learnt to understand what things the words, which I heard uttered in their respective places in various sentences, signified. And once I got my tongue around these signs, I used them to express my wishes. (PI §1)

Rejecting concept empiricism or the Augustinian picture of language is not a matter of rejecting a primitive empirical theory in child psychology. It is something that can be done on a priori grounds.[6] It is fairly obvious that concept empiricists have great difficulties in explaining how logical, mathematical and abstract vocabulary is acquired, for we cannot observe logical, mathematical and abstract objects or properties.[7] However, Sellars (and Wittgenstein) reject concept empiricism not only for logical, mathematical and abstract expressions, but claim that it even falls short of giving an adequate account of how we acquire primitive observational vocabulary. Even the acquisition of vocabulary for immediately observable properties such as colour, temperature, shape and size, is not adequately accounted for by concept empiricism.

[6] For this point see Geach (1971), 18. Geach attacks concept empiricism under the label 'abstractionism' (1971, 18-44).
[7] Cf. ibid., 22-32.

Concept empiricists seem to be committed either to a form of conceptual atomism or to the idea that we are somehow innately endowed with conceptual abilities. Both these ideas are inconsistent with Sellars's reasons for rejecting the myth of the given.

Let us first turn to *conceptual atomism*. According to concept empiricism the meaning of words is learned by simply associating these words with sensible properties. Unless we assume that learning the meaning of a word does not consist in the acquisition of a new conceptual ability, but merely in learning which sign design is associated with which concept, concept empiricism seems to imply that we could acquire concepts one by one. We simply associate 'yellow' with yellow things or sensations of yellow, 'red' with red things or sensations of red and 'round' with round objects or impressions of something round and thereby acquire the concepts *yellow*, *red* and *round*.

This idea is inconsistent with what I above characterized as Sellars's third and fourth reason for rejecting the myth of the given; namely, his claim that possessing a concept and knowing the meaning of a word presupposes both a recognition of at least some of the inferential consequences and antecedents of its application and an awareness of one's own competence in the application of the concept and the use of the corresponding word.

If a fledgling speaker has acquired a tendency to utter 'yellow' in the presence of yellow objects, this does not mean that she has acquired the concept *yellow*. It does not mean that she is able to judge that something is yellow; and even if she is able to utter the whole sentence 'This is yellow' reliably in the presence of yellow objects this does not mean that her words express the corresponding observational judgment. In order for that to be the case she has to (i) recognize at least some of the inferential consequences of applying the concept *yellow* to an object and (ii) be aware of the fact that her inclination to apply the concept *yellow* to an object is *ceteris paribus* a reliable indicator of the presence of something yellow (cf. EPM, 164-70). Unless the fledgling speaker meets these two

requirements her utterance of 'This is yellow' is at most expressive of certain discriminative abilities and she can be used as a colour indicator in the same way a thermometer can be used as a temperature indicator. Sellars takes these discriminative abilities to be a necessary condition for the possession of conceptual abilities but rejects the idea that they are a sufficient condition for it.

If the concept empiricist rejects conceptual atomism, then she is committed to the idea that we are *innately* endowed with conceptual abilities. Acquiring a language will not mean for her acquiring a range of conceptual abilities, but merely learning which word is used to express which concept. The concept empiricist will assume that the linguistic neophyte already possesses a conceptually structured awareness of her surroundings and merely has to find the right correlations between her concepts and the words of the language she is learning. Language learning will then proceed in the way described by Augustine. However, as Wittgenstein rightly points out, this view also implies that we have to think of the linguistic neophyte as though she already spoke some kind of innate 'language of thought' before she learned to speak a public language:

> Augustine describes the learning of human language as if the child came into a foreign country and did not understand the language of the country; that is, as if he already had a language, only not this one. Or again, as if the child could already *think*, only not yet speak. And "think" would here mean something like "talk to itself". (PI §32)

Just as conceptual atomism, this idea of an innate language of thought is inconsistent with Sellars's rejection of the myth of the given. It contradicts his fundamental assumption – characterized above as his second reason for rejecting the myth – that possessing a concept is a matter of being able to use a word of a public language.

4. Psychological nominalism

As we have seen, Sellars believes that, first, all those mental acts and states of awareness – paradigmatically perceptions – which can play a role in reasoning, possess a conceptually

structured content and that, second, conceptual abilities depend on the capacity to use language. Sellars calls the view resulting from the endorsement of these two claims 'psychological nominalism' (cf. EPM, 160). According to psychological nominalism, all awareness that can play a role in thought is 'a linguistic affair' (ibid.).[8]

Sellars's criticism of concept empiricism is supplemented by a positive theory of how concepts are acquired in the course of learning one's first language. According to Sellars we do not form concepts by, as it were, reading them off reality, nor are we innately endowed with conceptual abilities. We rather acquire them by being *trained* or c*onditioned* by those who teach us our first language to perform utterances in accordance with the rules of the language we are learning.

As I have discussed in the previous chapter, for Sellars the meaning of linguistic expressions is their 'functional role' in what he calls 'pattern-governed behaviour'. Pattern-governed behaviour is 'behaviour which exhibits a pattern, not because it is brought about by the intention that it exhibit this pattern, but because the propensity to emit behaviour of the pattern has been selectively reinforced, and the propensity to emit behaviour which does not conform to this pattern selectively extinguished' (MFC, 86-87).

Since Sellars thinks that language should be understood in terms of its cognitive role, i.e. as something in which we *think* and *reason* (cf. MFC, 81-85), rather than in terms of its role as an instrument of communication, he focuses on three kinds of linguistic performances corresponding to the cognitive acts of perception, inference and action. As I have pointed out in the previous chapter, the *first* performances are 'language entry

[8] The use of the label 'psychological nominalism' to refer to the thesis that all awareness depends on the ability to use language is a bit misleading, as the word 'nominalism' is normally used to refer to a metaphysical theory, namely the denial of the existence of abstract objects, and not to a theory in the philosophy of mind. However, in Sellars's defence it should be pointed out that his psychological nominalism is closely related to his endorsement of metaphysical nominalism (cf. EAE, 259). Abstract entities are, according to him, nothing more than linguistic roles and grasp of them does not consist in a relation between a thinker and an abstract object, but in the ability to employ a linguistic item instantiating that role. Grasping the abstract object *redness* consists in being able to use the word 'red' or any other word with an equivalent role.

transitions', i.e. linguistic responses to non-linguistic stimuli; the *second* are 'intra-linguistic moves', i.e. transitions from one utterance to another; and the third 'language departure transitions', i.e. utterances which are followed by non-linguistic performances such as raising one's hand (cf. MFC, 87-88). All utterances that have an equivalent norm-governed role within the network of language entry transitions, intra-linguistic moves and language departure transitions within their respective languages mean the same.

In order for a person's utterance of 'p' to count as meaning that p, it has to occur as part of that person's pattern-governed behaviour. The utterance of 'This → is yellow'[9] will only count as meaning that the object ostensively referred to is yellow if the word 'yellow' appears in the speaker's behavioural repertoire not only as a reaction to yellow objects but also within acts of inferring and utterances which are bound up with other linguistic and non-linguistic actions. The speaker has to be able to use 'yellow' not merely, as concept empiricism claims, as a reaction to yellow objects, but should also be capable of moving from, say, 'This → is yellow' to 'This → is coloured' and to follow utterances such as 'I'll look it up in the yellow book' *ceteris paribus* with looking something up in a yellow book. Thus, in order for her utterance of 'This → is yellow' to count as meaning what it usually does, namely that the object ostensively referred to is yellow, she has to be able to use 'yellow' in accordance with the rules of her language not only in language entry transitions, but also in intra-linguistic moves and language departure transitions.

The ability to speak a language and the correlated ability to use concepts in both speech and thought is acquired in essentially three *logically* but not *chronologically* distinct steps (cf. MFC, 85-89). *First*, the linguistic neophyte is trained or conditioned to display the appropriate pattern-governed behaviour. It is important to note that this first step is not a matter of *teaching* or *instructing* the fledgling speaker to act in accordance

[9] As in the previous chapter I use the arrow '→' to represent a deictic gesture.

with the rules of language, but of *training* or *conditioning* her to do so. Being instructed or taught to perform certain actions usually requires the pupil to have a conception of the actions she is learning to perform and the circumstances in which to perform them. In the case of acts of speech, such as the simple perceptual judgment 'This → is yellow', this would mean that she could already grasp that a certain kind of act of speech is appropriate in certain circumstances; for example, that 'This → is yellow' is the appropriate answer when queried about the colour of a yellow object in one's immediate vicinity. However, such a prior conception of the act of speech and the circumstances in which to perform it on the part of the linguistic neophyte is just an instance of the kind of pre-linguistic awareness that is precluded by Sellars's 'psychological nominalism'. Pattern-governed linguistic behaviour – unlike the kinds of activities which we are instructed or taught to do, such as playing poker or reading music – can be displayed by a person who has no conception of what she is doing. For, as we saw, pattern-governed behaviour is not 'brought about by the intention that it exhibit [a] pattern, but because the propensity to emit behaviour of the pattern has been selectively reinforced, and the propensity to emit behaviour which does not conform to [the] pattern selectively extinguished' (MFC, 86-87).[10]

The *second* step in learning one's first language is to acquire the ability to speak not only *in* language but also *about* language; that is, learning to speak in a meta-language. In this second step the fledgling speaker learns to not only express her thoughts, perceptions and wishes with the help of language but also to react to the utterances of others and herself with such utterances as 'X said that p' or 'I said that p'.

The second step paves the way for the final *third* step. In the third step the language learner acquires the ability to speak not merely in accordance with rules of language but

[10] It is for the same reason that Wittgenstein frequently speaks about language learning as training.

also *about* these rules. She thereby learns to recognize the linguistic mistakes of others and of herself and to correct violations of the rules of language as well as to alter these rules for her own purposes by proposing linguistic innovations.

Only after the language learner has gone through these three steps will she have reached linguistic maturity and will count as a fully competent speaker. For in order to count as a competent speaker she has to act not merely in *accordance* with the rules of language but she also has to be able to display some kind of grasp of them; an ability that is exercised, for example, when a person corrects linguistic mistakes, proposes new definitions or otherwise intentionally changes the rules of the language she is speaking.

However, the fact that the speaker should possess some kind of grasp of the rules guiding her use of language should not be taken to imply that whenever a mature speaker uses a bit of language she consults the rules of her language. Apart from certain exceptions, such as correcting linguistic mistakes by others and oneself, speaking a language remains at the level of pattern-governed behaviour. Sellars illustrates this with the help of the following example:

(A) Jones	All men are mortal. So, no non-mortals are men.
(B) Smith	If I am entitled to 'all men are mortal', I am entitled to 'no non-mortals are men'. I am entitled to the former, I state it thus: All men are mortal So, I am entitled to the latter, I state it thus: No non-mortals are men. (MFC, 88)

Only Jones is inferring *in* language that no non-mortals are men from the premise that all men are mortal. Smith is exhibiting a piece of practical reasoning about linguistic entitlements, which he then proceeds to exercise. If one assumes that all speech is a matter of acting in the light of the rules of language, as Sellars caricatures it in this example, one is committed to the myth of the given. For one has to assume that before we, as it were, clothe our thoughts in words, we first have to be aware of the rules that guide the use of the words we are employing, and, secondly, have to recognize the presence of the application

conditions of these words prior to actually using them. If we assume that these two conditions have to be fulfilled by any observation report, we arrive at the idea which Sellars calls 'the heart of the Myth of the Given', namely the claim that 'observation "strictly and properly so-called" is constituted by certain self-authenticating nonverbal episodes, the authority of which is transmitted to verbal and quasi-verbal performances when these performances are made "in conformity with the semantical rules of the language"' (EPM, 169).

Sellars's psychological nominalism consists not only in the idea that we acquire conceptual capacities only in the course of learning a language, but also in the idea that speaking is a *sui generis* form of exercising these capacities.[11] Language is not only a tool to transmit information, but is itself a medium in which we think and judge. Thinking of language exclusively as a kind of communicative tool, Sellars believes, can easily lead to a form of the myth of the given; namely to the idea that speakers form intentions to use language in order to communicate a specific piece of information on the basis of an awareness of facts, which is prior to and independent of all language use. Hence, it is very important for Sellars that we can *think* in language; what he calls 'thinking-out-loud' is just as much a form of thinking as thinking silently (MEV, 282).

His belief that conceptual abilities are acquired in the course of learning one's first language and his claim that thinking out loud is a genuinely cognitive activity lead Sellars to the conclusion that we learn to think silently by learning to think out loud. Conceptual abilities are acquired by learning to speak and the capacity to exercise these abilities silently or in private is, accordingly, acquired through exercising them out loud in public utterances. Only someone who can speak can have the kind of conceptually structured

[11] This view is obviously already implicit in his focus on the cognitive role of language. For Sellars's most sustained discussion of it, see his LTC.

thoughts that occur within the 'space of reasons' i.e. that can serve as premises and conclusions in reasoning and can be criticised in the light of norms of rationality.

5. Perception and self-knowledge

After having brought his criticism of the myth of the given to a conclusion, Sellars devotes the second half of EPM to giving an account of both perceptual knowledge and self-knowledge, which is in line with his psychological nominalism and his rejection of the myth of the given. Implicitly relying on his thesis that thinking out loud is a genuine form of thinking, Sellars explains how we learn to apply concepts to our environment in perceptual judgements and to our own mental goings-on in judgements expressing self-knowledge.

The nature of perception, in particular visual perception, was one of Sellars's main philosophical concerns and is the topic of a significant part of his published work (apart from EPM cf. e.g. SM, SK, PH, FMPP, SSOP). The details of his views are complicated and their interpretation is a matter of considerable controversy.[12] However, Sellars's core ideas on perception can be sketched in a few sentences.

We have already encountered his claim that all awareness is an exercise of conceptual capacities. In the case of perception this means that all acts of perceiving have, apart from their phenomenal character, a conceptual core. We learn to apply concepts to our environment by being trained to react to environmental stimuli with verbal behaviour. In order to count as a genuinely cognitive response to one's environment this verbal behaviour has to occur as part of pattern-governed linguistic behaviour, i.e. it has to be part of a network of not only language entry transitions but also of intra-linguistic moves and language departure transitions. Furthermore, in order for a verbal response to a perceptual

[12] The difficulties of interpreting Sellars's views on perception comes out quite clearly in the controversies between Brandom, McDowell, Rosenberg and deVries, who all criticize or defend slightly different interpretations of Sellars's views. Cf. Brandom (2002), Brandom (1997), McDowell (1998), McDowell (2009), deVries (2008), deVries and Coates (2009), Rosenberg (2007).

stimulus, to enter the 'space of reasons', i.e. to count as expressing a perceptual claim or judgement, the subject has not only to be able to react reliably to her environment with the appropriate verbal responses, but also has to be aware of her own reliability. A perceptual judgement can, according to Sellars, be backed up by the subject by a further argument highlighting the reliability of her own responses (cf. EPM, 168-69; SK, 340-45). Thus the ability to apply concepts to one's environment is acquired by, first, learning to react reliably to features of one's environment with appropriate verbal responses and, second, learning to place these responses in a network of further inferentially related verbal activity, which essentially includes the capacity to recognize the reliability of these reports.

Of course, this awareness of one's own reliability does not have to occur at a very sophisticated level. As I understand Sellars, the subject does not have to be able to do more than, for example, confidently say something like 'I know a red object when I see one' when her claim that she is seeing a red object is challenged in order to count as fulfilling the second requirement for perceptual knowledge.

His account of introspection, i.e. of judgements about our own mental states, Sellars develops in EPM by, as he puts it, 'making up a myth of [his] own' (EPM, 178), the so-called myth of Jones. The task of the myth of Jones is to give an account both of how we can acquire the concept of a mental event, in particular the concept of an occurrent thought and the concept of a sensation, and how we can come to use these concepts in introspective judgements. By telling us how we acquire the concepts of thought and sensation it also tells us what these concepts amount to. Thus the myth of Jones tells us what we mean when we ascribe a thought or a sensation to another person or to ourselves, thereby providing us with an analysis of our concept of thought.

Sellars tells the myth of Jones in two steps, the first discusses our concept of thinking and the second our concept of sensation. In what follows I will confine myself to the first

step of the myth of Jones, i.e. Sellars's discussion of our concept of thinking; I will ignore the second step, i.e. Sellars's discussion of our concept of sensation. For Sellars sensations do not have intentional content or representational purport. A discussion of them, therefore, does not fall within the scope of this thesis.

6. Thinking

The main purpose of the myth of Jones within the expository architecture of EPM is to give an account of two interrelated and, by Sellars, not clearly distinguished phenomena. *First*, it is meant to give an account of how we can acquire the concepts of thought and sensation, which is in line with his rejection of the myth of the given and his psychological nominalism. For him this means giving an account of how we can acquire the concept of an episodic mental event, such as an occurrent thought or a sensation, in the light of his insight that 'even such "simple" concepts as those of colors are the fruit of a long process of publicly reinforced responses to public objects… in public situations' (EPM, 176). *Second*, it aims to provide an explanation of how these concepts function in introspective judgements, i.e. judgements expressing awareness of one's own thoughts and sensations,[13] which both avoids the myth of the given and explains their special authority.

In his *Science and Metaphysics* (SM henceforth), Sellars retells the myth of Jones with a slightly different focus. SM, which Sellars himself calls a 'sequel' to EPM (cf. SM, viii), has a less narrowly epistemological focus than the earlier work and is only tangentially concerned with the myth of the given. Accordingly, Sellars turns his attention away from what I just claimed was the second explanatory goal of the myth of Jones, namely the attempt to give an account of introspective judgements which avoids the myth of the given (cf. SM, 155). In his discussion of thoughts in the third and sixth chapters of SM, Sellars

[13] My repeated use of the notion of *introspection* might ring alarm bells, as the term 'introspection' is often associated with the idea of a quasi-perceptual account of self-knowledge. As I use the term it should not carry this implication. When I am talking about 'introspective judgements' I merely want to single out judgements that refer to states of one's own mind.

turns his attention instead to three further explanatory goals, which were already present in EPM, but did not receive a similar pride of place. These are, first, to avoid 'logical behaviourism', second, to avoid a commitment to an outdated form of 'the synthetic *a priori*' and third, to stay clear of the 'the *global* argument from analogy' (cf. SM, 71 & 155).

Logical Behaviourism. As in EPM Sellars aims in SM to give an account of thinking that embraces the behaviouristic idea that 'overt behaviour can provide "logically adequate criteria" for mental episodes' (SM, ix; cf. also EPM, 189) while rejecting the claim of what he calls 'logical behaviourism' that all there is to thinking is overt behaviour and dispositions to behaviour. The label '*logical* behaviourism' is aimed to highlight the behaviouristic idea that the relation between ascriptions of acts of thinking to a person and ascriptions of certain behaviour or dispositions to behaviour is an *analytical* one. Sellars, on the other hand, believes that the relation between these two types of ascriptions is logically synthetic.

Synthetic A Priori. It is not at all clear what Sellars means when he says that his account of thinking is supposed to avoid the synthetic a priori. For he also claims that the relation between acts of thinking and their behavioural manifestation *is* an instance of the synthetic a priori, at least as he understands this notion (cf. SM, 68). In his 'Is there a 'Synthetic A Priori'?' (ITSA) he defines synthetic a priori truths as propositions that are conceptually true yet not analytic. Examples of such truths are propositions such as 'Nothing can be red and green all over at the same time' or 'Heavy objects fall to the ground when unsupported'. It is quite clear that he believes propositions expressing the relationship between thought and behaviour to be of exactly this logically synthetic yet a priori nature, as can be seen in his claim that the 'tie between the "overt" and "covert" is "logical" rather than "empirical" – yet "logical" in a way which involves no commitment

to logical behaviourism' (SM, 68). Thus, according to Sellars's own definition of the synthetic a priori, true statements expressing the relation between thought and behaviour are synthetic a priori truths. Accordingly, Sellars maintains, in apparent contradiction to his claim that he wants to avoid the synthetic a priori, that his account of thinking is 'a doctrine of synthetic necessary truth' (ibid.); which means for him synthetic a priori truth, since he takes necessity to be indicative of apriority (cf. ITSA, 300).

What he appears to reject, then, is the idea that the relation between thought and behaviour is an instance of the synthetic a priori in a more traditional sense of this notion. What he rejects under the label 'synthetic a priori' is the idea that there is a relation between thought and behaviour which can be known a priori and is entirely independent of our conceptual framework. The relation between thought and behaviour is conceptual because the rules of our language allow us to move from the statements describing certain behaviour to statements ascribing certain thoughts. It is not conceptual in the sense that we have a priori knowledge of language-independent synthetic necessary connections existing between thought and behaviour.

The Argument from Analogy. Sellars wants to give an account of our concept of thinking which provides an adequate story of how this concept can legitimately figure in both self- and other-ascriptions of acts of thinking. In particular, he wants to avoid the idea, famously defended by J. S. Mill (cf. Mill 1889, 243-44), that we only really know what thoughts and feelings are from our own case and apply the concepts of thought or feeling to other people only because we find them to be physically and behaviourally similar to ourselves. The account of thinking provided by the myth of Jones is intended to explain how we can apply a single concept to both ourselves and others, without claiming, as Mill does, that the ascription of thoughts to others is based on an 'argument from analogy'; an argument, which Sellars claims, has 'convincingly been shown to be incoherent' (SM, 71).

In EPM Sellars develops an account of our concept of thinking in essentially three steps, only the last of which is the myth of Jones itself.[14] He begins by expressing his agreement with what he calls the 'classical tradition' (cf. EPM, 177) in claiming that our concept of thought is the concept of a genuinely *episodic* inner occurrence and not a concept referring to mere behaviour and dispositions to behaviour. However, he distinguishes himself from the classical tradition by disengaging the idea that we have privileged access to our own thoughts from the idea that these thoughts are, as it were, 'self-disclosing'. It was usually held, for example by Descartes, that occurrent thoughts were inner episodes which could not occur without being known to occur by the subject. Chiefly responsible for this idea, according to Sellars, is a mistaken assimilation of thoughts to sensations and feelings. He suggests the 'classical tradition' typically thought that since pains and thoughts are varieties of one and the same species we are always aware of our own thoughts just as we are always aware of our own pains. Just as it makes no sense to say that someone has a pain but does not know it, so it makes no sense to say that someone has a thought but is not aware of it. Accordingly, Sellars rejects along with the idea that thoughts are 'self-disclosing' the idea that 'thoughts belong in the same general category as sensations, images, tickles, itches, etc.' (EPM, 177).

Instead he suggests that our concept of thinking is the concept of an inner episode which (i) has to be sharply distinguished from other non-conceptual inner goings-on such as sensations and feelings; (ii) should not be identified with verbal imagery; (iii) is an episode to which the first person has privileged but not infallible access; and (iv) our awareness of which is not a matter of inner perception. Sellars appears to think that claims (i) to (iv) are interrelated. If one thinks of thoughts as varieties of sensation one is easily led to the idea that our access to our own thoughts is infallible. If thoughts were a matter of

[14] For this way of partitioning Sellars's argument cf. Rosenberg (2007, 172-76).

verbal imagery, they would have a qualitative aspect, i.e. would be varieties of sensation. Finally, regarding thoughts as pieces of verbal imagery can easily lead us to the idea that we perceive our thoughts just as we hear public utterances. Sellars concludes:

> If we purge the classical tradition of these confusions [i.e. the assimilation of thoughts to sensations and the idea of their 'self-disclosing' nature, S.B.], it becomes the idea that to each of us belongs a stream of episodes, not themselves immediate experiences, to which we have privileged, but by no means either invariable or infallible, access… [W]e can "hear ourselves think", but the verbal imagery which enables us to do this is no more the thinking itself than is the overt verbal behavior by which it is expressed and communicated to others. It is a mistake to suppose that we must be having verbal imagery -- indeed, any imagery -- when we "know what we are thinking" -- in short, to suppose that "privileged access" must be construed on a perceptual or quasi-perceptual model. (EPM, 177-78)

The *second* step in Sellars discussion of thinking in EPM is the rejection of the Cartesian idea that '…both overt verbal behaviour and verbal imagery owe their meaningfulness to the fact that they stand to… thoughts in the unique relation of "expressing" them' (EPM, 177). Sellars not only rejects what in the previous chapter I have called the 'dual-process conception of meaning', but all semantic programs, such as H. P. Grice's,[15] which try to explain the meaning of linguistic utterances in terms of underlying communicative intentions. He rather suggests an analysis of the relation between thought and speech, which turns Gricean programs on their head. According to Sellars the meaning of utterances should not be understood in terms of the intentional content of underlying mental acts and states, but rather the intentional content of mental acts and states should be understood in terms of the meaning of linguistic utterances; as he puts it in EPM:

> My… problem is to see if I can reconcile the classical idea of thoughts as inner episodes which are neither overt behavior nor verbal imagery and which are properly referred to in terms of the vocabulary of intentionality, with the idea that the categories of intentionality are, at bottom, semantical categories pertaining to overt verbal performances. (EPM, 180)

[15] The *locus-classicus* for Grice's view on meaning is Grice (1957).

The idea that the intentionality of thought should be explained in terms of the intentionality or meaning of linguistic utterances, leads Sellars to the view that our concept of thinking is *analogical.* The character of this analogical concept is explained in the third and final step of Sellars's explanatory strategy, the so-called 'myth of Jones'.

In this *third* step Sellars imagines a pre-historic community of people who speak a 'Rylean' language, whose fundamental descriptive vocabulary refers to publicly observable objects and their properties. The 'Ryleans' do not possess any concepts referring to inner episodes. They neither possess concepts for thought or other conceptual mental acts nor are they capable of speaking about non-conceptual mental episodes, such as pains or sense-impressions.

In EPM Sellars leaves it somewhat unclear whether the Ryleans use *any* psychological vocabulary or whether they merely lack expressions for the genuinely episodic mental occurrences whose status he seeks to explicate. In later publications he makes clear that 'the strategy of the argument [involves] conceding that a distinguishable stratum of discourse about the mental can be reconstructed along Rylean lines' (SM, 97; cf. also ibid., 151 & 159; SK, 328). I believe we should understand Sellars as stipulating that the Ryleans already possess what he in EPM calls a 'modified Rylean account' (EPM, 180) of thinking and are what he later calls 'verbal behaviourists' (cf. e.g., MFC, SK, SM).[16] They do not speak about thoughts and sensations as we do, but they still ascribe dispositions to verbal behaviour to one another. They even have a primitive concept of occurrent thinking which they use to ascribe special kinds of utterances or proximate short-term propensities to such utterances to each other. What is special about these utterances is that they are not performed with any communicative intention but are cases of what *we*, who already have the concept of a thought, would call 'thinking out loud':

[16] Further strong evidence for the claim that the Ryleans are already verbal behaviourists can be found in a correspondence with Hector-Neri Castañeda which is published on the internet (cf. CSCA, 03.04.1961).

> According to V[erbal] B[ehaviourism], thinking 'that-p', where this means 'having the thought occur to one that-p', has as its *primary* sense *saying* 'p'; and a *secondary* sense in which it stands for a short term proximate propensity to say 'p'. (MFC, 83)

Sellars aims to explore what further conceptual resources would have to be added to the Rylean language in order for them to be in a position to acquire *our* concept of thought (EPM, 179). He thereby, first, tries to give an answer to the question of what conceptual resources are presupposed by *our* concept of thought and, second, tries to show how the Ryleans could acquire this concept in a series of steps which is epistemically sound and, in particular, avoids steps, such as Mill's argument from analogy, which according to Sellars have been shown to be flawed (cf. SM, 155).

Addressing the first question means explaining what conceptual resources the Ryleans need to have 'in order that they might come to recognize each other and themselves as animals that *think*' (EPM, 179). *First*, a central element of Sellars's inferential semantics is the claim that the ability to use ordinary empirical vocabulary such as 'green', 'heavy' and 'mass' presupposes grasp of the kinds of properties and relations made explicit by the use of modal terms (cf. IM, CIL). This means in particular that the capacity to speak about ordinary physical objects in space and time presupposes grasp of at least some of their causal powers and some of the causal relations obtaining between them, a claim succinctly summarized in the title of one of his earliest publications: 'Concepts as Involving Laws and Inconceivable without Them' (CIL). Accordingly, Sellars assumes that the Ryleans can not only use ordinary empirical vocabulary but also modal vocabulary and the subjunctive conditional (cf. EPM, 178). An ability which is already implicitly acknowledged by claiming that they are 'verbal behaviourists', as this implies that they are able to speak about long- and short-term behavioural *dispositions*.

Second, in addition to the basic empirical vocabulary and the correlated modal and causal vocabulary, Sellars stipulates that the Ryleans already possess semantic vocabulary

(cf. EPM, 179). As we have seen, according to Sellars the meaning of an utterance is not determined by the intentions with which it is performed, but by its rule-governed role within the speaker's behaviour. Accordingly, grasping the meaning of an utterance and being able to speak about it does not require grasping the thoughts, beliefs and intentions that brought it about, but merely depends on the ability to recognize its role within pattern-governed behaviour. The ability to recognize this role and to form vocabulary pertaining to meaning, reference, truth and intentionality can be granted to the Ryleans without thereby conceding to them the ability to speak about the kind of occurrent mental states, whose status is to be explained.

Third, Sellars stipulates that the Ryleans are able to explain observable phenomena with the help of something like scientific theories, which postulate unobservable causes. He claims:

> [W]e may suppose these language-using animals to elaborate, without methodological sophistication, crude, sketchy, and vague theories to explain why things which are similar in their observable properties differ in their causal properties, and things which are similar in their causal properties differ in their observable properties. (EPM, 183)

These theories are usually built around a *model*, as according to Sellars most scientific theories are (cf. EPM, 182). The postulated unobservable entities and their interaction are modelled on a selected set of observable entities and their behaviour. The model is accompanied by a *commentary*, which qualifies and restricts the extent of the analogy between the theoretical postulates and the non-theoretical phenomena on which they are modelled.

Sellars maintains that with these three conceptual building-blocks in place the Ryleans are, at least in principle, in a position to acquire *our* concept of thinking. Accordingly, the next step in his 'anthropological science fiction' (EPM, 178) aims to show, how the

concept of thinking could have been introduced within the Rylean community in a series of epistemologically unobjectionable steps (cf. EPM, 183-89).

For this purpose Sellars stipulates that among the Ryleans a genius called Jones appears (cf. EPM, 183). Jones wants to know why his fellow Ryleans (and he himself) behave intelligently 'not only when their conduct is threaded on a string of overt verbal episodes -- that is to say, as we would put it when they "think out loud" -- but also when no detectable verbal output is present' (EPM, 186). Furthermore, he wants to explain the constant changes in our behavioural dispositions, in particular the constant flux of changing short-term propensities to display verbal behaviour, which even the Ryleans recognize (cf. SM, 159); for, as we have seen, the Ryleans are verbal behaviourists and therefore have a rudimentary concept of thinking, according to which a thought is an immediate short-term propensity to perform a certain type of utterance. Accordingly, they recognize, but cannot explain, that 'waking life is a steady stream of… changes in one's verbal propensities' (SM, 73).

Jones develops a theory according to which this 'steady stream of changes in one's verbal propensities' as well as the intelligent actions which surround it, are the *causal* result of postulated inner episodes which he calls 'thoughts' (cf. EPM, 186-87). His model for these postulated causally efficacious inner episodes is overt verbal behaviour, in particular the kind of overt verbal behaviour which we, who already possess the concept of thinking, would call 'thinking out loud'. The relationship of these postulated inner episodes to overt behaviour and behavioural dispositions is comparable to the relationship between the unobservable postulates of a scientific theory and the observable phenomena they are supposed to explain. Accordingly, Sellars describes the relationship between Jones's theory of thoughts and the psychological explanations provided by verbal behaviourism as similar to the relationship between the explanations of the behaviour of physical objects

through micro-physical theories on the one hand, and through descriptions that ascribe certain propensities to these objects on the other. In *The Structure of Knowledge* (SK), which was published about 20 years after EPM, Sellars describes the relationship between verbal behaviourism and a full-blown theory of thought with the following words:

> Can we not regard classical theories of mental acts construed as pure occurrents (as contrasted with short-term propensities to think-out-loud) as theories in a sense which is analogous to micro-physical theories? Indeed, cannot we regard our common-sense conception of thought processes as such a theory? Such a theory would be designed to explain propensities to think-out-loud as micro-physical theory is designed to explain the powers and propensities which we know things to have at the perceptual level… (SK, 328)

Since they are verbal behaviourists, the Ryleans are aware of the fact that propensities to utter certain sentences are generally followed by certain actions or propensities to utter other, inferentially related, sentences. Furthermore, along with a primitive conception of thinking, verbal behaviourism provides the Ryleans with primitive concepts of the, according to Sellars, genuinely dispositional phenomena of belief and desire (cf. LTC, TA).[17] This allows them to give explanations of each other's behaviour in terms of *reasons*. What they are unable to do, is to give proper *causal* explanations of their behaviour and their verbal propensities. This is remedied by Jones's theory by postulating 'inner utterances' which are the causes of our constantly changing verbal propensities and the causal antecedents of the overt behaviour with which these propensities are known to be connected.

Jones's theory carries over to the postulated mental episodes the semantic concepts of truth, reference and meaning which characterize overt utterances. Just as he used to speak about the meaning, truth and reference of overt utterances, Jones's theory now allows him

[17] Roughly a belief that *p*, according to Sellars, is a disposition to think that *p* in situations that elicit one's epistemic commitments. To have a desire that *p* is roughly to have a disposition to form, in appropriate circumstances, intentions of the form 'It shall be the case that *p*'. (In his philosophy of action Sellars uses 'shall' in a technical sense in which it always expresses an intention (cf. TA, 105)). According to *verbal behaviourism*, therefore, a belief that *p* is a disposition to utter 'p' or have a short term propensity to utter 'p' in situations that elicit one's epistemic commitments, and a desire is a disposition to utter 'It shall be the case that *p*' or have a short term propensity to do so in appropriate circumstances.

to speak about the meaning, truth and reference of the newly postulated episodes of 'inner speech'. A thought, according to Jones, has the same normative-functional role within the 'game of reasoning', i.e. within a network of other thoughts and the overt behaviour which is their causal consequence, as the utterance on which it is modelled has within a language game. And since the propositional content of an utterance is determined by its rule governed role within language games, the thought which is modelled on it shares this content.

Furthermore, Sellars thinks that characterizing an utterance as having a specific meaning, i.e. as playing a specific role, implies that it stands in a complex network of *causal relations* with other linguistic and non-linguistic behaviour. His reason for this belief appears to be an idea, which in his 'Truth and 'Correspondence'' he calls a 'principle' (TC, 216) and succinctly states as follows:

> Espousal of principles is reflected in uniformities of performance (ibid.)

Since using words with a specific meaning means accepting the rules or 'principles' governing their use, using a word with a particular meaning requires, according to this principle, that one displays certain 'uniformities of performance'. However, Sellars also makes clear that this principle should not be understood as implying that whenever one engages in an activity governed by rules, such as playing a game or speaking a language, one has to exactly follow all the rules pertinent to what one is doing:

> I am not claiming that to *follow* a principle, i.e. act on a principle, is identical with exhibiting a uniformity of performance that accords with the principle. I think that any such idea is radically mistaken. I am merely saying that the espousal of a principle or standard, *whatever else it involves*, is characterized by a uniformity of performance. (ibid)

These remarks occur in the context of a discussion of language, and I take it that what Sellars is suggesting is something like this: When I utter a sentence 'p', this utterance is

governed by rules, standards or principles, which, among other things, determine a myriad of inferential consequences 'q', 'r', 's', etc. of 'p'. I may count as accepting the rules governing the use of 'p' and therefore as *saying* or *claiming* that p, although I (i) do not, after uttering 'p', continue by uttering 'Therefore q', 'Therefore r', Therefore s', etc. and (ii) might occasionally even draw conclusions from 'p' that are not licensed by the rules of language. The idea should be familiar from the discussion of Wittgenstein's views on rule-following in the second chapter. Just as I might count as saying that p although I do not state all its consequences or draw the wrong conclusions from it, so I might count as following the command 'Add 2!' although I do not perform all the infinitely many additions that the command orders me to do, or may occasionally make a mistake in its execution.

Still, Sellars believes that his idea that 'espousal of principles is reflected in uniformities of performance' implies that describing someone as saying something implies that her utterance stands in a network of regularly correlated linguistic and non-linguistic behaviour, which can be understood in terms of natural laws. The relevant natural laws are obviously not strict causal laws but *ceteris paribus* laws. An utterance such as 'This banana is yellow' is *ceteris paribus* correlated with the presence of a yellow banana and, depending on the speaker's epistemic interests, is *ceteris paribus* followed by an utterance expressing an inferential consequence of its content, for example the utterance 'Therefore it is ripe'. This utterance, in turn, might, depending on the desires and intentions of the speaker, lead *ceteris paribus* to an utterance expressing an intention, such as 'Therefore I shall eat it', which is *ceteris paribus* followed by the action of eating the banana referred to. These *ceteris paribus* laws which connect utterances, in particular those constituting acts of thinking out loud, with further utterances expressing an inferentially related content

and with non-linguistic actions, are for Sellars sufficient to say that the former are the *cause* of the latter (cf. AAE, 199-200).

Along with the semantic properties thoughts are conceived to share these causal properties with the acts of thinking out loud on which they are modelled. What holds for the overt act of thinking out loud 'This banana is yellow' holds equally for the 'inner utterance' of the same sentence in 'mentalese', as Sellars sometimes calls the inner language postulated by Jones's theory (cf. MFC, 89; IKTE, 455 & 456). Given the right interests and background beliefs, the 'inner utterance' of this sentence equally tends to lead to the inner utterance of 'Therefore it is ripe'. And the inner utterance of 'Therefore I shall eat it' just as the act of thinking out loud expressing the same content *ceteris paribus* causes the action of eating the banana referred to. The inner utterances or thoughts thereby causally explain quiet intelligent behaviour. And since thoughts, according to Sellars, are complex causal events that have a tendency to manifest themselves in verbal behaviour they not only explain quiet intelligent behaviour but also our verbal propensities and their actualizations in acts of thinking out loud.

As is, according to Sellars, common practice for scientific theories formulated in terms of a model (cf. EPM, 182), Jones provides a commentary on his model which restricts and qualifies the analogy between the entities of the model and the theoretical postulates. First, the commentary makes clear that although the newly introduced thoughts share some of the causal characteristics of overt utterances they do not share all of them. The occurrence of a thought does not involve the production of sounds or of marks on paper. Nor are thoughts 'the wagging of a hidden tongue' (EPM, 187). The analogy between thoughts and overt utterances is restricted to the semantic properties of these utterances and the causal role they have within human behaviour which underlies these semantic properties. Jones's theory leaves it entirely open how these normatively and causally characterized episodes

are realized. It makes no ontological claims and is accordingly compatible with a dualistic as well as a monistic ontology (cf. EPM, 187).

Second, the commentary emphasizes that not all kinds of utterances are included in the model, but only those utterances which we, who already have the concept of a thought, would characterize as cases of thinking out loud. In contrast to many other kinds of acts of speech, which are performed in the presence of an audience in order to 'achieve perhaps devious, perhaps noble, ends' (SM, 75), acts of thinking out loud are *not* performed 'on purpose'. It is an essential feature of thoughts that they usually do not occur on purpose either. Although it is correct to say that we often perform intentional mental actions, such as thinking through a mathematical or philosophical problem, the individual steps of our reasoning as well as the volition which – according to Sellars – initiates it do not occur because we intend them to do so.

At this point it becomes clear why Sellars should be interpreted as assuming that the Ryleans already possess the conceptual framework of verbal behaviourism. Without any psychological vocabulary, Jones would find it difficult to single out the utterances which are acts of thinking out loud. He would not be in a position to distinguish utterances which are performed 'on purpose' from utterances which are not. Verbal behaviourism, however, provides the conceptual tools to do this. Acts of thinking out loud are simply those utterances which do not occur within a network of dispositions to perform utterances expressing intentions to perform linguistic actions.

Third, Jones's theory does not suppose that thoughts have any experiential component (cf. EPM, 188). Thoughts are not a variety of experience or sensation, as Sellars already claims at the outset of his discussion of our concept of thinking in EPM. Indeed, as the myth of Jones is told, the Ryleans acquire the concept of thinking before they acquire the

concept of a sensation. Accordingly, Jones is in no position to characterize thoughts as varieties of sensation or experience.

Fourth, the fact that thoughts are causally responsible for our overt verbal behaviour does not mean that the meaning of overt utterances will, after the introduction of Jones's theory, be analyzed in terms of them (cf. EPM, 188). Jones recognizes that the semantic concepts, with which he characterizes the newly postulated entities, are primarily applied to overt linguistic behaviour. After all, the myth of Jones is aimed to validate Sellars's claim that 'the categories of intentionality are, at bottom, semantical categories pertaining to overt verbal performances' (EPM, 180). It does so by showing that we can imagine a society which has the concept of linguistic meaning before it has the concept of intentional mental acts.

Fifth, the commentary accompanying Jones's model for thinking stresses that thoughts are 'inner' only in an analogical sense of this term. As Sellars points out they are ''in' language-using animals as molecular impacts are 'in' gases, not as 'ghosts' are in 'machines'' (EPM, 187). What Sellars wants to highlight with this rather cryptic remark is that by talking about thoughts as 'inner' we are not giving them any spatial location but are merely highlighting their role as theoretical postulates, i.e. their role as occurrences which explain observable behaviour but – contra analytical behaviourism – cannot be identified with it. 'Inner' in this context merely means 'imperceptible'. Sellars claims in SM that the widespread talk of thoughts as 'inner' is highly misleading (cf. SM, 169). Beginning with his earliest publications he repeatedly stresses the Aristotelian point that thoughts can only be ascribed to the person *as a whole* (cf. APM; SM, 169-70; MP, 214-16). And in SM he points out that even though the attribution of thoughts usually serves as a causal explanation of other thoughts or overt behaviour it does not imply that one part of a person's body sets another part of it in motion (cf. SM, 170).

We are now finally in a position to say what thoughts, according to Sellars, are: *A thought is an exercise of a person's conceptual abilities. It is ascribed to the person as a whole and it is characterized only with respect to its semantic and correlated causal properties, but not with respect to its ontological realization. The concept of a thought is an analogical concept; it is formed by analogy with the overt linguistic behaviour by which thoughts are said to be 'expressed'. Thoughts explain overt behaviour in the same way the postulates of a scientific theory explain the behaviour of physical objects.*

One upshot of this characterization is that acts of thinking out loud are genuine acts of thinking, as I have already pointed out in the above discussion of Sellars's psychological nominalism. For they share all the features in terms of which Jones's theory characterizes thoughts. They are occurrent acts of persons possessing propositional content and are connected through a variety of *ceteris paribus* causal laws with the person's linguistic and non-linguistic behaviour. In his late paper *Mental Events* (MEV) Sellars indeed makes clear that he has always held 'that the members of a certain class of linguistic events are thoughts' (MEV, 282).

This explains the *prima facie* quite puzzling claim Sellars makes in both EPM and SM (cf. EPM, 186; SM, 151) that Jones's theory is particularly aimed at explaining *silent* intelligent behaviour. 'After all', one might object, 'not only what a man does, but also what he says can often be quite puzzling'. Verbal behaviour stands just as much in need of explanation as non-verbal behaviour. Now Sellars's claim that acts of thinking out loud are a variety of thoughts means that if a person were to accompany all her actions by acts of thinking out loud these actions would not stand in need of the kind of rational and causal explanation Jones's theory aims to provide. For the acts of thinking out loud share the

causal and semantic properties of thoughts proper and therefore provide the desired explanation of the actions they accompany.

However, the thesis that thinking out loud is a genuine form of thinking is somewhat difficult to reconcile with Sellars's repeated claim that the purpose of Jones's theory is not only to explain intelligent behaviour, but also to provide a causal explanation for both propensities to think out loud and their actualization in public acts of speech (cf. EPM, 178 & 186-87). The claim that thoughts provide causal explanations for acts of thinking out loud seems to imply that thinking out loud is not a form of thinking but merely its causal effect. My impression is that Sellars sees the causal explanation Jones's theory provides of thinking out loud and short term propensities to think out loud as fundamentally different from the causal explanation it provides of intelligent behaviour. Unlike other behaviour, thinking out loud is not strictly speaking a causal effect of thoughts but merely the 'overt manifestation' (SM, 151) of a causally complex event. This reading is supported by the fact that Sellars appears to believe thoughts have a natural tendency to manifest themselves in speech, as can be seen in the case of babbling infants (cf. EPM, 178).

After this presentation of Jones's theory, I will now turn to what Sellars calls the 'dénouement' (cf. EPM, 188) of his piece of 'anthropological science fiction', namely his account of introspective judgements. Having developed his theory Jones teaches it to his fellow Ryleans. Equipped with the new theory they are not only capable of applying it to each other but can also use it to ascribe thoughts to themselves. Sellars explains introspection by suggesting that the Ryleans could train each other – and themselves – to make direct reports about their own thoughts, even in the absence of the behavioural evidence that was used to introduce the concepts of the theory. After training has been successfully completed, these reports will be counted among the *criteria* others use to

ascribe thoughts to a person. The training the linguistic neophyte receives in order to acquire the ability to perform direct introspective judgements is similar to the training she receives for direct observational judgments. She is trained to perform introspective judgements when the theory suggests the presence of a particular thought, just as she is trained to make observational judgments when she is in a particular observational situation. The only difference is that in the case of thinking the trainers have to infer the presence of a particular thought from the behaviour of the trainee, while in the case of observation the trainers can themselves *see* the observational phenomena the trainee is being conditioned to respond to. Sellars summarizes his view in the following famous passage:

> [W]hen Tom, watching Dick, has behavioral evidence which warrants the use of the sentence (in the language of the theory) "Dick is thinking 'p'" (or "Dick is thinking that p"), Dick, using the same behavioral evidence, can say, in the language of the theory, "I am thinking 'p'" (or "I am thinking that p.") And it now turns out -- need it have? -- that Dick can be trained to give reasonably reliable self-descriptions, using the language of the theory, without having to observe his overt behavior. Jones brings this about, roughly by applauding utterances by Dick of "I am thinking that p" when the behavioral evidence strongly supports the theoretical statement "Dick is thinking that p"; and by frowning on utterances of "I am thinking that p", when the evidence does not support this theoretical statement. Our ancestors begin to speak of the privileged access each of us has to his own thoughts. *What began as a language with a purely theoretical use has gained a reporting role.* (EPM, 189)

This story requires the truth of Jones's theory – this is what Sellars indicates with his question 'need it have?' If the causally efficacious inner episodes Jones postulates did not exist, then the linguistic neophyte could not be trained to respond to them. At the heart of introspective judgements is a stimulus-response connection between the reported act of thought and the introspective judgement, just as at the heart of observational judgements there is a stimulus-response connection between the environmental circumstances and the observational judgement. Obviously, linguistic training could not set up this connection in the absence of a stimulus. Therefore, the postulated episodes have to exist.

The special authority accorded to the first person in the ascription of thoughts is explained by four interrelated factors. First, only the subject herself is with respect to her own thoughts in a situation similar to the observer responding to phenomena in her visual

field. Only she can be trained to react *directly* to the stimulus. Other people always *infer*, at least in a loose sense of the term, the presence of a particular thought from a set of behavioural criteria, which – although they are conceptually connected to the thought they indicate – do not *analytically* imply its presence. Second, this explains why the first person is granted by the linguistic community the ability to make authoritative judgements about her own thoughts, even in the absence of behavioural criteria. Third, the fact that one can be trained to react directly to one's own thoughts with speech acts stating their occurrence also explains why these acts themselves will be counted among the criteria for the ascription of the same thoughts by others, once the speaker has reached linguistic maturity. Fourth, as Sellars only explains in a correspondence with Hector-Neri Castañeda (CSCA) but not in EPM itself, self-ascriptions of thoughts are usually 'self-referentially incorrigible'; for when they are themselves cases of thinking out loud they contain the thought they are ascribing. The act of thinking out loud 'I think that *p*' contains the act of thinking that p just as the utterance of 'I am uttering 'p'' contains the utterance of 'p' (cf. CSCA, 11.03.1962).

However, a mere stimulus triggered linguistic response is not by itself a judgement. Sellars's account of introspective reports parallels his account of observation reports (cf. CSCA, 03.04.1961). We saw above that, according to Sellars, our linguistic responses to the things we see and hear have to be not merely reliably connected to the presence of the object we observe; they also have to be taken up into the 'space of reasons'. The subject has to recognize at least some of the inferential commitments she is undertaking by making a particular observation report, and she has to be able to use the vocabulary constituting the report in a variety of different circumstances. Furthermore, she has to be aware of the reliability of her own linguistic responses to the environing stimuli.

According to Sellars the same is true in the case of introspective reports. In order for a linguistic response to an event in our own mind to count as a genuine exercise of concepts the subject has to (i) be able to recognize the inferential significance of what she is saying, (ii) be able to employ the psychological vocabulary she is using in a variety of different circumstances, and (iii) be aware of her own capacity to use the employed psychological vocabulary reliably in response to states of her own mind. These three demands amount to the claim that in order to be able to make claims about our own acts of thoughts we have to know Jones's theory. We have to know under what circumstances to apply the psychological vocabulary we are using to others and ourselves, and we have to be aware of the explanatory significance of the use of this vocabulary. Accordingly Sellars concludes:

> As I see it, this story helps us understand that concepts pertaining to such inner episodes as thoughts are primarily and essentially *intersubjective*, as intersubjective as the concept of a positron, and that the reporting role of these concepts -- the fact that each of us has a privileged access to his thoughts -- constitutes a dimension of the use of these concepts which is *built on* and *presupposes* this intersubjective status. (EPM, 189)

In order to bring this section on thinking to a close I will now return to the explanatory targets Sellars set for his account of thinking. This means addressing four questions. *First*, is the account of thinking provided by the myth of Jones in line with Sellars's psychological nominalism and his rejection of the myth of the given? *Second*, does it provide an account of introspective judgments, which both explains their special authority and avoids the myth of the given? *Third*, does it avoid the 'argument from analogy'? *Fourth*, does it provide an account of the relation between thinking and overt behaviour which avoids both logical behaviourism and the synthetic a priori in the traditional sense of this term?

Before I turn to these questions, however, it should be noted that with the myth of Jones Sellars provides a theory of thought which vindicates what I claimed were the first

and second steps of his account of thinking as presented in EPM. *First*, the inner episodes postulated by Jones's theory fit the 'revised classical analysis of our common sense conception of thoughts' (EPM, 178) Sellars seeks to defend. For the concept of thinking developed in Jones's theory is the concept of an inner episode which (i) is sharply distinguished from other non-conceptual inner goings-on such as sensations and feelings; (ii) is distinct from verbal imagery; (iii) is an episode to which the first person has privileged but not infallible access; and (iv) our awareness of which is not a matter of inner perception.

Claim (iv) might appear somewhat problematic because of the close analogy I highlighted above between statements making introspective claims, on the one hand, and observation reports, on the other. However, I take it that despite this analogy Sellars's views on introspection are still sufficiently different from his views on perception to allow him to say that he holds a non-perceptual account of self-knowledge. For (a) in his account of introspective judgements sensations or images play no mediating causal role as they do in his account of perception; (b) observation reports – unlike introspective reports – are not 'self-referentially incorrigible'; and (c) observation reports do not have the kind of disparity between the first and the third person that is an essential feature of introspective reports.

Second, Sellars succeeds in reconciling the classical idea of thoughts as inner intentional episodes, which are causally responsible for intelligent behaviour with 'the idea that the categories of intentionality are, at bottom, semantical categories pertaining to overt verbal performances' (EPM, 180). He does so by telling a coherent piece of 'anthropological science fiction' (EPM, 178) that illustrates how we could have acquired the concept of meaning prior to the concept of thought.

Still, on its own, telling this story is only a necessary but not a sufficient argumentative step to show that our concepts of intentionality have their primary use as semantic concepts pertaining to overt speech. In order to show that this claim is not only possibly true but actually true, Sellars also has to show that a story in which people first have the concept of inner intentional episodes and then acquire the concept of meaning is incoherent. I believe that when Sellars tells the myth of Jones in EPM he takes himself to have already demonstrated this latter point in his discussion of the myth of the given.

There are two ways in which we can conceive of people as already having the concept of inner intentional episodes prior to the concept of meaningful linguistic utterances. First, we could imagine a society of people, let us call them 'contra-Ryleans', who are able to speak about themselves and others as thinking that p or intending to do A without being able to say such things as 'X said that p' or ''p' means *q*'. Such a 'contra-Rylean' society is for Sellars inconceivable; for, as we have seen above, in order for a person to count as a competent speaker she has to be able to not only speak *in* language but also to speak *about* language. This implies that she has to have some semantic concepts.

Second, one might assume that people already have the concept of meaningful inner episodes before they learn to speak. According to this view, the linguistic neophyte is already able to ascribe thoughts to people, and on that basis constructs theories, which explain their verbal behaviour as aiming to communicate these thoughts to others. This suggestion, however, is a form of the Augustinian picture of language learning and is rejected by Sellars. It assumes that the linguistic neophyte is endowed with innate conceptual abilities, which allow her to ascribe thoughts to others and construct theories about their behaviour. This assumption is a version of the myth of the given, since it ascribes awareness of facts to people, which is independent of and prior to the use of language and symbolic representation.

I now turn to the explanatory targets Sellars set for his account of thinking: *First*, in telling the myth of Jones Sellars provides an account of thinking, which is in line with his psychological nominalism and his rejection of the myth of the given. Just as his psychological nominalism requires, Sellars's account of thinking shows that awareness of our own and other people's thoughts requires possession of the concept of thought and that this concept is acquired through learning the use of a word in the course of a 'long process of publicly reinforced responses… in public situations' (EPM, 176).

Furthermore, since the possession of the concept of thought requires at least an implicit grasp of Jones's theory, neither our own nor other people's thoughts are given to us in the way Sellars considers objectionable. As we have seen, Sellars argues that we cannot have perceptual knowledge without (a) applying concepts to perceptual situations, which are learned in the course of acquiring one's first language, (b) knowing at least some of the inferential consequences of the concepts we are employing, and (c) being aware of our own command of these concepts. Similarly, he believes that we cannot be aware of either our own or other people's thoughts without (a_t) possessing the concept of thought, (b_t) knowing at least some of the inferential consequences of the application of this concept – a knowledge which is, of course, an immediate consequence of the required knowledge of Jones's theory; and (c_t) being aware of our own command of the concept of thought.

Second, his account of thinking enables Sellars to explain the special authority accorded to the first person in the ascription of thoughts in a way that avoids commitment to the myth of the given. The first person, like everyone else, has to meet conditions (a_t) to (c_t) before she can make authoritative judgements about her own thoughts. Hence, her special authority is not due to the fact that these thoughts are somehow directly presented to her mind's eye in a way that would involve the myth of the given. It is rather explained by three facts. First, only the first person's judgements about her acts of thought are self-

referentially incorrigible. When A judges that she thinks that p, her judgement contains an act of thinking that p. When other people judge that A thinks that p, on the other hand, there judgement, although it also contains an act of thinking that p, does not contain A's act of thinking that p.[18] Second, only the first person can make judgements about her own thoughts in the absence of criteria, although these will only count as authoritative as long as they generally cohere with her other behaviour. Third, only the first person is with respect to her own thoughts in a situation similar to the observer responding to phenomena in her visual field. Only she can react directly to her own thoughts, other people have to infer the presence of her thoughts from behavioural criteria, which are conceptually connected to them but do not analytically imply their presence.

It is important to note that this last difference between first and third person ascriptions of thoughts is only a difference in degree, not in principle. For Sellars perceptual judgements are spontaneous exercises of concepts in response to perceptual stimuli, which are reliably connected with the presence of the properties and objects referred to in the judgement. However, he grants that there is no reason in principle why we cannot occasionally make these judgements, even if the concepts employed in them do not refer to perceptible objects and qualities. A scientist, for example, who has sufficiently interiorized a theory, might spontaneously and reliably judge certain microphysical events to have taken place when she sees their effects. If she does not infer their occurrence from the effect, Sellars is willing to say that there is a sense in which she perceives these microphysical events. In the same way we might acquire the ability to perceive what other people are thinking. After sufficient training we may be able to spontaneously and reliably react to other people's thoughts with true judgements of the form 'A thinks that p'. Hence,

[18] Obviously, only present-tense ascriptions of thoughts are self-referentially incorrigible. When A judges that yesterday morning she thought that p, her judgement does not contain the thought she is ascribing to herself. However, this fits the fact that the special authority of the first person with respect to her past thoughts is significantly diminished compared to her authority with respect to her present thoughts.

the third difference between third and first person ascriptions of thought resides merely in a difference in the kinds of situation that allow them to make such spontaneous and reliable judgements. While the first person will normally always be able to spontaneously and reliably report her own thought, the third first person will be able to do so only on very few occasions.

Third, since we do not have access to our own thoughts prior to and independently of our access to other people's thoughts, Sellars's account of our knowledge of other minds avoids Mill's arguments from analogy. Mill assumes that we can acquire the concept of thinking by simply being aware of our own thoughts. We then ascribe thoughts to other people because we notice that their behaviour is similar to our own. According to Sellars's account of our concept of thought, we do not understand other people's thoughts in analogy to our own. Rather, the ability to be aware of one's own thoughts presupposes the ability to ascribe thoughts to other people, for it presupposes a command of Jones's theory, which involves the ability to ascribe thoughts to people on the basis of behavioural criteria.

Finally, by explaining our concept of thinking in analogy to theoretical concepts Sellars demonstrates how the relationship between thoughts and their behavioural manifestation can be both conceptual and logically synthetic (cf. SM, 69). Since theoretical concepts are introduced in terms of the observable phenomena they are supposed to explain, they are conceptually connected to these phenomena. However, since the theoretical entities referred to by theoretical concepts are supposed to explain observable phenomena, theoretical concepts are not mere abbreviations for complex observable phenomena and are therefore not analytically definable in terms of them.

Furthermore, the account of the conceptual relationship provided by the myth of Jones does not rely on the conception of the synthetic a priori Sellars rejects. Although true statements expressing relationships between thoughts and behaviour are logically synthetic

a priori truths, they do not express some sort of language independent eternal verities that are recognized through 'rational intuition'. Their truth is due to the fact that they express material inferential relations between behavioural and mental vocabulary. These inferential relations can be recognized by anyone who knows the relevant vocabulary, since being able to employ this vocabulary requires being able to draw these inferences. In other words, the truth of statements specifying relations between thought and behaviour is explained by the fact that they make explicit linguistic rules that have to be recognized by anyone employing the relevant vocabulary; it is not explained by their being some sort of language independent synthetic necessary connections between thought and behaviour that are recognized by means of 'rational intuition'.

7. Wittgenstein's criticism

In the first chapter I attributed to the early Wittgenstein the view that thinking took place in a mental symbolism. Thinking, according to the early Wittgenstein, consists in the operation of mental signs that correspond to the words of public languages (cf. TLP 4.1121) and 'have the same sort of relation to reality as words' (NB, 130). I have discussed this view under the label 'language of thought'.

In the second chapter I presented two, I believe fairly compelling, objections to this view. First, thinking is, as Wittgenstein puts it, the 'last interpretation' (BB, 34). In the case of speech or any other kind of symbolic representation we can always ask how the symbols we employ are meant, or how they are to be interpreted. In the case of symbolic representation a gap can always open up between the symbol and its content. This is because knowing what symbols someone employed does not imply knowing what these symbols stand for. For example, knowing what sentence someone uttered does not imply knowing what she meant by it. This is different in the case of thinking. Knowing what

someone thinks implies knowing the content of her thought; as Wittgenstein puts it: '[If] a thought is observed there can be no further question of an understanding' (PG, 143).

Second, if thinking consisted in the operation of mental signs, the attribution of a thought to a person would imply that somewhere in her body or mind some sign had been produced. Hence, it should make sense to ask where the thought is located and what the bodily organ of thought is. This, however, Wittgenstein argues is not the case. Asking for the location of an act of thought does not make sense (cf. PI §361).

It might seem that these objections to the Tractarian account of thinking should also count against Sellars's account, as there is clearly a sense in which Sellars too believes in a 'language of thought'. After all, according to the myth of Jones, our concept of an act of thought is modelled on our concept of a meaningful act of speech.

However, there is a decisive difference between Sellars's account of thinking and the account presumed in the *Tractatus.* Although Sellars understands thoughts in analogy to public utterances, he does not subscribe to the idea that thinking consists in the operation of mental signs. For Sellars the analogy between thought and speech is explained in purely normative-functional terms. Attributing the thought that p to a person means attributing an occurrent mental state to her that plays the same normative-functional role in the 'game of reasoning' as 'p' plays in language games. It does not mean attributing to her some operation in mental signs.

The purely normative-functional character of Sellars's analysis of the concept of thought allows him to avoid the second objection the later Wittgenstein raised against the Tractarian conception of thinking and to claim that the attribution of thoughts to a person does not imply that a 'bodily part of a person is the 'organ' of conceptual thinking' (SM, 170). The fact that Jones's theory attributed not merely semantic but also causal properties

to thoughts does not imply, Sellars suggests, that thoughts must somehow be locatable in specific parts of our body:

> [It] does not seem logically necessary that what sets part of a person in motion must be a change in another part. It is not logically necessary, in particular, that in order for conceptual activity to cause a change in a part of a person the activity must be a change in another *part* of the person. (ibid)

Furthermore, Sellars suggests, quite in line with Wittgenstein's thinking, that taking the attribution of a thought to imply the attribution of literally 'inner' goings-on, such as brain-processes, is a matter of mistaking the scientific explanation of our cognitive capacities with an analysis of our ordinary concept of a thought (cf. SM, 169-70). However, unlike Wittgenstein, who, as we saw, does not believe that such an explanation is likely to be forthcoming, Sellars is confident that it eventually will (cf. SM, 174).

Finally, Sellars's account of thinking cannot only accommodate the phenomenon Wittgenstein highlights by speaking about thought as the 'last interpretation', it actually helps to explain it. According to Sellars the meaning of a linguistic representation is its normative-functional role. Hence, talking or thinking about the meaning of such a representation means talking or thinking about its normative-functional role. We do this, according to Sellars's analysis of expressions of the form '…means---' and '…stands for---', by putting the expression constituting the representation or some other expression with the same normative-functional role in dot quotes. However, we can also think or speak about a linguistic representation under a different guise; namely by thinking or speaking about its specific 'sign-design'.

The important difference between thought and speech is that we cannot think or speak about thoughts under these two different guises. Our concept of a thought is, according to Sellars, the concept of an occurrent mental phenomenon with a specific normative-functional role. The only way we can refer to an act of thought both in speech and in thought is by specifying its normative functional-role. If we specify the normative

functional-role of someone's thought incorrectly we do not misinterpret the thought but fail to refer to it. And since specifying the normative-functional role of a representation means specifying its content this means that whenever we succeed in referring to someone's thought at all we thereby automatically specify the content of her thought. Hence Wittgenstein is right when he says that 'if a thought is observed there can be no further question of an understanding' (PG, 143).

Conclusion

I began this essay by discussing, in the first chapter, the account of intentionality the early Wittgenstein developed in the *Tractatus*. In the second and third chapter I then discussed various objections to this account and others like it, which can be found in Wittgenstein's later writings. I argued that, according to the later Wittgenstein, thinking does not consist in the operation of mental signs and that the meaning of utterances does not depend on underlying processes of meaning and understanding, but is wholly determined by their *use* or *role* in language games. Finally I suggested that the Tractarian idea of an internal 'pictorial relationship' between language and world is mistaken. First, as both Wittgenstein and Sellars argue, intentional acts and states do not stand in a relation to their content. Second, the necessary relation between intentional phenomena and the objects, events and states of affairs at which they are directed is not explained by an ineffable metaphysical relation between language and mind, on the one hand, and the world, on the other, but, as the later Wittgenstein argues, by the grammatical fact that we may use the expressions we use to specify the content of intentional acts and states to speak about the objects at which they are directed.

Finally, I began the fourth chapter by pointing out that the later Wittgenstein never developed a systematic account of intentional mental phenomena. In particular, he never developed a systematic account of occurrent thinking, which could be said to replace the mistaken account he appears to have presupposed when writing the *Tractatus*. Most of the chapter was then devoted to a discussion of Sellars's account of thinking. I concluded by suggesting that this account could accommodate the objections against the Tractarian account of thought, which I discussed in the second chapter. More specifically, Sellars's analysis of our concept of thought is compatible with the insights (i) that the ascription of

thoughts does not imply that there is a part of our body in which thoughts are located and (ii) that thoughts are not the kind of representational phenomena that can be interpreted.

Throughout my discussion of both Wittgenstein and Sellars I have expressed my own views predominantly on questions of interpretation and have remained mostly silent on matters of philosophical substance. However, I shall conclude this essay by briefly highlighting two potentially significant problems I see with Sellars's account of thinking. My criticisms of these problems are not intended to amount to definitive objections to Sellars's views but are supposed to point towards lacunae in his thinking, which I believe require further exploration. What I say will be Wittgensteinian in spirit but will not rely on anything specific Wittgenstein has said.

First, Sellars's reasons for holding that thought and speech *causally explain* intelligent behaviour strikes me as unconvincing. His argument is of roughly the following form:

(a) Thinking and speaking require the recognition or 'espousal' of norms or principles governing thought and speech.

(b) The recognition or 'espousal' of norms or principles is reflected in 'uniformities of performance' (TC, 216).

(c) *Therefore*, thinkers and speakers display uniformities of performance.

(d) The uniformities of performance displayed by thinkers and speakers can be described in terms of *ceteris paribus* laws.

(e) *Therefore*, the pieces of conduct constituting these uniformities of performance, i.e. individual acts of speech or thought, stand in relations of cause and effect.

I think steps (a) to (c) are valid and would also be endorsed by Wittgenstein. A person will only count as saying or thinking something if she displays certain 'uniformities of

performance'. An utterance, for example, will only count as an expression of intention if the speaker's expressions of intention tend to be associated with behaviour suitable to realize the intentions expressed and tend not to be associated with behaviour apt to thwart them. The problematic step is premise (d); in particular because Sellars thinks that the 'principles' or rules reflected in uniformities of thought and speech are rules of inference expressing the inferential properties of the contents of individual acts of thought and speech (cf. e.g. NAO, 130). Thus, he thinks that when I think that p then I will *ceteris paribus* also think that q, that r and that s, if q, r and s are inferential consequences of p. I am afraid that this idea strikes me as simply implausible. I fear that more often than not when I think a certain thought I do not right away think of all or even some of its inferential consequences.

In a passage already quoted in the preceding chapter Sellars himself seems to provide a reason for rejecting this idea:

> I am not claiming that to *follow* a principle, i.e. act on a principle, is identical with exhibiting a uniformity of performance that accords with the principle. I think that any such idea is radically mistaken. I am merely saying that the espousal of a principle or standard, *whatever else it involves*, is characterized by a uniformity of performance. (TC, 216)

Sellars is right in saying that the idea that acting on a principle or rule is identical with exhibiting a uniformity of performance that accords with it is radically mistaken. But if so, then it is also mistaken to claim that the uniformities of performance characterizing the rule-governed activities of thinking and speaking reflect the norms of inference applying to individual acts of thought and speech to such a degree that we can say that a thought *ceteris paribus* causes its inferential consequences.

The second fundamental problem with Sellars's account of thinking is his suggestion that our concept of thought is analogous to theoretical concepts in the sciences, and,

accordingly the explanation of human behaviour in terms of thoughts analogous to theoretical explanations. Although Sellars does not wish to say that our concept of thought *is* a theoretical concept (cf. EPM, 183), he thinks the similarity to theoretical concepts is close enough to make it illuminating to give an account of it by analogy with the latter and to understand the explanation of human behaviour in terms of thoughts in analogy to theoretical explanations. I believe this is a mistake.

Scientific theories can be replaced by explanatorily more powerful successor theories. If our concept of thought is part of a psychological theory then it should be conceivable that it is abandoned along with the theory in which it has its place; for example, if the common sense theory of thought is replaced by an explanatorily more powerful theory of human behaviour provided by the neurosciences. That something like this is not only conceivable but will or ought to happen has been argued by 'eliminative materialists', most prominently Sellars's own students Paul and Patricia Churchland.[1] According to eliminative materialists our ordinary psychological vocabulary forms a theory of 'folk psychology'. This theory will be replaced and its theoretical vocabulary completely abandoned once the neurosciences have made enough progress to provide an explanatorily more powerful alternative.

Although Sellars never commented on eliminative materialism in his writings, it is quite clear that he would reject it. First, the metaphysics of Jones's theory of thoughts is different from the metaphysics of ordinary scientific theories. Whereas the thoughts of Jones's theories are attributes of persons, which are the ontologically fundamental particulars in what Sellars calls the 'manifest image man-in-the-world' (cf. PSIM, 6-18), the postulates of scientific theories are, according to Sellars, particulars and not attributes of particulars (cf. PSIM, 19). Second, Sellars appears to think that our ordinary

[1] See e.g. Churchland (1981). Both Churchlands studied with Sellars in the 1960s. Paul Churchland wrote his dissertation under Sellars's supervision.

psychological vocabulary is in some sense constitutive of who we are; it is a central part of a conceptual framework 'in terms of which, to use an existentialist turn of phrase, man first encountered himself – which is, of course, when he came to be man' (PSIM, 6). A person, Sellars appears to think, is *essentially* a creature conceiving of herself in terms of our ordinary psychological vocabulary. Hence this vocabulary cannot be abandoned or 'eliminated' (cf. PSIM, 39-40) without man ceasing to be man.

I think Sellars is right about these two points. Thoughts are not particulars but attributes of persons. And it seems true to say that it is a constitutive property of a person that she is able to think of herself as having thoughts, beliefs and desires. It is doubtful that we would think of creature unable to conceive of herself in these terms as a person, at least if we think of the concept of a person as a metaphysical concept which applies to a creature if and only if she is rational. However, as long as one maintains that there is a close analogy between thoughts and theoretical postulates and between psychological explanations and theoretical explanations, invoking these two points as reasons for rejecting eliminative materialism seems *ad hoc.* Eliminative materialists appear to simply think through to its natural conclusion a thought, which originates with Sellars's analogy. And while eliminative materialists are happy to embrace the conclusion that our ordinary psychological explanations will eventually be replaced by explanations provided by the neurosciences, I think this conclusion should really be seen as a *reductio ad absurdum* of the idea that leads to it. Hence, one should reject this idea, i.e. one should reject Sellars's suggestion that thoughts are analogous to theoretical postulates and that ordinary psychological explanations are analogous to theoretical explanations in the sciences.

The idea that our concept of thought is a theoretical concept within a theory of 'folk psychology' is usually considered to be an essential part of Sellars's account of thinking and has been enormously influential in the philosophy of mind ever since the publication

of EPM in 1956. It might therefore seem as though a rejection of this idea would amount to a complete rejection of Sellars's account of thinking and of the mind more generally.

I do not think that this is right. The explanatory role in Sellars's account of the analogy between thoughts and theoretical postulates is actually quite small. Sellars's purpose in comparing the concept of thought to concepts for theoretical entities is (a) to explain the fact that the ability to speak about one's own thoughts presupposes the ability to speak about the thoughts of others (cf. EPM, 183) and (b) to highlight the fact that there is a non-analytic conceptual relation between ascriptions of thoughts and descriptions of behaviour (cf. SM, 69).

It seems to me that Sellars could easily retain insights (a) and (b), even if he insisted that our concept of thought was merely an analogical concept, without making the additional claim that the analogy between acts of thought and acts of speech was the kind of analogy operative in scientific theories, which are formulated in terms of a model. He could still maintain that (a) the ability to speak about one's own thoughts presupposes the ability to speak about the thoughts of others, and (b) that there is a non-analytic conceptual relation between thought and behaviour. Indeed it seems to me that the conceptual relation between thought and behaviour is obscured rather than illuminated by comparing our concept of thinking to a theoretical concept. The fact that scientific theories can be replaced shows that the relation between statements about the behaviour of theoretical objects and statements about observable phenomena, whose truth is evidence for the truth of the former, is not as close as the conceptual relation between ascriptions of thoughts and descriptions of behaviour. The conceptual or 'grammatical' relation between talk about a person's thoughts, beliefs, desires and intentions and talk about her behaviour is far more intimate than the view that our mental concepts are theoretical concepts within a theory of

‘folk psychology’ can account for. Still recognising the closeness of this relation does not mean endorsing logical behaviourism.

Bibliography

Sellars's works in the alphabetical order of their abbreviations:

AAE 1973. Actions and events. *Nous* **7**, 179-202. Reprinted in EPH. References are to the latter appearance.

AE 1963. Abstract Entities. *Review of Metaphysics* 16, 627-7. Reprinted in SR. References are to the latter appearance.

APM 1949. Aristotelian philosophies of mind. In: R. W. SELLARS, M. FARBER and V. J. McGILL, eds. *Philosophy for the Future*. New York: Macmillan Co. Quoted from PPPW.

BBK 1960. Being and Being Known. *Proceedings of the American Catholic Philosophical Association*, 28-49. Reprinted in SPR. References are to the latter appearance.

CIL 1948. Concepts as involving laws and inconceivable without them. *Philosophy of Science* **15**, 287-315. Reprinted in PPPW. References are to the latter appearance.

CSCA 1961-62. Correspondence between Wilfrid Sellars and Hector-Neri Castañeda. CHRUCKY, A., ed. Published on the internet: *www.ditext.com/sellars/corr.html*.

EAE 1963. Empiricism and Abstract Entities. in: SCHILP, P. A., ed. *The Philosophy of Rudolf Carnap.* Open Court, 431-68. Reprinted in EPH. References are to the latter appearance.

EPH 1974. *Essays in Philosophy and Its History*. Dordrecht: Reidel, 1974.

EPM 1956. Empiricism and the Philosophy of Mind. In: FEIGL, H. and SCRIVEN, M., eds. *Minnesota Studies in The Philosophy of Science, Vol. I: The Foundations of Science and the Concepts of Psychology and Psychoanalysis*. Minneapolis: University of Minnesota Press, 253-329. Reprinted in revised form in SPR. References are to the latter Appearance

FMPP 1981. Foundations for a metaphysics of pure process. *The Monist* **64**, 3-90.

IKTE 1978. The role of imagination in Kant's theory of experience. In: JOHNSTON, H. W. *Categories: a colloquium.* University Park, PA: Pennsylvania State University Press, 231-45. Reprinted in SR. References are to the latter appearance.

IILO 1963. Imperatives, intentions, and the logic of 'ought'. In: CASTANEDA, H. and NAKHNIKIAN, G., eds. *Morality and the Language of Conduct*. Wayne State University Press, 159-214.

IM 1953. Inference and meaning. *Mind* 62, 313-38. Reprinted in SR. References are to the latter appearance.

ITSA 1953. Is there a synthetic a priori? *Philosophy of Science* **20**, 121-38. Reprinted in a revised form in his SPR. References are to the latter appearance.

IV 1964. Induction as vindication. *Philosophy of Science* **31**, 197-231. Reprinted in EPH. References are to the latter appearance.

LRB 1949. Language, rules and behavior. In: HOOK, S., ed. *John Dewey: Philosopher of Science and Freedom*. New York: The Dial Press, 289-315. Reprinted in PPPW. References are to the latter appearance.

LTC 1969. Language as thought and as communication. *Philosophy and Phenomenological Research* ***29***, 506-27. Reprinted in SR. References are to the latter appearance.

MEV 1981 Mental events. *Philosophical Studies* **39**, 325-45. Reprinted in SR. References are to the latter appearance.

MFC 1974. Meaning as functional classification. *Synthese* **27**, 417-37. Reprinted in SR. References are to the latter appearance.

NAO 1980. *Naturalism and Ontology*. Reseda, California: Ridgeview Publishing.

NS 1962. Naming and Saying. *Philosophy of Science* **29**, 7-26. Reprinted in SPR. References are to the latter appearance.

PH 1963. Phenomenalism. In: SPR, 60-105.

PPPW SICHA, J. F., ed, 1980. *Pure pragmatics and possible worlds: the early essays of Wilfrid Sellars*. Reseda, California: Ridgeview Publishing Company.

PSIM 1962. Philosophy and the scientific image of man. In: COLODNY, R., ed. *Frontiers of science and philosophy*. Pittsburgh: University of Pittsburgh Press, 35-78. Reprinted in SPR. References are to the latter appearance.

SK 1975. The Structure of Knowledge: (1) Perception; (2) Minds; (3) Epistemic Principles. In: H. CASTANEDA, ed. *Action, knowledge and reality: studies in honor of Wilfrid Sellars*. Indianapolis: Bobbs-Merrill, 295-347

SM 1967. *Science and metaphysics: variations on Kantian themes*. London: Routledge and Kegan Paul.

SPR 1963. *Science, perception and reality.* London: Routledge and Kegan Paul.

SR BRANDOM, R. B., SCHARP, K, eds. 2007. *In the space of reasons: selected essays of Wilfrid Sellars*. Cambridge, Mass.: Harvard University Press.

SRLG 1954. Some reflections on language games. *Philosophy of Science* **21**(3), 204-28. Reprinted in revised form in SPR. References are to the latter version.

SSOP 1982. Sensa or sensings: reflections on the ontology of perception. *Philosophical Studies* **41**, 83-111.

TA 1966. Thought and action. In: K. LEHRER, ed. *Freedom and Determinism*. New York: Random House, 105-39.

TC 1962. Truth and 'correspondence'. *The Journal of Philosophy* **59**, 29-56. Reprinted in SPR. References are to the latter appearance.

VR 1976. Volitions re-affirmed. In: M. BRAND and D. WALTON, eds. *Action Theory.* Dordrecht: Reidel, 47-66.

WSNDL 2009. *Wilfrid Sellars Notre Dame Lectures: 1969-1986 – The Bootleg Version*. Edited with an introduction by P. AMARAL. Published online: http://www.ditext.com/amaral/wsndl.pdf

Wittgenstein's works in the alphabetical order of their abbreviations:

References are to page numbers unless otherwise indicated.

BB 1974. *The blue and brown books: preliminary studies for the "Philosophical Investigations"*. 2nd edn. Oxford: Basil Blackwell.

CL 2008. *Wittgenstein in Cambridge : letters and documents 1911-1951*. Oxford: Blackwell.

LW I 1992. *Last writings on the philosophy of psychology - volume I*. Oxford: Basil Blackwell.

LWL 1980. *Wittgenstein's lectures, Cambridge, 1930-1932 : from the notes of John King and Desmond Lee*. Oxford: Basil Blackwell.

NB 1979. *Notebooks, 1914-1916*. Oxford: Blackwell.

OC 1975. *On certainty*. Oxford: Blackwell.

PG 1974. *Philosophical grammar*. Oxford: Basil Blackwell.

PI 2009. *Philosophical investigations*. 4 edn. Chichester: Wiley-Blackwell.

PPF 2009. *Philosophy of psychology – a fragment*. Part two of PI.

PR 1975. *Philosophical remarks*. Oxford: Basil Blackwell.

RFM 1978. *Remarks on the foundations of mathematics*. Oxford: Blackwell.

TLP 2001. *Tractatus logico-philosophicus*. London: Routledge.

WVC *Wittgenstein and the Vienna Circle: conversations.* Shorthand notes recorded by F. Waismann and edited by B. McGuinnes. Oxford: Basil Blackwell.

Z 1981. *Zettel.* Oxford: Blackwell.

Works by other authors:

AMMERELLER, E., 2001a. Die Abbildende Beziehung. Zum Problem der Intentionalität im Tractatus. In: W. VOSSENKUHL, ed, *Ludwig Wittgenstein, Tractatus logico Philosophicus,* 111-140.

AMMERELLER, E., 2001b. Wittgenstein on intentionality. In: H. GLOCK, ed, *Wittgenstein: a critical reader.* Oxford: Blackwell Publishers, 59-93.

BAKER, G. P. and HACKER, P. M. S., 1984. *Scepticism, rules and language.* Oxford: Blackwell.

BAKER, G. P. and HACKER, P. M. S., 2005a. *Wittgenstein: understanding and meaning - part I: essays.* Oxford: Blackwell Publishing.

BAKER, G. P. and HACKER, P. M. S., 2005b. *Wittgenstein: understanding and meaning - part II: exegesis §§1-184.* Oxford: Blackwell Publishing.

BAKER, G. P. and HACKER, P. M. S., 2009. *Wittgenstein: rules, grammar, and necessity: essays and exegesis of 185-242.* Oxford: Wiley-Blackwell.

BLOCK, I., 1981. *Perspectives on the philosophy of Wittgenstein.* Oxford: Basil Blackwell.

BRADLEY, F. H., 1893. *Appearance and reality: a metaphysical essay.* London: Swan Sonnenschein.

BRANDOM, R., 2002. The centrality of Sellars's two-ply account of observation to the arguments of 'Empiricism and the Philosophy of Mind'. In: R. BRANDOM, *Tales of the mighty dead : historical essays in the metaphysics of intentionality.* Cambridge, Mass: Harvard University Press, 348-367.

BUDD, M., 1989. *Wittgenstein's philosophy of psychology.* London: Routledge.

CHILD, W., 2011. *Wittgenstein.* London: Routledge.

CHURCHLAND, P. M., 1981. Eliminative materialism and the propositional attitudes. *The Journal of Philosophy,* **78**(2), 67-90.

COATES, P. and DEVRIES, W. A., 2009. Brandom's two-ply error. In: W. A. DEVRIES, ed, *Empiricism, perceptual knowledge, normativity, and realism : essays on Wilfrid Sellars.* Oxford: Oxford University Press, 131-145.

CRANE, T., 2001. *Elements of mind*. Oxford: Oxford University Press.

CRANE, T., forthcoming. Wittgenstein and intentionality. Available on the internet: *http://web.mac.com/cranetim/Tims_website/Online_papers_files/Wittgenstein%20and%20intentionality.pdf*

DAVIDSON, D., 2001. *Essays on actions and events.* Oxford: Clarendon Press.

DEVRIES, W. A., 2008. McDowell, Sellars, and sense impressions. In: J. LINDGAARD, ed, *Experience, norm, and nature.* Oxford: Blackwell Publishing.

DEVRIES, W. A., 2005. *Wilfrid Sellars.* Chesham: Acumen Publishing.

DIAMOND, C., 2002. Peter Winch on the Tractatus and the unity of Wittgenstein's philosophy. In: A. PICHLER and S. SÄÄTELÄ, eds, *Wittgenstein: the philosopher and his works.* Frankfurt;: Ontos, 141-171.

DIAMOND, C., 1988. Throwing away the ladder. *Philosophy,* **63**(243), pp. pp. 5-27.

DIAMOND, C., 1989. Rules: looking in the right place. In: D.Z. PHILLIPS and P. WINCH, eds, Basingstoke: Macmillan, 12-34.

DIAMOND, C., 1991. *The realistic spirit: Wittgenstein, philosophy, and the mind.* Cambridge, Mass: MIT Press.

DUMMETT, M., 1991. *Frege and other philosophers.* Oxford: Clarendon Press.

FANN, K.T., 1967. *Ludwig Wittgenstein : the man and his philosophy.* New York: Dell.

FODOR, J.A., 1975. *The language of thought.* Cambridge, Mass.: Harvard University Press.

FODOR, J.A., 2008. *LOT 2: the language of thought revisited.* Oxford: Oxford University Press.

FOGELIN, R.J., 1976. *Wittgenstein.* London: Routledge & Kegan Paul.

FREGE, G., 1918-19. The thought. In: M. BEANEY, ed, 1997. *The Frege reader.* Oxford: Blackwell. Cited after the original pagination.

FREGE, G., 1989. Briefe an Ludwig Wittgenstein aus den Jahren 1914-1920. Edited by J. BERGER and A. JANIK. *Grazer Philosophische Studien,* **33**, pp. 5-33.

GEACH, P. T., 1971. *Mental acts: their content and their objects.* 2 edn. London: Routledge & Kegan Paul.

GLOCK, H., 2006. Thought, language, and animals. *Grazer Philosophische Studien,* **71**(1), 139-160.

GLOCK, H., 1996. *A Wittgenstein dictionary.* Oxford: Blackwell Publishers.

GOETHE, J.W., 1857. *Hermann and Dorothea.* London: War and Lock.

GRICE, H.P., 1989. *Studies in the way of words.* Cambridge, Mass.: Harvard University Press.

HACKER, P. M. S., 1986. *Insight and illusion: themes in the philosophy of Wittgenstein.* Oxford: Clarendon Press.

HACKER, P. M. S., 1990. *Wittgenstein: meaning and mind.* Oxford: Basil Blackwell.

HACKER, P. M. S., 1996. *Wittgenstein: mind and will.* Oxford: Basil Blackwell.

HACKER, P. M. S., 2001. *Wittgenstein: connections and controversies.* Oxford: Clarendon Press.

HACKER, P. M. S., 2007. *Human nature: the categorial framework.* Oxford: Blackwell.

HOLT, E. B., MARVIN, W. T., MONTAGUE, W. P., PERRY, R. B., PITKIN, W. B. and SPAULDING, E. G., 1912. *The new realism: cooperative studies in philosophy.* New York: Macmillan.

HYMAN, J., 2003. Pains and Places. *Philosophy,* **78** (303), 5-24.

KENNY, A., 1984. *The legacy of Wittgenstein.* Oxford: Blackwell.

KENNY, A., 2009. Cognitive Scientism. In: H. GLOCK and J. HYMAN, eds, *Wittgenstein and analytic philosophy: essays for P. M. S. Hacker.* Oxford: Oxford University Press, 250-262.

KRIPKE, S. A., 1982. *Wittgenstein on rules and private language: an elementary exposition.* Oxford: Blackwell.

KUKLICK, B., 2002. *A history of philosophy in America, 1720-2000.* Oxford: Clarendon Press.

KÜNNE, W., 2009. Wittgenstein and Frege's Logical Investigations. In: H. GLOCK and J.HYMAN, eds, *Wittgenstein and analytic philosophy: essays for P. M. S. Hacker.* Oxford: Oxford University Press, 26-62.

MALCOLM, N., 1986. *Nothing is hidden : Wittgenstein's criticism of his early thought.* Oxford: Blackwell.

MALCOLM, N., 1989. Wittgenstein on Language and Rules. *Philosophy,* **64**(247), 5-28.

MCDOWELL, J., 1984. Wittgenstein on following a rule. *Synthese,* **58**(3), 325-363.

MCDOWELL, J., 1998. Having the world in view: Sellars, Kant, and intentionality. *The Journal of Philosophy,* **95**(9), 431-491.

MCDOWELL, J., 2009. Why is Sellars's essay called "*Empiricism* and the philosophy of mind". In: J. MCDOWELL, ed, *Having the world in view: essays on Kant, Hegel, and Sellars.* Cambridge, Mass.: Harvard University Press, 221-238.

MCGINN, M., 2006. *Elucidating the Tractatus: Wittgenstein's early philosophy of logic and language.* Oxford: Clarendon Press.

MCGINN, M., 2010. Wittgenstein and internal relations. *European Journal of Philosophy,* **18**(4), 495-509.

MILL, J. S., 1889. *An examination of Sir William Hamilton's philosophy.* 6 edn. London: Longmans, Green and Co.

MOORE, G. E., 1919. External and internal relations. *Proceedings of the Aristotelian Society,* **20**, pp. pp. 40-62.

ROSENBERG, J. F., 2007. *Wilfrid Sellars: fusing the images.* New York: Oxford University Press Inc.

RUNDLE, B., 2001. Meaning and Understanding. In: H. GLOCK, ed, *Wittgenstein: a critical reader.* Oxford: Blackwell Publishers, 59-93.

RUSSELL, B., 1921. *The analysis of mind.* London: Allen & Unwin.

SCHROEDER, S., 2006. *Wittgenstein: the way out of the fly-bottle.* Cambridge: Polity Press.

SCHULTE, J., 2010. Some remarks on 'Wittgenstein's Early Philosophy of Mind'. In: J. COTTINGHAM and P. M. S. HACKER, eds, *Mind, method, and morality: essays in honour of Anthony Kenny.* Oxford: Oxford University Press, 330-351.

SELLARS, R. W., 1969. *Reflections on American philosophy from within.* Notre Dame: University of Notre Dame Press.

SELLARS, R. W., 1916. *Critical realism: a study of the nature and conditions of knowledge.* New York: McNally & Company.

STROUD, B., 1996. Mind, meaning, and practice. In: H. D. SLUGA and D. G. STERN, eds, *The Cambridge companion to Wittgenstein.* Cambridge: Cambridge University Press.

WINCH, P., 1987. Language, thought and world in Wittgenstein's Tractatus. In: P. WINCH, *Trying to make sense.* Oxford: Basil Blackwell, 3-18.

WRIGHT, C., 2001. *Rails to infinity: essays on themes from Wittgenstein's Philosophical Investigations.* Cambridge, Mass.: Harvard University Press.

WRIGHT, C., 1980. *Wittgenstein on the foundations of mathematics.* London: Duckworth.