Wittgenstein and Davidson on Language, Thought, and Action

Edited by Claudine Verheggen
Wittgenstein and Davidson are two of the most influential and controversial figures of twentieth-century philosophy. However, whereas Wittgenstein is often regarded as a deflationary philosopher, Davidson is considered to be a theory builder and systematic philosopher par excellence. Consequently, little work has been devoted to comparing their philosophies with each other. In this volume of new essays, leading scholars show that in fact there is much that the two share. By focusing on the similarities between Wittgenstein and Davidson, their essays present compelling defences of their views and develop more coherent and convincing approaches than either philosopher was able to propose on his own. They show how philosophically fruitful and constructive reflection on Wittgenstein and Davidson continues to be, and how relevant the writings of both philosophers are to current debates in philosophy of mind, language, and action.

Claudine Verheggen is Professor in the Department of Philosophy at York University in Toronto. She has published numerous articles on issues in mind and language, and she is co-author of Donald Davidson’s Triangulation Argument: A Philosophical Inquiry (2016).
Wittgenstein and Davidson on Language, Thought, and Action

Edited by
Claudine Verheggen
York University, Toronto
In memory of Sue Larson
Contents

List of Contributors ix
Acknowledgements x

Introduction 1

Claudine Verheggen

1 Holism in Action 8

Robert H. Myers

2 Davidson’s Wittgensteinian View of Meaning 28

Paul Horwich

3 Davidson and Wittgenstein 46

A Homeric Struggle?

Åsa Wikforss

4 Rule-Following and Charity 69

Wittgenstein and Davidson on Meaning Determination

Kathrin Glüer

5 Davidson’s Treatment of Wittgenstein’s Rule-Following Paradox 97

Claudine Verheggen

6 Davidson and Wittgenstein on Meaning and Understanding 123

Barry Stroud

7 The Search for “The Essence of Human Language” in 139

Wittgenstein and Davidson

Jason Bridges

8 Wittgenstein and Davidson on Animal Minds 159

Hans-Johann Glock

9 Wittgenstein and Davidson on First-Person Authority and the 186

Univocality of Mental Terms

William Child
viii Contents

10 The Harmony of Thought and Reality
   *Wittgenstein and Davidson vs McDowell*
   TIM THORNTON
   205

11 Davidson, Russell, and Wittgenstein on the Problem of Predication
   JOSÉ L. ZALABARDO
   226

Index
   250
Contributors

Jason Bridges University of Chicago
William Child University of Oxford
Hans-Johann Glock University of Zürich
Kathrin Glüer Stockholm University
Paul Horwich New York University
Robert H. Myers York University, Toronto
Barry Stroud University of California, Berkeley
Tim Thornton University of Central Lancashire
Claudine Verheggen York University, Toronto
Åsa Wikforss Stockholm University
José L. Zalabardo University College London
I am very grateful to all the authors of this book for their contributions. I also want to thank the participants in a workshop on Wittgenstein and Davidson, which I organized in Toronto in March 2015, and York University for sponsoring it. Preparation of the volume was made possible in part by a research fellowship from the York University Faculty Association. Thanks are also due to my editor, Hilary Gaskin, for her support of the project and her patience in seeing it realized. Finally, I am very grateful to Olivia Sultanescu for her help at every stage of the project, from the initial review of the chapters to the preparation of the index.

I dedicate this book to the memory of Sue Larson, who first made me realize, through the numerous conversations I had with her in the 1990s, how much Wittgenstein and Davidson had in common, and who thus was the first inspiration for this book.
Introduction

Claudine Verheggen

I’m not interested in constructing a building so much as in having a perspicuous view of the foundations of possible buildings. (Wittgenstein 1980, 7)

Maybe those long hours I spent years ago admiring and puzzling over the *Investigations* were not spent in vain. (Davidson 1999, 286)

Ludwig Wittgenstein and Donald Davidson are two of the most formidable figures of twentieth century philosophy, equally influential, equally controversial. It would be hard to overstate the importance and the extent of the influence they have had and continue to have on contemporary philosophers working in a great many areas of philosophy. Yet, they are typically taken to be engaged in two radically different ways of doing philosophy. On the one hand, Wittgenstein is widely understood to be a deflationary philosopher, who recommends that philosophical problems be dissolved rather than solved, and thus that no constructive philosophical thesis be advanced. Davidson, on the other hand, is widely acknowledged to be a theory builder, a systematic philosopher par excellence, whose views about the nature of language and thought are intended to have consequences for most areas of philosophy. Given these alleged differences, it is perhaps no wonder that, though the writings of each philosopher have generated volumes of commentaries, some of which compare them to other leading philosophers, there has so far been no book comparing Wittgenstein and Davidson.

The purpose of this volume is to demonstrate that this is a serious lacuna, resulting, at least in part, from a serious misreading of both philosophers. There is in fact much that Wittgenstein and Davidson share (as Davidson himself was increasingly recognizing in the last fifteen years of his life) and much to be learnt from investigating them side by side. In one way or another, the chapters in this volume address these commonalities or “family resemblances”, as it is tempting to call them. Some of the chapters aim at establishing these resemblances, thereby reinforcing particular claims being advanced by both philosophers. Some of the chapters use the arguments of one philosopher to improve on the views of the other. All aim either at presenting more compelling defences of the views of one or the other, or both, of these philosophers, or at developing
more coherent and convincing versions of their views than either philosopher put forward on his own. Thus the volume is not primarily exegetical, though the authors have also taken great care to get the texts right. Rather, the volume demonstrates how philosophically fruitful and constructive reflection on Wittgenstein and Davidson continues to be, and how relevant the writings of both philosophers are to current debates.

Not only do family resemblances between Wittgenstein and Davidson get established and elaborated upon by most authors in the volume, some family resemblances are also to be found among the authors themselves. In addition, interestingly, certain family resemblances between Wittgenstein and Davidson asserted by some authors are denied by others. The volume thus reveals, as Wittgenstein would have put it, “a complicated network of similarities overlapping and criss-crossing: similarities in the large and in the small” (Wittgenstein 1953, §66). In what follows, I shall draw an outline of the most significant of these similarities, describing thereby the structure and main themes of the book.

It starts where Davidson started, with his causal theory of action introduced in 1963. Interest in this theory has recently been revived along with interest in one of the works Davidson was reacting to, viz., *Intention*, by the famously Wittgensteinian Elizabeth Anscombe. Two worries that have always plagued Davidson’s theory concern the problem of deviant causal chains and the problem of weakness of the will. Robert Myers argues that these worries can to a large extent be overcome once the theory is outfitted with a more thoroughly holistic account of pro-attitudes, instead of the standard Humean account usually attributed to Davidson. The problems of deviant causal chains and weakness of the will can then be seen as problems of detail, as Davidson always suggested, and the differences between Anscombe’s version of the non-causal theory and Davidson’s version of the causal theory can be seen as significantly reduced, as Davidson always intended.

The chapters that come next address one or the other facet of Wittgenstein’s and Davidson’s discussion of natural language and its relation to thought and to reality. A primary interest in these topics was shared by the two philosophers throughout their lives. This focus was true of both the early Wittgenstein of the *Tractatus* and the later Wittgenstein of *Philosophical Investigations*. And it was true of both Davidson’s early views on radical interpretation, the interpretation from scratch of a person’s language and thoughts, and his later views on triangulation, the idea of two creatures responding simultaneously to each other and to the world they share. The family resemblances emphasized by all authors are between the later Wittgenstein and Davidson. But different authors focus on different periods of Davidson’s work. What they all recognize is that Wittgenstein and Davidson both see an important connection between the way we use our
words and what we mean by them, but they exploit this connection in importantly different, though overlapping, ways.

Paul Horwich is in fact reticent to attribute to Davidson a use-theory of meaning, for he thinks that Davidson would take it to be incompatible with his truth-theoretic view of meaning, in particular, his truth-conditional account of compositionality. But he also thinks that Davidson is wrong to insist on this truth-conditional account—a superior, deflationary account is in sight, which is compatible with a Wittgensteinian use-theory of meaning. Moreover, it would be easy for Davidson to adopt the deflationary account, since his own account of interpretation possesses all the seeds of a use-theory of meaning. Indeed, Horwich sees an important parallel between Davidson’s idea that interpretation is guided by the principle of charity, and thus based on agreement between interpreter and interpretee, and Wittgenstein’s idea that interpretation is guided by use, and thus based on the shared practice of accepting sentences. This “transformation” of Davidson into Wittgenstein would have him abandon his non-deflationist theory of truth and embrace an explanation of truth in terms of meaning, rather than the other way around, as Davidson himself would have it. But is such a transformation necessary?

Some philosophers think not. One of these is Åsa Wikforss, who argues that a truth-conditional theory of meaning is perfectly compatible with a meta-semantic theory that puts communication, and therefore use, at its centre. Thus, according to her, the view of a use-theorist like Wittgenstein is perfectly compatible with the view of a formal semantics theorist like Davidson. Indeed Davidson defends both views. The key is to recognize that the truth-theory is supposed to be the answer to the question what one could know that would enable one to interpret another person’s words, which is different from the meta-semantic question how one could know it, to which radical interpretation is the answer. It is at this foundational, meta-semantic level that Wittgenstein and Davidson have much in common. Wikforss argues that Davidson’s principle of meaning determination, the principle of charity, is similar to Wittgenstein’s, since both recommend that facts about use be mapped onto meanings and semantic correctness conditions, and thus both emphasize the need for agreement in beliefs and judgments for there to be communication. As she says, “the device of the radical interpreter is meant precisely to illustrate how use determines meaning”—on this much she and Horwich agree. The discussion of meaning determination brings out, according to Wikforss, another similarity between Wittgenstein and Davidson. Contra many commentators, she maintains that Wittgenstein did not believe that meaning is determined by conventions, just as Davidson did not. However, Wikforss ends up remarking, Wittgenstein and Davidson’s insistence that meaning is determined by use entails that meanings cannot be “perfectly objective”.

This is also the conclusion reached by Kathrin Glüer, who finds in Davidson’s writings an answer to Wittgenstein’s rule-following paradox,
a paradox which seems to pose a serious threat to the idea that linguistic expressions can ever be meaningful. Glüer argues that it is thinking of meaning as use, or as determined by use, that leads Wittgenstein to the paradox, for, she asks, if meaning is use, how could meaning determine “a potential infinity of objectively correct applications?” She then considers the account of meaning determination Davidson offers in his writings on radical interpretation as a possible answer to Wittgenstein’s paradox. Radical interpretation is based on a premise shared by Wittgenstein and Davidson, viz., that language is essentially public. Thus, according to the account of radical interpretation, meanings are determined in two steps: first by detecting the interpretee’s attitudes of holding true uninterpreted sentences in given circumstances, and then by having these attitudes determine meanings via the principle of charity. The interpreter, as Glüer emphasizes, plays an essential role in determining meaning, and thus only what speakers are “sensitive” to is eligible to be meant. The problem of objectivity seems therefore to remain with us. Does it, however?

Claudine Verheggen’s paper might well be read as suggesting that it does not, not because Davidson succeeds in solving it but, rather, because, following Wittgenstein, he succeeds in dissolving it. Verheggen, too, tackles Wittgenstein’s rule-following paradox and argues that the considerations that lead him to the paradox are similar to the considerations that lead Davidson to develop his account of triangulation and thus the claim that one’s possession of language and thoughts requires one to interact linguistically with others and the world around them. She further argues that, with Wittgenstein’s help, Davidson is in a position to vindicate the premise that language is essentially public, which in turn solidifies Davidson’s approach to meaning and his account of meaning determination. Key to the argument is the claim that, according both to Wittgenstein and to Davidson, when reflecting on the connection between meaning and use, we should not conceive of use in non-semantic terms, but we should think of it as the meaningful use of words by people engaged in communication. Thus Verheggen argues that Wittgenstein and Davidson are both fierce non-reductionists, though she also believes that their non-reductionism is compatible with non-quietism.

Though he does not use the word ‘non-reductionism’, Barry Stroud argues that this is the fundamental idea shared by Wittgenstein and Davidson. Stroud stresses the importance they both bestow on ostension in learning language and for the words learnt to mean what they do. He also stresses the claim that for neither of them is ostension sufficient for language learning. What is needed, in addition, is mastery of the grammatical structure of language. This is a mastery that can be described, following Davidson, by a truth-conditional theory of meaning, which, according to Stroud, finds its equivalent in Wittgenstein’s talk of a speaker’s grasp of the “place in language, in grammar” that is assigned to different kinds of word. Thus Stroud, too, sees no conflict between a truth-conditional theory of meaning
and the idea that use and meaning are essentially connected. But it is crucial to
keep in mind that the relevant use can only be described intensionally, which is to
say that, ultimately, we can account for speakers meaning what they do by their
words only by saying what the words they use are used to mean.

Jason Bridges further investigates the themes developed in the two pre-
vious chapters with an in-depth examination of Wittgenstein’s search for the
“essence of human language”. Bridges construes this as the search for an
explanation of linguistic meaning that would be “external” to our under-
standing of meaning in use. And he articulates the many ways in which
Wittgenstein takes this search to be futile. There is no way we could achieve
such an external perspective, e.g., by positing “meanings” as external objects,
or mental processes alleged to give “life” to otherwise “dead” signs. Nor
would appeal to the notion of reference help, for our understanding of what
words refer to is itself dependent on our understanding of what speakers use
their words to mean. Contra the two previous authors, Bridges is not con-
vincing, however, that Davidson does not part company with Wittgenstein on
this issue. Though he grants many similarities between Wittgenstein’s and
Davidson’s approach to meaning and acknowledges Davidson’s non-
reductionist aspirations, he finds in Davidson’s account of radical interpreta-
tion a desire still to achieve an external perspective on meaning.

With the next two chapters we turn our attention to the relation between
language and thought. Hans-Johann Glock addresses the question whether
there can be thought without language and, in particular, whether non-
linguistic animals can have thoughts. He highlights the many similarities that
form the background of Wittgenstein’s and Davidson’s views on the topic,
centrally, their third-person approach to mental phenomena and their insistence
that thoughts be manifestable in behaviour. He then argues that Davidson is
wrong in maintaining that possession of thoughts of any kind requires language
and its manifestation in linguistic behaviour. A better, more moderate position
can be found in Wittgenstein, according to whom non-linguistic animals can
have “beliefs, desires and intentions of a simple kind, namely, those that can be
expressed in non-linguistic behaviour.” In particular, non-linguistic animals
can have beliefs about the perceptible features of their environment, which can
be attributed on the basis of their reactions to their environment. However, the
simple beliefs that are attributed are “neither intensional nor conceptual nor
holistic in the way attributions of the same thoughts to linguistic subjects are.”
On this Wittgenstein can be seen to agree with Davidson. Another similarity
highlighted by Glock concerns Wittgenstein’s and Davidson’s views of first-
and third-person ascriptions of mental predicates. This is the focus of the
next chapter.

William Child brings out the essential connection between use and mean-
ing acknowledged by Wittgenstein and Davidson by focusing on their
accounts of mental terms and concepts and, in particular, on the questions whether, and if so how, mental terms can have the same meaning in their first-person and third-person uses, since we ascribe mental predicates to others on the basis of their behaviour and we ascribe them to ourselves without evidence. Davidson maintained that Wittgenstein did not even try to meet the explanatory challenge prompted by the asymmetry. Child, however, retorts that Davidson’s explanation is itself less than adequate, coming down, as it does, to regarding people’s self-ascription of mental properties without reference to their behaviour as a “basic, unanalysable fact”. Child then goes on to argue that a satisfactory explanation can in fact be uncovered in Wittgenstein’s writings, an explanation, moreover, that Davidson might have found congenial. The key is to pay attention to first-person and third-person uses of mental terms and to realize that these uses are mutually interdependent. Specifically, on the one hand, we teach children to replace their non-linguistic expressions of sensations with linguistic expressions in circumstances in which we have ascribed the sensations to them on the basis of their behaviour. On the other hand, in the case of many mental properties, we do not ascribe them to others unless they can apply them to themselves without reference to their behaviour. Importantly, contra many commentators on Wittgenstein, though not all of those contributing to this volume, Child thus attributes to Wittgenstein the non-quietist view that there are philosophical questions to which substantive answers can be given.

Tim Thornton, however, is one of those who emphasize the “therapeutic” side of Wittgenstein and thus the idea that philosophical problems are better dissolved than solved. He thinks, moreover, that Davidson can be understood as agreeing with Wittgenstein, at least when it comes to the question “how mind can impact the world, or the world can impact on the mind”, and that together they can address this question in a more satisfactory, and more therapeutic, way than John McDowell, who takes himself to be a follower of Wittgenstein in his therapeutic endeavours. Thornton argues that McDowell’s account of the lack of a gap, or the “harmony”, between thought and reality “has increasingly taken the shape of a philosophical theory of perception of decreasing intuitiveness”, not to mention its apparent commitment to idealism. According to Thornton, Wittgenstein’s and Davidson’s accounts fare much better, stressing, as they both do, the role of language in shaping the mind, and, as Davidson in particular does, the role played by the world in our possession of language and thoughts. Once these roles are given their proper due, the question of contact between mind and world becomes obsolete.

José Zalabardo, too, maintains that Davidson did not think that an explanation of the relation between subjects and the world could be obtained. He reaches this conclusion while considering the early writings of Wittgenstein in the Tractatus, and Davidson’s last writings, posthumously published as Truth
and Predication. Zalabardo’s focus is on the problem of predication both in its metaphysical form, “How are particulars related to universals?” and in its semantic form, “How are names or other singular terms related to predicates?” (Davidson 2005, 77). He chastises Davidson for not considering the debates Wittgenstein and Russell were having on this issue in the second decade of the twentieth century, and he proceeds to offer an extensive discussion of these exchanges. He ends with a presentation of Davidson’s own solution to the semantic version of the problem of predication. Like most authors in the volume, Zalabardo emphasizes the role played by linguistic use in explaining linguistic meaning. Like many, he does not think that the connection between use and meaning is in conflict with a truth-conditional theory of meaning. The book thus ends by reaffirming one of the main family resemblances between Wittgenstein and Davidson and by reinforcing one of the main family resemblances to be found among its contributors.

On the whole, the volume forcefully displays how engaged with contemporary issues Wittgenstein and Davidson can be understood to be and how fruitfully contemporary philosophers are engaging with them both. I believe that, to a large extent, it also demonstrates that Wittgenstein and Davidson can both be regarded as systematic philosophers, and that constructive philosophical theorizing is compatible with, in Wittgenstein’s famous words, “leaving everything as it is” (Wittgenstein 1953, §124).

References

1 Holism in Action

Robert H. Myers

In his landmark paper, “Actions, Reasons, and Causes” (1963), Donald Davidson argued against a wide range of philosophers whom he regarded as followers of Wittgenstein, including, perhaps most notably, Elizabeth Anscombe, that reasons for actions are causes of actions, and that rationalizing explanation is a species of causal explanation. By “reasons” he of course meant, not normative reasons, that is, considerations counting in favor of actions, but motivating reasons, that is, psychological states appropriately explanatory of actions. Davidson’s thesis was that motivating reasons explain actions by rationalizing them, by revealing what agents intend to accomplish in performing them, and that this requires that they be causes of those actions. Famously, he stated this thesis as two claims about what he called primary reasons for actions: (1) “In order to understand how a reason of any kind rationalizes an action it is necessary and sufficient that we see, at least in essential outline, how to construct a primary reason.” (2) “The primary reason for an action is its cause” (1963, 4). As he quickly elaborated: “$R$ is a primary reason why an agent performed the action $A$ under the description $d$ only if $R$ consists of a pro-attitude of the agent towards actions with a certain property, and a belief of the agent that $A$, under the description $d$, has that property” (1963, 5).

Almost immediately, however, Davidson found himself having to acknowledge that any causal account of action along these lines faces two very difficult problems. For one thing, there is, downstream from the primary reason, the problem of deviant causal chains; it is not enough for the primary reason to cause the action, it must cause it in the right way. For another thing, there are, upstream from the primary reason, equally vexing problems concerning the role to be played by normative judgment; it is not enough for the primary reason to cause the action in the right way, it must have been generated in the right way itself. Still, while he simply despaired (1973, 79) of solving the first of these problems, and always remained alive to the difficulty of the second, Davidson maintained throughout his career that both should be regarded as problems of detail – that is to say, not as problems that call into question the basic idea that primary reasons for actions are causes of those actions, but as problems that emerge when one takes up this idea and attempts to construct a complete
account of intentional action on its basis. His thought seems to have been that, since complete accounts in philosophy are invariably difficult to construct, no special occasion for concern exists here.

This assessment of the situation was widely accepted for many years, and remains popular even today. Recently, however, there has been resurgent interest in Anscombe’s broadly Wittgensteinian approach to these issues, and, with it, in the possibility that Davidson’s difficulties are in fact fatal to his view. As I explain in Section 1, I agree that these difficulties are fatal to Davidson’s view as it is often interpreted and as Davidson himself usually presented it in his classic papers on philosophy of action. But I believe that, on a different interpretation, Davidson’s view fares better, indeed well enough to justify treating these difficulties as problems of detail. Furthermore, I believe that this different interpretation is the one closest to Davidson’s real meaning, and in fact that we can find many signs of it even in the classic papers, although signs that admittedly are obscured by other remarks seeming to invite the more familiar interpretation. As I argue in Section 1, the pivotal question here is what Davidson took pro-attitudes to be – a question about which, at least in his earliest papers on philosophy of action, he sent mixed messages, sometimes sounding as if he advocated what I shall call the standard view, sometimes sounding as if he meant to be embracing a more holistic account.

According to what I shall call the standard view of what pro-attitudes are, a pro-attitude towards actions with a certain property is, at bottom, a disposition to perform actions one believes likely to have that property.\(^1\) If this was Davidson’s view of the matter, then I agree that the difficulties confronting his causal theory of action are in fact fatal. But given the early mixed messages, we should be careful not to assume too quickly that this ever was his view of the matter. From the very beginning, after all, he often wrote as if he took people’s pro-attitudes to be more sensitive to their normative beliefs. I argue in Section 2 that we can find in Davidson’s later writings on what he called triangulation even clearer signs of a less purely functional and more thoroughly holistic account of what pro-attitudes are, and that this holistic account of what pro-attitudes are goes a long way towards vindicating his optimism about the causal theory of action. It also significantly reduces the differences separating Davidson’s version of the causal theory of action from versions of the non-causal theory like Anscombe’s. But since a causalist variant on Anscombe’s theory was very much what Davidson wanted, this counts in favor of my interpretation, not against it.

Although the details will have to wait until Section 2, the idea here is that pro-attitudes have what I call a systemic aim. They aim, as a system, to get normative matters right, so as to bring one’s actions into line with whatever

\(^1\) One influential discussion of this account can be found in Smith 1987.
normative truths there are. This is compatible with the fact that people’s pro-attitudes, considered individually, are sometimes oblivious or even opposed to their normative beliefs, much as David Velleman and others have insisted.\(^2\) The suggestion is only that people’s pro-attitudes are typically sensitive to their normative beliefs, not that every one of their pro-attitudes is always and essentially shaped by them. As we shall see, however, this is all it takes for Davidson to answer his two worries about the causal theory of action. If pro-attitudes are purely functional states, there are limits on how much they can help to turn mere behavior into intentional action. Too much work has to be done by the manner in which they cause behavior and by the manner in which they have themselves been generated by normative judgment. But as soon as we shift to a more holistic account of what pro-attitudes are, it becomes easier to understand how they can pull their weight in a causal theory.

Whether this shift to a more holistic account of what pro-attitudes are is enough to overcome other objections that have been leveled against Davidson’s causal theory is a harder question, raising larger issues than I can profitably take up within the confines of this chapter. Followers of Anscombe frequently complain that, by treating reasons for actions as causes of actions, Davidson made it impossible to understand how agents could have knowledge of their actions that was not merely observational or inferential but importantly practical.\(^3\) As we shall see, a more holistic account of what pro-attitudes are does give Davidson something interesting to say in response to this charge. It enables him to offer what we might call a balanced account of agential knowledge, according to which agential knowledge is observational or inferential and practical in roughly equal measure. And while this will almost certainly not be enough to placate many of Anscombe’s followers, it does introduce some pressure on them to consider whether they are demanding too much. Which side of this dispute should ultimately be expected to prevail is not a question that I will attempt to answer in this chapter. I will be more than happy if the arguments in this chapter simply establish that causal theories of action are still in the race.

1 Davidson’s Early Problems

Let’s begin with the problem of deviant causal chains. Davidson’s famous example goes as follows:

A climber might want to rid himself of the weight and danger of holding another man on a rope, and he might know that by loosening his hold on the rope he could rid himself of the weight and danger. This belief and want might so unnerve him as to cause him to

\(^2\) Velleman 1992. For more discussion, see Myers and Verheggen 2016, Chapter 6.

\(^3\) Some recent examples can be found in Ford, Hornsby and Stoutland 2011.
loosen his hold, and yet it might be the case that he never chose to loosen his hold, nor did he do it intentionally. It will not help, I think, to add that the belief and the want must combine to cause him to want to loosen his hold, for there will remain the two questions how the belief and the want caused the second want, and how wanting to loosen his hold caused him to loosen his hold. (1973, 79)

On the standard view of what pro-attitudes are, the climber’s desire to rid himself of the weight and danger is a disposition to perform actions he believes likely to rid himself of the weight and danger. Given his belief that he could rid himself of the weight and danger by loosening his hold on the rope, one would therefore expect him (all else being equal) to loosen his hold on the rope, and of course as Davidson tells the story this is precisely what he proceeds to do. But he seems not to do it for this (or indeed for any other) reason; his action seems not to be intentional even though it has been caused by a primary reason. So what has gone wrong?

A natural response would be to note that the climber’s primary reason seems only to have initiated a process that resulted in his action and not also to have monitored and guided that process as it unfolded. For even an action as simple as this could go wrong in a variety of ways. Suppose, for example, that when the climber loosens his hold on the rope, he does not loosen it enough; the rope begins to slip a bit, but not enough to rid him of the weight and danger. As Davidson tells the story, one might expect the climber to be relieved to see this. After all, he never chose to loosen his hold on the rope, he did so only because his desire and belief made him nervous, so one might expect him now to be trying to regain his grip. On the other hand, however, if he actually had loosened his hold on the rope with the intention of ridding himself of the weight and danger, one might expect him now to be redoubling his efforts. This suggests that a solution to the problem of deviant causal chains might be found by requiring that primary reasons be not simply initiating causes of actions but also what we might call monitoring causes of them.4 Of course, more would have to be said about what sort of causal relation this is. But one might understandably hope to treat this, much as Davidson does, as a difficult problem of detail, and not, as his opponents would have it, as a deeper problem calling the causal theory itself into question.

But the problem with this suggestion is that there is no way primary reasons could be monitoring causes of actions, if they bring together single pro-attitudes

---

4 Harry Frankfurt may mean to be ruling out possibilities like this when he says:

It is integral to the causal approach to regard actions and mere happenings as differentiated by nothing that exists or that is going on at the time the events occur, but by something quite extrinsic to them – a difference at an earlier time among other sorts of events entirely. (1978, 70)

But it is difficult to see what the grounds for this claim might be. It is certainly true that primary reasons could be monitoring causes of actions only if they evolve as actions unfold. As we shall see in just a moment, however, there are some very good grounds for thinking they do.
with single beliefs and if pro-attitudes are as the standard view claims. As we have already noted, on that view, the climber’s primary reason is a disposition to perform actions he believes likely to rid himself of the weight and danger, combined with a belief that he could rid himself of the weight and danger by loosening his hold on the rope. So understood, what more can his primary reason do but initiate a putative loosening of his hold on the rope? As matters begin to unfold, he might well notice that they either are or are not going as he wanted, and this might cause him to acquire new pro-attitudes, or new beliefs about what he could now do to satisfy the original pro-attitude, in which case he might find himself with a new primary reason, a new combination of pro-attitude and belief, that will initiate a new action. But this new action will be vulnerable to the worry about deviant causal chains just as the original action was. There is nothing in the staccato launching of new actions that could constitute the requisite monitoring of an unfolding action. It therefore seems safe to conclude that solving the problem of deviant causal chains would require a different account of what primary reasons are; and so the question is what a different account of primary reasons might be and whether Davidson might have had some such account in mind.

An obvious first step would be to replace the standard account of what pro-attitudes are with an account that regards pro-attitudes not as dispositions to perform actions but as dispositions to monitor and adjust behavior, so the climber’s desire to rid himself of the weight and danger could be treated, not as a disposition to perform actions he believes likely to have that result, but as a disposition to monitor and adjust his ongoing behavior in light of his evolving beliefs about how to achieve that result.5 Along with this, we could also replace Davidson’s account of what primary reasons are with an account that regards primary reasons not as static pro-attitude and belief pairs but as evolving pro-attitude and belief clusters, so the climber’s primary reason for loosening his hold on the rope would be his desire to rid himself of the weight and danger, not just paired with his belief that he could achieve that result by loosening his hold on the rope, but rather clustered with this belief and various others, such as (in our elaborated example) his belief that he is not loosening his hold on the rope enough and thus needs to redouble his efforts in order to rid himself of the weight and danger. The claim might then be that, as Davidson tells the story, the climber’s action is not intentional because his primary reason is merely the initiating cause; he is precisely not monitoring and adjusting his ongoing behavior in light of his evolving beliefs about how to achieve his desired result but just letting go.

5 Notice that this is not yet to introduce the more holistic account of what pro-attitudes are that I will ultimately be attributing to Davidson. Even someone who is satisfied with a less holistic account of what pro-attitudes are should be willing to acknowledge that pro-attitudes are not just simple dispositions to perform actions (whatever that could mean) but rather complex dispositions to monitor and adjust behavior with a view to achieving that result.
Although these proposals evidently take us some distance from Davidson’s position as it is usually understood, I suspect he would in fact have accepted them as at most just friendly amendments and perhaps even just minor clarifications. I will have much more to say about his account of pro-attitudes in due course. Regarding his account of primary reasons, it is important to note that while Davidson often described the component belief as a general belief that any action meeting description \( d \) will have the desired property, whatever it is, this is not how he described it when he first defined what a primary reason is. What he described there, to quote him again, is “a belief of the agent that \( A \), under description \( d \), has that property” (1963, 5) – in other words, a singular belief about the action being performed, not a general belief. Although this was undoubtedly a bit sloppy, I do not believe it was confused; I see it rather as tacitly acknowledging the fact that primary reasons for action must be regarded as comprising multiple and evolving beliefs. Pairing a pro-attitude with a suitably related general belief might suffice to initiate an action, but in order to function as part of the monitoring cause, the pro-attitude must subsequently work together with singular beliefs, beliefs about whether and how the action needs adjusting in order to achieve the desired result.\(^6\)

Now it might be objected that the threat of deviant causal chains would remain even if the climber’s primary reason were understood to be not just the initiating cause of his action but also its monitoring cause. Thus suppose, as before, that his desire to rid himself of the weight and danger, when combined with his belief that he could rid himself of the weight and danger by loosening his hold on the rope, so unnerves him that he loosens his hold, but not enough to achieve the desired result. And now suppose that, realizing this, he recognizes that he will need to redouble his efforts if he is to finish the job, and that this once again unnerves him, causing his hold to loosen even more. On the face of it, there seems nothing to prevent this comedy of errors (or should we say, successes?) from continuing until he finally ends up ridding himself of the weight and danger, exactly as he desires; but still he will neither have chosen to loosen his hold nor loosened his hold intentionally. Thus it might be objected that the adjustments we have proposed making to Davidson’s accounts of pro-attitudes and primary reasons have not brought us any closer to a solution to the problem of deviant causal chains. And of course this might in turn be offered as further proof that the problem of deviant causal chains is not merely a problem of detail but instead a problem fatal to the causal theory of action. But I believe this would be the wrong lesson to draw from our discussion so far.

\(^6\) I owe this talk of “general” and “singular” beliefs to Hunter 2015. Hunter himself denies that singular beliefs can be part of primary reasons, presumably because he thinks primary reasons cannot evolve over time, and so can serve only as initiating causes of actions and not as monitoring causes. But I think that is a mistake on a par with holding that Davidson believed actions don’t unfold over time but somehow just happen.
I say this because it seems to me that the nature of the problem has changed. So long as we were working with the idea that the climber’s desire to rid himself of the weight and danger is nothing more than a disposition to perform actions he believes likely to have that result, there did not seem to be any prospect of meeting what is surely a necessary requirement on intentional action – the aforementioned requirement that its primary reason be not just its initiating cause but also its monitoring cause. But once we shift to the idea that the climber’s desire is precisely a disposition to monitor and adjust his ongoing behavior in light of his evolving beliefs about what it would take to achieve that result, there is no obvious reason to deny that this requirement will be met in most cases. The fact that there will undoubtedly still be some deviant cases in which it will not be met is neither here nor there, since we will now be in a position to explain what they deviate from. My suggestion, therefore, is that Davidson’s response to the challenge of deviant causal chains makes sense and is defensible provided that we interpret him as abandoning the standard accounts of what pro-attitudes and primary reasons are in favor of something like the modified accounts that I have been attributing to him here. So let us turn now to consider whether these or similar modifications to his views can also overcome the problems that arise upstream from primary reasons when we ask what role normative judgment plays in generating them.

Davidson never highlighted these upstream problems as succinctly as he highlighted the downstream problem of deviant causal chains, but his entire discussion of weakness of the will is premised on the thought that there are problems here. As he said:

Another common line to take with incontinence is to depict the akrates as overcome by passion or unstrung by emotion … This … leads to the view that one never acts intentionally contrary to one’s best judgement, and so denies [principle] P3; there are no incontinent actions in the sense we have defined. (1969, 27–28)

It is not enough to show that an incontinent action was rationalized and caused by some wayward desire, because incontinent action is intentional and intentional action requires normative judgment to have played a role in its genesis. But what role? Surprisingly, Davidson goes on to maintain that there must be a sense in which one’s primary reason reflects one’s best judgment about one’s reasons for action. In order to accommodate the possibility of incontinent action, he then introduces a distinction between normative judgments that are

---

7 As Ralf Stoecker (2013) helpfully reminds us, it was never Davidson’s goal to spell out fully sufficient conditions for intentional action in purely causal terms. That could not have been his goal, for he thought the whole point of causal concepts was to figure in explanations that are not nomological. So we needn’t deny that people’s dispositions to monitor and adjust their behavior could cause their behavior to be “monitored” and “adjusted” in deviant ways. We need simply maintain that this account of pro-attitudes, unlike the standard account, gives us some prospect of explaining what the right ways are.
conditional and those that are unconditional. Incontinent action is intentional, he insists, because its primary reason does reflect one’s unconditional or “all out” judgment, even though it fails to reflect one’s conditional or “all things considered” judgment, and is to that extent irrational. Surely, however, this is not actually a plausible requirement on intentional action. So the question once again is whether the problems here are in fact fatal to Davidson’s causal theory of action, or whether we can find in the end good grounds to treat them as nothing more than problems of detail.

I say this is not a plausible requirement on intentional action because it seems perfectly obvious to me that people can and often do act intentionally, not merely against their conditional or “all things considered” judgments about what they have most reason to do, but also against their unconditional or “all out” judgments. Davidson might of course have denied this, but all the examples he deployed to show that people can act intentionally in cases of the former sort would seem effective in cases of the latter sort as well. Are we seriously to suppose that Austin could intentionally help himself to somebody else’s portion of ice cream if his judgment acknowledging the reasons against doing so was merely conditional but not if it was fully unconditional? Surely he would have scoffed at the suggestion that his powers were limited in this way! Indeed, Davidson’s requirement on intentional action seems so obviously implausible to me that I think we really must ask ourselves some hard questions about why he should have felt it necessary to propose any such thing. And the answer, I think, must lie with his muddled views about what pro-attitudes are. Presumably it was because he had not clearly disavowed the standard view of what pro-attitudes are that he believed there was a problem about intentional action calling for his requirement on intentional action as its solution. If only he had seen more clearly that the standard view of what pro-attitudes are is mistaken, he would have realized that there is no need for the problematic requirement on intentional action that he proposed.

On the standard view of what pro-attitudes are, a pro-attitude towards actions with a certain property, call it p, is, at bottom, just a disposition to perform actions one believes likely to have property p. Since nothing is presupposed, either about how one came to have this particular disposition, or about how one generally comes to have dispositions of this sort, the mere fact that this pro-attitude helps rationalize and cause one’s action (even in the right way) does nothing to establish that one’s action is intentional. Something more is needed, presumably something to the effect that the

---

8 I am hardly alone in this. See, e.g., Bratman 1979, Mele 1987, and Stroud 2013 for discussion.
9 The reference here is to the example discussed in Davidson 1969, at p. 29.
pro-attitude was generated in the right way, whatever exactly that turns out to be, from one’s deliberations about what one has most reason to do.\textsuperscript{10} But then the problem is that, no matter what we stipulate about the manner in which one’s intentional actions must be sensitive to one’s normative judgments, it seems obvious that particular actions could in fact fail to be sensitive to one’s normative judgments in the prescribed manner and yet still be perfectly intentional. So we are stymied; the standard view of what pro-attitudes are creates a puzzle about intentional action, while the possibility of incontinent action guarantees that the only plausible solutions have no chance whatever of succeeding. If Davidson really meant to be endorsing the standard view of what pro-attitudes are, his account of these matters was therefore doomed to fail; for in that case his only realistic chance of explaining how intentional action is possible was to deny that “pure” cases of incontinence are possible after all.\textsuperscript{11}

On the other hand, however, suppose Davidson meant to reject the standard view of what pro-attitudes are in favor of a more thoroughly holistic account, an account according to which one’s pro-attitudes typically are sensitive to one’s normative beliefs because they aim, as a system, to get normative matters right. On this account of what pro-attitudes are, the simple fact that a pro-attitude helps rationalize and cause one’s action (in the right way) typically is all it takes to establish that one’s action is intentional. Even if the particular pro-attitude in question fails to accord with one’s judgments (even one’s unconditional judgments) about what one has most reason to do, the fact that it is the product of a system aiming to get normative matters right would seem to be enough to quiet worries on this score.\textsuperscript{12} More would of course need to be said to explain why a general system of pro-attitudes aiming to get normative matters right sometimes gives rise to particular pro-attitudes failing to accord with one’s normative beliefs. But since there is no reason in principle to think this should be impossible, there would be no existential threat here to causal theories of action; the questions and problems that remain would be, in the relevant sense, questions and problems of detail, though nonetheless interesting and important for all that. Indeed, Davidson’s own subsequent attempts to explain incontinent actions by invoking partitions of the mind and the distinction between mental causes and motivating

\textsuperscript{10} There are philosophers who reject this presumption, of course. See Velleman 1992 for one interesting and influential example.

\textsuperscript{11} “Pure” cases of incontinence are cases where the agent acts against her unconditional judgment.

\textsuperscript{12} Of course, if it’s not a product of the system, but, say, the result of a blow to the head, then matters will be different. The point again is just that, on the more holistic account of what pro-attitudes are, intentionality is the norm and deviancy is the exception. Even when people fail to act on their better judgment, it is typically not because their aim to get normative matters right has been bypassed, but rather, as Pamela Hieronymi (2009) puts it, because settling on what to believe is one thing and settling on what to do is another.
reasons provide good illustrations of how interesting and important these matters can be.\textsuperscript{13}

Still, while it may be true that Davidson’s subsequent discussions of incontinent action lend some support to this interpretation, it is hard to deny that in his original (1969) discussion he was largely working with the standard view. Not entirely, perhaps, for he did embrace “a mild form of internalism” (p. 29) according to which one’s normative judgments are typically, though not always or necessarily, reflected in one’s pro-attitudes, and it is not easy to see how this claim could be defended on the standard view of what pro-attitudes are.\textsuperscript{14}

However, had he really meant to embrace a more thoroughly holistic account of what pro-attitudes are, he surely would have realized that the possibility of incontinent action is not as troubling as he was saying, and that his distinction between conditional and unconditional judgments is more relevant to discussions of failures of reasoning than it is to discussions of failures of will. I think the best we can say is that he was unclear in these early papers about exactly what pro-attitudes, and hence primary reasons, are. While he certainly gave some signs of understanding that the standard view of what they are is mistaken, he gave no serious indication of having decided on the account with which he thought it should be replaced. Perhaps it was only to be expected, therefore, that his defense of the causal theory of action should have been based on the standard view, despite the disastrous consequences this fact was bound to have both for his account of deviant causal chains and for his account of weakness of the will.\textsuperscript{15}

In any case, inasmuch as his early defense of the causal theory of action was based on the standard view of what pro-attitudes are, I think we must agree that the downstream problem of deviant causal chains and the upstream problems involving weakness of the will are fatal to it. Recent followers of Anscombe are quite right to reject this version of the causal theory, and in fact we can only wonder why it took so long for the tide to turn against Davidson on this issue. As I have been intimating, however, I think the causal theory of action should be based on a more holistic account of what pro-attitudes are, and that even in his earliest papers Davidson gave some indications that he appreciated this fact.

\textsuperscript{13}See Davidson 1982b and 1986. (In saying these attempts are important, I don’t mean to be endorsing them. Partitioning the mind may not be necessary.)

\textsuperscript{14}In fact, he at first insisted that they must be so reflected and only later, on the next page, acknowledged that that was too strong. This is one of several places where Davidson might seem to have flirted with the possibility that normative beliefs actually are, or entail, pro-attitudes. Usually, however, he was quick to acknowledge that that would be going too far, and that the relations between these states are more complex. Normative beliefs necessarily have some rational bearing on pro-attitudes, and because of this are typically reflected in them, but they do not entail them.

\textsuperscript{15}This is harsher than the judgment I rendered in Myers and Verheggen 2016, Chapter 6, but Davidson’s mishandling of these two issues clearly warrants it.
and saw at least roughly how the reconstruction should go. What I now want to argue is that he subsequently embraced this more holistic account much more decisively, and that it in fact does go a long way toward vindicating his early optimism about the causal theory. Whether it makes the causal theory safe against all the criticisms recent followers of Anscombe have leveled against Davidson is not something I shall attempt to establish here. My goal in the next section is simply to establish that the reconstructed causal theory is a much more formidable contender than most non-causalists seem to realize, and that this is the version of the theory they should be focusing on when discussing Davidson’s work, instead of always beating up on the simpler version of the theory that is prevalent in his early papers.

As I hope is clear, my suggestion is not going to be that there are two theories in Davidson, the early and the late, with the shift between them marking some fundamental departure from his original account. Insofar as there is a shift, I see it as a move from lesser clarity to greater, not from one clearly held theory to another. And as often happens in such cases, there is no “eureka” moment to point to, just incremental developments eventually adding up to something different, and better. In what follows, I shall focus on two developments that I believe are absolutely central to Davidson’s later writings but, somewhat paradoxically, almost never brought into the spotlight and properly examined, not even by Davidson himself. The first development has to do with Davidson’s understanding of the doctrine he christened the holism of the mental as it applies particularly to pro-attitudes. As we shall see, while the standard view allows that pro-attitudes are holistic to some extent, Davidson eventually came to think it is not holistic enough. The second development has to do with Davidson’s understanding of the role triangulation plays in determining propositional contents, and the lessons he thought we can learn about pro-attitudes from the fact that triangulation plays this role. As we shall see, Davidson eventually came to regard the triangulation argument as speaking in favor, not simply of a more holistic account of pro-attitudes, and thereby of the causal theory of action, but also of a kind of realism and even causalism about normative reasons and their relation to normative beliefs.

2 Davidson’s Eventual Solution

Davidson introduced the doctrine of the holism of the mental in a number of different ways, some quite a bit more helpful than others. Here are two early examples, the second of which is considerably more helpful than the first:

Beliefs and desires issue in behaviour only as modified and mediated by further beliefs and desires, attitudes and attendings, without limit. (1970, 217)
… we make sense of particular beliefs only as they cohere with other beliefs, with preferences, with intentions, hopes, fears, expectations, and the rest. It is not merely, as with the measure of length, that each one tests a theory and depends upon it, but that the content of a propositional attitude derives from its place in the pattern. (1970, 221)

The second passage is more helpful than the first because it makes it clearer that the holism of the mental is a doctrine, not just about the attribution of propositional attitudes, but also about their content. As Davidson later put it:

… the content of any given attitude depends on its place in the whole network [of attitudes that one has]. (1995a, 15)

Moreover:

Not only does each belief require a world of further beliefs to give it content and identity, but every other propositional attitude depends for its particularity on a similar world of beliefs. In order to believe that the cat went up the oak tree I must have many true beliefs about cats and trees, this cat and this tree, the place, appearance and habits of cats and trees, and so on; but the same holds if I wonder whether the cat went up the oak tree, fear that it did, hope that it did, wish that it had, or intend it to do so. (1982a, 98–99)

So the doctrine is intended to apply to pro-attitudes as well as beliefs. And certainly the standard view of what pro-attitudes are is compatible with this claim. If a pro-attitude towards actions with property $p$ is a disposition to perform actions one believes likely to have property $p$, then the content of this pro-attitude depends on the contents of those beliefs and, through them, on the contents of many others.

The question, however, is whether this takes the holism of the mental far enough. For on the standard account, the contents of a person’s pro-attitudes typically depend only on the contents of her descriptive beliefs and not at all either on the contents of her normative beliefs or on the contents of other pro-attitudes that she has. And yet, as Davidson came to realize, the typical case in fact seems very different:

Ordinary desires like wanting to win at chess compete with further values, they are conditioned by experience, and grow stale with repeated frustration. In other words, to have one desire, it is necessary to have many. (1990, 89)

Once again, though, this is not terribly well put. The suggestion might seem to be that the contents of a person’s pro-attitudes are directly dependent on one another, and indeed Davidson later maintained that the holism of the mental is “intra-attitudinal” as well as “inter-attitudinal” (1995a, 13–17). But I think a moment’s reflection reveals that what he actually had in mind cannot have been quite so simple.

Of course, a desire could explicitly be to act on some other desire or on some normative belief. But those aren’t the typical cases.
What exactly is one disposed to do insofar as one wants to win at chess? If one is not disposed to perform just any action one regards as likely to increase one’s chances of winning, that is presumably because one believes one’s reasons for seeking to win are not reasons to perform just any action that would promote that result, and one’s disposition is only to perform actions one believes one has reason to perform. Thus one might not be disposed to play raw beginners, despite believing it would greatly increase one’s chances of winning, because one believes one’s reasons for seeking to win have to do with the development of skills one would not develop by playing raw beginners. Of course, in that case, one probably also has a desire to develop those skills, so it may look as though the content of one’s desire to win at chess is directly shaped by the content of that other desire, in at least something like the manner to which Davidson seemed to be alluding in the passage quoted above. But in fact the contents of both desires are being shaped by one’s normative beliefs. This is not to say either that normative beliefs are pro-attitudes or that pro-attitudes are entailed by normative beliefs. Although Davidson occasionally said such things, he was usually careful to treat beliefs and pro-attitudes, even normative beliefs and pro-attitudes, as distinct existences. He always remained committed to what has become known as the Humean theory of motivation; it was what we might call the Humean – i.e., the standard – account of desire that he eventually put firmly behind him, holding instead that one’s pro-attitudes are typically shaped by one’s normative beliefs because they aim, as a system, to get normative matters right. As I have mentioned, premonitions of this shift exist even in the very early papers. I suspect, for example, that Davidson listed “moral views” among pro-attitudes in “Actions, Reasons, and Causes” (1963, 4) because he was already assuming they typically shape pro-attitudes. But it was only later, in thinking through his holism, that his dissatisfaction with Humean accounts of pro-attitudes became clear.

Now it might be objected that it is still not terribly clear. After all, as we acknowledged above, the standard account of what one’s pro-attitudes are does make them out to be somewhat holistic. It makes their contents dependent on the contents of one’s descriptive beliefs, although (again, typically, anyway) not on the contents of one’s normative beliefs. Moreover, proponents of this account of what pro-attitudes are could certainly allow that competition among them often influences their contents over time. If one desires both to win at

17 For a fuller discussion of this example, including responses to a number of possible rejoinders, see Myers and Verheggen 2016, pp. 143–147.
18 One well-known case has already been discussed in note 14 above. For another, less well-known case, see 1984, pp. 25–26.
19 Of course, this would require the further assumption that moral views are, or involve, normative beliefs. But Davidson undoubtedly was assuming that as well.
chess and to develop skills of the sort required to win against reasonably strong
opponents, it would hardly be surprising if the former desire were to transform
over time, becoming precisely a desire to win at chess against reasonably strong
opponents. Is it really clear that Davidson’s examples of and remarks concern-
ing intra-attitudinal holism were intended to be getting at anything more than
this? If he actually had a novel view of the nature of desire, why did he never
take the time to say what it was?

Surely, however, this last question, though understandable, is unfair. It is
understandable, because it would be nice to have a simple statement of the more
holistic account of desire to contrast with the standard claim that a desire to
perform actions with property $p$ is, at bottom, a disposition to perform actions
one believes likely to have property $p$. It is not fair to insist on this, however,
because the point of the more holistic view is precisely to deny that pro-
attitudes can be understood so atomistically. All we can offer, according to
the holist, are general claims about the form pro-attitudes typically take in
consequence of the fact that they share a certain systemic aim. Now, admittedly,
Davidson could have been clearer on this point. He could have said more
explicitly that people’s pro-attitudes aim, as a system, to get normative matters
right and that individual desires to perform actions with various properties are
thus typically dispositions to bring one’s behavior into line with the reasons one
believes one has to perform such actions. Still, the examples he offered were
invariably along these lines:

We would not want to make money unless we believed it would put us in a position to
obtain things we need or value; we would not want to go to the opera unless we thought
we would enjoy it (or that someone else wants us to be there, etc.). (1995a, 16)

In fact, if anything, he tended to overstate the position, neglecting to stress that
this is a systemic aim and that individual pro-attitudes very often stray from the
mark to greater or lesser degrees. But given how often he did think to mention
such possibilities, it cannot seriously be supposed that this is something he ever
meant to deny.

Further confirmation that Davidson took the holism of the mental to rule out
standard accounts of what pro-attitudes are can be found in his application of
the triangulation argument. The broad idea here is that the contents of people’s
propositional attitudes are in the first instance fixed by their triangulations with
one another and objects in their shared environment. Objects cannot, by
themselves, fix the contents of the attitudes that people have towards them;
that requires, in addition, the active participation of the particular people whose
attitudes they are. And this content fixing is not something a solitary person
could do on her own; it requires the back and forth of multiple people trying to
get one another straight. Now of course there is a great deal that could be said
both about the nature of the argument that Davidson offers for these claims and
about its ultimate merits. For our purposes here, however, it suffices to point out two ways in which this “interpersonal externalism” lends support to strong versions of the holism of the mental. For one thing, in assuming that people typically are motivated to engage in the back and forth triangulation requires, it attributes to them a systemic interest in getting things right; for another thing, since people cannot triangulate on one aspect of their shared environment without simultaneously triangulating on others, the view ensures that contents will be holistic through and through.

Now it must be acknowledged that, when Davidson first developed this argument, his focus was exclusively on the contents of descriptive beliefs. Pro-attitudes and normative beliefs were never mentioned. And it might be thought that he was right to ignore them, since it would be a mistake to suppose that normative properties exist on which people could possibly triangulate. But Davidson would not have agreed. Before ever broaching the triangulation argument, he tended to treat normative beliefs in ways more suggestive of realism than of either expressivism or constructivism. Later in his career (1995b), with the general argument largely in place, he explicitly took up the question of its particular application to normative matters, making his realist commitments clear. If the boys described by Gilbert Harman (1977) really were torturing that cat, Davidson would have attributed normative properties to the situation and invoked them in explaining people’s responses to it. Of course, many philosophers, following Harman, would dismiss this out of hand, arguing that causal explanations appealing to normative properties are at best quite unnecessary and at worst simply unintelligible. However, such charges are hard to sustain without begging the question, since they almost always depend on the assumption that pro-attitudes are exactly as the standard account says. If pro-attitudes are purely functional states, as the standard account claims, the idea that they are explained by and answerable to normative properties is indeed extremely hard to accept. But if this is not assumed, the dialectical situation changes very dramatically, and Davidson’s claim that their contents are typically fixed through triangulations on normative properties becomes hard to resist.

---

20 See Myers and Verheggen 2016, Part One, for a general defense of Davidson’s triangulation argument; its application to pro-attitudes is discussed in Part Two.

21 “Interpersonal externalism” is the helpful term Claudine Verheggen uses to describe Davidson’s non-communitarian version of social externalism. See Myers and Verheggen 2016, Chapter 3.

22 While others might also draw the distinction between all-things-considered and all-out judgments, the importance Davidson gives it is certainly suggestive of realism. So too is his life-long insistence that no single normative principle could ever be expected to capture the full story about reasons for action.

23 I discuss this at length in Myers and Verheggen 2016, Chapter 7. The key is to appreciate how the fact that pro-attitudes share a systemic aim makes it possible for situations to possess normative properties. Once this is understood, worries about metaphysical “queerness” and explanatory role become more tractable.
This is not the place to defend Davidson’s realism about normative properties. But it does show how far he was intent on taking holism in the case of pro-attitudes. If he had been prepared to live with the more limited holism afforded by the standard account of what pro-attitudes are, he could have avoided many of these issues. He would still have had to explain why people are so frequently motivated to participate in the triangulating behavior he deemed necessary to fix the contents of their descriptive beliefs; so he would have had to attribute to them an interest in getting things right. But he might have denied that this is a systemic aim of their pro-attitudes.24 Alternatively, he might have allowed that it is a systemic aim of their pro-attitudes, while maintaining that its presence marks only a minor departure from the standard account of what pro-attitudes are because it is solely an interest in getting descriptive matters right. The fact that Davidson did not try to make do with either of these arguments, but instead went all the way with normative properties on which people literally can triangulate, proves that he thought pro-attitudes are much more holistic than the standard account recognizes. So it is some such more holistic account of pro-attitudes that we should be attributing to Davidson when we pass our final judgment on his causal theory of action. And this, as I have argued, makes his theory much safer than it originally seemed both against worries concerning weakness of the will and against worries involving deviant causal chains.

Before we leave this issue, it is worth considering how Davidson’s causalism about normative properties corroborates our speculations about the form that a reconstructed causal theory of action should take. What might be meant in saying that the wrongness of torturing a cat could cause a person to believe that there is at least some reason for bystanders to intervene? Davidson’s suggestion would certainly not have been that there is some law to the effect that every well-placed witness to this scene would necessarily have come to that conclusion. For one thing, he would have insisted that people who had not triangulated on normative properties at all would not have acquired the conceptual wherewithal to form such a belief. But even if a person did have the concepts needed to form such a belief, and was perfectly well positioned to recognize that what the boys were doing was wrong, Davidson would have denied that a correct assessment of the situation was bound to result. He would have insisted that law-governed explanations are not possible at this level of description, and he certainly would have resisted any attempt to define normative properties in terms of the responses that particular sorts of people have to situations in particular sorts of circumstances. He was not assuming that secondary qualities should serve as the

24 That is to say, he might have tried to treat it just as one pro-attitude among all the others.
model for normative properties, and in any event did not believe that second-
ary qualities were relational properties in this sense.25

For our purposes, however, the important point here is not this negative one
about the absence of strict laws but a positive one about the nature of normative
judgment. Just as it is not possible to acquire one conceptual capacity without
acquiring many others, so too is it not possible to exercise one conceptual
capacity without exercising many others. Thus Davidson’s claim would never
have been that someone equipped with the concept of torture might simply
understand that there is reason to intervene in the torture of the cat. Much as
actions take time, and can always go wrong, so too is normative judgment more
complex than this suggests. These complexities may not always show up in the
phenomenology, but they are there beneath the surface all the same. To have the
concept of torture is to understand that something could appear to be torture and
yet not be. Thus someone who exercised her capacity to recognize cases of
torture must, in doing so, also have exercised her capacity to recognize things
that can sometimes look like torture. She must have concluded, for example,
that the cat was alive and not already dead, and that the boys were really setting
fire to it and not merely pretending to. And because concepts can be applied
only as such judgments are made, the causal role that was played by the fact that
it was torture must be a limited one. It may well be true that this observer would
not have formed the belief that there was reason to intervene if there had not in
fact been reason to intervene. But it is equally true that she would not have
formed this belief if she had not been making various related judgments with
a view to getting the question right.

I am belaboring this point because I think it is important to see how this
general approach to things gets extended to cover the whole of Davidson’s
theory of action. His picture is not one of links in a chain, with normative
properties connected to normative beliefs, normative beliefs connected to
motivating desires, and motivating desires connected to intentional actions,
and with each new link in the chain coming into being in one fell swoop,
solely and simply because the link that immediately preceded it had come into
being before. Such a picture, if it makes any sense at all, clearly does leave
something out. In fact, it would not be much of an exaggeration to say that it
leaves everything out, since what it leaves out is precisely the additional
causal factors helping to make actions more than mere behaviors, desires
more than mere dispositions, and beliefs more than mere representations – the
basic and indeed animating interest in getting things right that, as I have
argued, Davidson regarded as a systemic aim shared by people’s desires as
well as their beliefs.

25 See Davidson 1995b, p. 47. For more on Davidson’s metaethical views and commitments, see
Myers and Verheggen 2016, Chapter 7.
Perhaps it will be objected that to abandon the picture of links in a chain is to abandon the claim that actions can be understood in purely event-causal terms. Isn’t it, after all, the agent who is making her judgments, forming her intentions and monitoring her actions with a view to getting things right, not her desires or beliefs? How, indeed, can we say that actions are events appropriately caused by desires and beliefs, once we acknowledge that this causing unfolds over time and involves the oversight of agents? Viewed in this light, it might look as if the Davidson I have described here has disavowed causalism and joined the followers of Wittgenstein and Anscombe. But that is not so. It is true that the Davidson I have described here did not regard an action as an event that is fully caused by some single pair of desire and belief. A single pair of desire and belief may be its initiating cause, but its monitoring cause will be that desire plus the agent’s evolving beliefs about how things are going. In fact, the motivating desire might also evolve over the course of the action, along with the agent’s beliefs about her reasons for acting that way. After all:

During any continuous activity, like driving, or elaborate performance, like swimming the Hellespont, there are more or less fixed purposes, standards, desires, and habits that give direction and form to the entire enterprise, and there is the continuing input of information about what we are doing, about changes in the environment, in terms of which we regulate and adjust our actions. (Davidson 1963, 12–13)

It is also true that the Davidson I have described here was careful not to attribute an agent’s interest in getting things right to her individual desires or individual beliefs. But since he did attribute both the general interest in getting things right and the particular task of monitoring actions to the agent’s desires and beliefs considered as a system, I see no reason to fear that, by acknowledging that his causal theory of action needed to be made more complicated in these ways, Davidson was somehow abandoning his causalist commitments.

What, finally, of the objection that agents possess a special sort of knowledge – more practical than observational – of their intentional actions as they are in the course of performing them? As I suggested at the beginning, I think the more holistic account of pro-attitudes does give Davidson something interesting to say here, although perhaps not enough to placate everybody. If people act intentionally only when their reasons serve as monitoring causes of their behavior, they know what they are doing as they are in the course of doing it. Consider once again the unfortunate climber, but imagine this time that he really is letting go of the rope with the intention of ridding himself of the weight and danger. He needs observation to know that he has not yet fully succeeded in letting go; but not to know that he is letting go and so must redouble his efforts. The climber’s

26 This is Frankfurt’s worry (1978, 72). He thinks actions unfold under the guidance, not of propositional attitudes, but of agents.
knowledge that he is letting go is practical because it is grounded in the fact that his reason to let go is the monitoring cause of his behavior.\(^{27}\)

This may not be enough to placate some followers of Wittgenstein and Anscombe because they may conclude that it does not add anything new to what we have had all along. After all, it is typically agreed on both sides that agents know what they intend, and that self-knowledge of this sort is not observational but of some other kind. And of course nobody has ever denied that agents can come to know through observation that things in fact turned out in one particular way and not in some other. So it might be objected that no new, especially practical sort of knowledge is possessed in knowing that one’s actions have or have not unfolded as one had originally intended. The suggestion here, however, is not simply that agents can know, after their actions are done, that the intentions on which they were acting were or were not realized. The suggestion is rather that agents must know what they are doing as they are doing it, so they can monitor and adjust their behavior in light of their observations. The climber does not simply know that his intention was to let go; he knows that this is the intention on which he is currently in the course of acting. He may not know why this is the intention he is acting on, but he must know that he is acting on it in order to be acting on it. Followers of Wittgenstein and Anscombe may complain that this is still just knowledge of intention, and that especially practical knowledge is supposed to be not of intention but of action. To that, however, I think we could justifiably reply that they are in danger of setting the bar for practical knowledge too high. But that’s all for another time. All I claim to have established here is that Davidson’s causal theory of action needs to be understood in connection with his thoroughly holistic account of what pro-attitudes are, and that, when it is, it becomes much less vulnerable to objections than followers of Wittgenstein and Anscombe have suggested – and perhaps even less vulnerable than Davidson himself originally feared.

References


\(^{27}\) See Hunter 2015 for some closely related but interestingly different ideas about Davidson’s position on the question of practical knowledge.
Davidson’s Wittgensteinian View of Meaning

Paul Horwich

1 Introduction

Given that a fundamental division amongst philosophers of language is between those (in the majority) who advocate one or another heavily truth-theoretic account of meaning and those (a distinct minority) whose view of meaning is entirely use-theoretic, and given that Donald Davidson has been perhaps the most influential champion of the former approach and Ludwig Wittgenstein of the latter, the above title might well prompt bemusement. How can its author be so badly deluded or so perversely provocative?

Let me begin therefore by articulating my thesis without rhetoric. I’m certainly not claiming that Davidson’s express view of meaning was just the same as Wittgenstein’s. He’d surely insist that Wittgenstein’s definition of “the meaning of a word” as “its use in the language” is inadequate. He’d argue that meanings conceived of as uses can’t be reconciled with the engendering of sentence-meanings by word-meanings – so can’t be reconciled with the feature of natural language that explains how we are able to understand the countless complex expressions of our mother tongue that we’ve never come across before and whose meanings we haven’t been specifically taught. One of Davidson’s signature contentions is (roughly speaking) that in order to accommodate this so-called compositionality of meaning, we’ll need to articulate the facts as to what are meant by the expressions of a language as facts about their reference and truth conditions.

Moreover, he insists that the notion of truth needed in these articulations cannot be the deflationary conception (endorsed by Wittgenstein) whereby this concept is implicitly defined by the schema, “The proposition that \( p \) is true \( \leftrightarrow p \)”. For this presupposes a prior notion of proposition, hence a prior notion of meaning.

And it’s not just compositionality that Davidson thinks must be theorized in terms of inflationary truth. It’s also the epistemology of translation and interpretation – the right method for discovering which meanings (articulated truth-theoretically) a foreigner’s expressions possess.
All this may appear to amount to a thorough refutation of my title. But it doesn’t—because, as I said, I’m not claiming that Davidson’s opinions coincide with Wittgenstein’s. What I want to suggest rather is that the deepest and best component of Davidson’s overall view of meaning is a neo-Wittgensteinian use-theory, and that the rest of his picture can be and should be dispensed with. One of the missteps, I will argue, is Davidson’s truth conditional account of compositionality. For there’s a non-truth-theoretic account of this phenomenon that’s superior in many respects. Another misstep—especially needless—is his adamant opposition to the deflationary conception of truth. For, independently of the weakness of his arguments against it, I’ll show that he is wrong to think that his own account—even the truth-theoretic component of it—requires a more robust conception.

So I have to admit that this paper would more accurately and soberly (but pedantically) have been called something like “Davidson*’s Wittgensteinian View of Meaning”—where Davidson* is a somewhat idealized version of the man himself.¹

2 Transforming Davidson into Wittgenstein

I’ll start by trying to show that Davidson’s methodology of interpretation can be seen as a combination of two things:

(i) The Wittgensteinian idea that we should interpret a foreign speaker by pairing each of her words with a word of our own that is used, fundamentally, in the same way.

(ii) A reformulation of this idea in truth-theoretic terms. Specifically, Wittgenstein’s requirement that the pairing preserve basic dispositions of word-use (and therefore yield agreement in basic beliefs-forming dispositions) becomes the requirement that our interpretation of the words of the foreign language have the result that as many as reasonably possible of the sentences she accepts express beliefs that are true.

As Davidson puts his idea: we are to interpret the foreigner’s expressions charitably. We are to assign them reference conditions, satisfaction conditions, and hence truth conditions, in such a way as to “optimally” maximize the number of truths amongst the sentences she holds true. This optimal maximization is achieved in part by not expecting a given belief of the foreigner’s to be true when the evidence on which it was based was misleading. So, for example, the fact that I accept “Mars rotates” and Galileo did not accept “Marte gira” needn’t be a reason for me to avoid interpreting his “gira” with my “rotates”; because, if

¹ For Wittgenstein’s views see Sections 1–86 and 138–242 of his Philosophical Investigations (1953). For Davidson’s views see the essays collected in his Inquiries into Truth and Interpretation (1986).
I know how lousy his telescope was, I’d expect such divergences of opinion between us.

Here’s why I say that interpretation guided by Davidson’s principle of charity and interpretation guided by Wittgenstein’s use theory of meaning are not as different as they might initially seem to be.

In the first place, it would be implausible to suppose that the correct specification of the meanings of the expressions in S’s (foreign) language is the one that optimally maximizes the number of actual truths amongst the sentences that S accepts. For since we may be unsure of many things that S is not unsure about, and since we know that many of the things we think are true are in fact mistaken, we simply aren’t able to tell which truth conditions must be attributed to S’s sentences in order to maximize (or optimally maximize) the number of actual truths S expresses; so, if it’s actual truth that matters, we couldn’t be at all sure which interpretation of her language is the right one. Even if S is obviously speaking our language (since S and her community live in Australia, produce the same sentences as we do in the same circumstances, and appear to understand us perfectly) there would be the possibility that we don’t really agree with them. For perhaps our own judgments are very often mistaken; so there would be room for a non-homophonic interpretation of their speech that would leave them with many more true beliefs than we have. Therefore Davidson’s proposal must involve invoking, not what’s actually true, but what we interpreters take to be true. The right interpretation for me to give of S’s language is supposed to be the one that optimally maximizes S’s truthfulness by my lights.2 But that reduces to the one that optimally maximizes agreement between us.

So the concept of truth can to some extent be removed from Davidson’s picture. Only to some extent because, from his perspective, an interpretation must be based on an assignment of referents and satisfaction conditions to the foreigner’s words – the right assignment being the one that optimally maximizes the extent to which she and I hold-true sentences with the same truth conditions.

But let us set aside for a minute Davidson’s insistence on truth-theoretic articulations of meanings and his further reliance on the concept of truth. I’ll return to these important matters later. Let’s work for now with a naïve, non-truth-theoretic way of putting his proposal: to a first approximation

---

2 Davidson himself is quite clear and consistent on this point. See, e.g., his “Truth and Meaning”, in Inquiries into Truth and Interpretation, p. 27 (“The linguist . . . will attempt to construct a characterization of truth for the other which yields, so far as possible, a mapping of sentences held true (or false) by the other on to sentences held true (or false) by the linguist”). And see also p. xvii of his Introduction to that volume (“[The Principle of Charity] counsels us quite generally to prefer theories of interpretation that minimize disagreement”).
I’m to attribute meaning to the sentences S accepts in such a way that the beliefs they express are what I also believe

where

$$S \text{ believes that } p = \text{ For some } u, \text{ S accepts } u \text{ and } u \text{ means for } S \text{ what } \text{“}p\text{” means for me.}$$

This must be modified to incorporate Davidson’s recognition that the desideratum isn’t simply maximization of agreement, but optimal maximization of agreement. Thus we have:

I’m to attribute meaning to the sentences S accepts in such a way as to maximize agreement between the beliefs S has in any given epistemic circumstances and the beliefs that I would have in those circumstances.

And, given this formulation, all that’s now needed to boil his method down to Wittgenstein’s is to appreciate that Davidson’s emphasis on optimization amounts to the Wittgensteinian idea that the only agreements crucial to interpretation are in the basic rules or propensities of word-use governing which sentences are accepted and when.3 For example, in order for me to correctly interpret S’s “gira” with my “rotates” what’s required is that my basic use of “rotates” is the same as S’s basic use of “gira”.

To repeat: the principle:

I must interpret S in such a way as to optimally maximize the amount of truth in S’s beliefs

which, as we’ve seen, amounts to the principle

I must interpret S is such a way as to optimally maximize S’s agreement with me

becomes

I must interpret S in such a way as to maximize agreement on those practices of belief acquisition and revision that are epistemologically fundamental.

And putting this in terms of language rather than thought, we arrive at the Wittgensteinian principle:

I am to interpret S’s words, $$w_1, w_2, \ldots, w_k$$, in terms of my “$$f_1$$”, “$$f_2$$”, ..., “$$f_k$$”, just in case S’s basic acceptance practice with respect to her sentences is just like mine – except for the fact that S’s involves her words where my practice involves the corresponding words of mine.

3 Although Wittgenstein equates the meaning of a word with its “use”, it’s clear I think that he has in mind its “basic use”. – That is, not the total use of the word (including every event in which a sentence containing that word is accepted), but rather a rule or propensity of use that, in conjunction with other factors, explains that total use. For further discussion of his view of meaning, see Chapter 4 (“Meaning”) of my Wittgenstein’s Metaphilosophy (2013).
The central idea here is that each person’s verbal output is explained in part by fundamental propensities (or implicitly followed rules) of word-use that determine when sentences are to be accepted (and when not) as a function of further factors, including perceptual inputs and inferential dispositions.

For example, it may be that we English speakers have basic linguistic tendencies along roughly the following lines:

(i) To accept “That is red” in response to the sort of visual experience normally provoked by observing a clearly red surface
(ii) To accept “p” when “p and q” is accepted
(iii) To accept “x is water ↔ x has the underlying nature of the stuff in our seas, rivers, lakes and rain”
(iv) To accept instances of “<p> is true iff p”
(v) To accept the conditional, “∃Φ T*(Φ) ⇒ T*(neutrino)” – where “T*(neutrino)” is a formulation of neutrino theory.4

And in that case, if S’s overall use of w1 derives from the tendency described by substituting that word for “red” in (i), then S’s w1 and our “red” have the same meaning. If S’s use of w2 derives from the tendency described by substituting that word for “true” in (iv), then S’s w2 and our “true” have the same meaning. And so on. As Davidson says, interpretation rests on fundamental agreement.

No doubt many of his followers will react against this attempt at assimilation to Wittgenstein. They are likely to protest that Davidson’s militantly truth-theoretical account of interpretation, even if it appears to bear a certain similarity to Wittgenstein’s approach, is an enormous advance, insofar as it’s designed to cohere with Davidson’s ground-breaking and uniquely adequate strategy for explaining the compositionality of meaning.

In what follows I’ll respond at three levels to such protests. After reviewing Davidson’s line on compositionality, I’ll survey the respects in which it’s actually quite far from adequate. Then I’ll describe an alternative way of trying to accommodate that phenomenon – one that’s much simpler and free of the difficulties afflicting Davidson’s strategy. Finally, I’ll show how what’s right and important in his ‘principle of charitable interpretation’ is distorted and obscured by his need to make it cohere with his misguided account of compositionality.

3 Davidson’s Innovation

What kind of assumptions about the words of a foreign speaker would put us in a position to interpret each of the unlimited number of things she might say, and

---

4 This is an instance of the idea (developed by Russell, Carnap, Ramsey, and Lewis) that accepting a theory, “T(Φ)”, can be factored into a pair of independent commitments: first, a substantive claim – “∃Φ T(Φ)” – to the effect that there exists some phenomenon with the characteristics specified by the theory; and second, the decision to call that phenomenon (if it really exists), “T” – a decision that amounts to accepting the conditional, “∃Φ T(Φ) ⇒ T(Φ)”. 

Paul Horwich
how might such assumptions be verifiably correct? Like Quine, Davidson supposed that answers to these questions would constitute a more-or-less complete philosophical account of meaning. However, unlike Quine, who took for granted that the hard issues here were confined to the second question – “How, if at all, can determinately correct translation manuals be found?” – Davidson focussed unprecedented attention on the first one – “How could interpretive conclusions regarding the infinitely many complex expressions of a language be derived, as they surely would have to be, from finitely many interpretive assumptions about its simple terms?” That is, what form would our hypotheses about the meanings of someone’s words and sentences have to take in order that the latter be obtainable from the former?

Davidson’s approach to this previously unrecognized problem was ingenious and proved to be extremely seductive: we should solve it by piggybacking on Tarski’s work on truth. For Tarski showed us how truth conditions for the sentences of a formal language could be deduced in predicate logic from premises specifying the referents of its names and atomic predicates, and from further premises specifying, for each logical connective, how the referent (or truth-value) of any complex that is formed with it depends on the referents (or truth-values) of the connected expressions. Therefore, if we identify a word’s possessing the meaning it does with its having a certain referent (or, in the case of a connective, with the fact about how the referents/truth-values of the complexes formed with it depend on the truth-values/referents of the connected expressions), and if we identify a sentence possessing the meaning it does with its having a certain derivable truth condition, then Tarski-style deductions of truth conditions become precisely what we were looking for: namely, derivations of sentence-meanings on the basis of assumptions about word-meanings.

Consider for example how to arrive in such a way at an interpretation of the Italian sentence, “Marte gira”. We start with the pair of semantic premises:

5 For Quine’s account of meaning and translation see his Word and Object (1960). For Davidson’s account see especially his “Truth and Meaning” and “Radical Interpretation”, both in Inquiries into Truth and Interpretation.

6 For the sake of a simplified exposition, the sort of meaning on which I’m focussed in this paper is the literal semantic meaning, within a given language, of an expression type. In that vein, the central Davidsonian question is how a person’s knowledge of the meaning, in that sense, of a complex expression is derived from her knowledge of the meanings of its words. A further question – obviously related – concerns what statement is made by a given sentence-token (i.e. which proposition it expresses), and asks how that is inferred from features of the component word-tokens, from their syntactic mode of combination, and from the conversational contexts in which the sentence appears. I won’t be addressing that further question here – except to note the plausibility of supposing that we arrive at what is expressed by a given sentence-token on the basis, in part, of our knowledge of the literal semantic meaning of the type to which it belongs. Thus the question at issue here is the more basic one.

7 The preceding paragraph sketches Davidson’s initial formulation of his proposal. See Section 4 for why and how this version came to be modified.
The singular term, “Marte”, refers-in-formal-Italian to Mars

(k) (The result of applying the formal predicate, “gira(x)”, to a singular term that refers-in-formal-Italian to k, is true-in-formal-Italian ↔ k rotates)

And we also have the syntactic premise

The result of applying “gira(x)” to “Marte” is “gira(Marte)”

From these assumptions we infer that

“gira(Marte)” is true-in-formal-Italian ↔ Mars rotates

Then we invoke the fact that

The formal Italian, “gira(Marte)”, gives the meaning of the ordinary Italian sentence, “Marte gira”

which puts us in a position to conclude that

“Marte gira” is true-in-ordinary-Italian ↔ Mars rotates

and thereby to interpret the ordinary Italian sentence.

More generally, Davidson conjectured that, given the addition of axioms specifying the truth-theoretic imports of further terms – not merely names and predicates and the classical logical constants – then we could infer, for every sentence of a natural language, what it means (and hence explain why it means what it does) by showing, along Tarskian lines, why its logical formalisation (or regimented equivalent) has the truth condition that it does.

4 Troubles for Truth-Theoretic Semantics

The research projects engendered by the Davidsonian paradigm fell into two groups. First there were concerted attempts to show how the strategy really could be applied to all ordinary sentences, including those built with linguistic devices whose logical formalization had not previously been established. How, for example, might we deduce the truth conditions of sentences containing adverbs, or that-clauses, or counterfactuals, or attributive adjectives, or conditional-probability constructions, on the basis of premises concerning the referents of their words? To that end, how could such sentences be formalised in first-order predicate logic? Over the last forty odd years clever solutions have been found to several problems of this sort, although many kinds of sentence still remain intractable.

8 The route through ‘formal Italian’ is necessary because Tarskian deductions of truth conditions can be carried out only for sentences in some logically regimented (i.e. formalised) part of the language. Ordinary sentences are then dealt with by attributing to them the same truth conditions as their formalisations.
The second set of issues that had to be addressed were foundational. For instance, does the truth condition of a sentence really suffice to determine its meaning? After all (and as Davidson himself pointed out) if the biconditional, “if and only if”, is construed materially (as it is in predicate logic), then, for example, “snow is white” is true if and only if grass is green (since the sentences on each side the “if and only if” are both true). However “snow is white” obviously doesn’t mean that grass is green! So his strategy will require either some non-material construal of “u is true if and only if p”, or else some modal operation on the above overly weak material construal of it that will yield something strong enough to ensure “u means that p” and hence “If S accepts u, then S believes that p”.

Davidson himself made several attempts to solve this problem. His last and best proposal was that we should interpret a foreign utterance, u, with our sentence, “p”, just in case “u is true iff p” is a theorem of whichever truth-theory for the language of u optimizes agreement between speakers of that language and ourselves.9 And, to its credit, this proposal would indeed preclude our interpreting “Schnee ist weiss“ with “Grass is green” – because even though, understood materially, “Schnee ist weiss“ is true iff grass is green, it won’t be deducible from the truth-theory that optimizes agreement between German speakers and ourselves.

However, not all clearly incorrect interpretations are precluded. This is because, if the optimal truth-theory for a foreign language deductively entails that u is true iff p – and if “q” is logically equivalent to “p” – then that truth-theory will also entail that u is true iff q. But surely it would be wrong, for example, for us to interpret “Schee ist weiss” with our “It’s not the case that it’s not the case that snow is white”.

Another foundational question is whether natural language sentences really do have the same meanings as the best logical reformulations one might find for them. For example, is it plausible that “John might win” has precisely the same meaning as “(∃x)[PossWorld(x) & WinsAt(John, x)]”, and that the sense of “Mary is walking in the park” is identical to that of “(∃x)[Walking(x) & By(x, Mary) & InPark(x)]”? Even if the members of such pairs necessarily have the same truth value, the structural and semantic differences between them are so substantial that one might well doubt whether they could qualify as exact synonyms of one another.

Moreover, it’s one thing to allow that interpretations of logical sentences provide interpretations of the natural language sentences they formalize, but it’s another thing to show how specific theses of that sort can be justified. And this epistemological requirement presents a further difficulty for Davidson’s semantics. Obviously his answer can’t be that we give separate derivations of truth conditions for the natural sentence and its logical formulation, and discover that they yield the same result. That’s because, from his perspective, the

---

9 See “Radical Interpretation”, esp. pp. 138–9 of Inquiries into Truth and Interpretation.
truth condition of the former can be established only via the assumption that it has the same truth condition as the latter. It would appear that such an assumption could be justified only by finding that those two sentences have, in some deeper sense, “the same meaning”. But how could such a sense be accommodated within Davidson’s austere truth-theoretic framework?

The expectation that all these technical and foundational difficulties will eventually be overcome derives largely from the conviction that, since there is no decent alternative to the Davidsonian truth-theoretic view of compositionality, it must be more-or-less right.

I would suggest, however, that there is a decent alternative – a simple, deflationary alternative – one whose correctness would nonetheless cohere with Davidson’s principle of charitable interpretation. The existence of this option undermines the motivation for the Davidsonian research program in compositional semantics and makes it unnecessary to swallow its various implausible commitments.

5 A Deflationary Account of Compositionality

The alternative I have in mind may be described as “deflationary”, for its upshot is that Davidson’s problem (of how to derive interpretations of complex expressions) has an obvious and untheoretical solution. This solution assumes the principle of compositionality: that a complex expression’s meaning what it does is determined by the parts’ meaning what they do and by how those parts are combined. It assumes, in other words, that two expressions have the same meaning if and only if they are structurally isomorphic and if each component of one of them means the same as the corresponding component of the other.

For illustration, look again at how to reach an interpretation of “Marte gira”. We can begin with premises specifying the meanings of its words:

- “Marte” in Italian means the same as our “Mars”
- “_ gira” in Italian means the same as our “_ rotates”.

Then, in light of the principle of compositionality, we can infer

The result of applying “_ gira” to “Marte” in Italian means the same as the result of applying our “_ rotates” to our “Mars”.

And finally, given the syntactic facts

The result of applying “_ gira” to “Marte” = “Marte gira”

The result of applying “_ rotates” to “Mars” = “Mars rotates”

we can deduce the interpretation

“Marte gira” in Italian means the same as our “Mars rotates”.

Paul Horwich
And, in general, whenever some foreign expression is constructed by imposing a certain combinatorial procedure (i.e. a certain syntactic structure) on certain words, then we can interpret it in our language with the expression that results from imposing exactly the same procedure on synonyms of those words.\textsuperscript{10}

For thousands of years good translations and interpretations have been given by people who did not have the benefit of Davidson’s theory of compositionality. So how could they manage? Well, it’s a fair bet that they employed the naïve but perfectly adequate technique just sketched. And it’s also a fair bet that Frege, Wittgenstein, Quine and nearly all other pre-Davidsonian philosophers of language regarded the correctness of that technique as too obvious to be worth spelling out.\textsuperscript{11}

If this technique will in fact do, then Davidson’s approach and its attendant difficulties can be put behind us. We can abandon the struggle to find a conception of “truth condition” that’s sufficiently strong to capture meaning. We will then have obviated the need to cram every natural language construction into the narrow and gratuitous mould of predicate logic. And in that case there will be no reason to claim, rather implausibly as we have seen, that the predicate logic formalisation of a natural language sentence will exactly preserve its semantic meaning. Nor will there be need to resolve the tension between, on the one hand, having to assume that such pairs have the same truth conditions and, on the other hand the unavailability (given Davidsonian commitments) of any conception of meaning that would allow such assumptions to be justified and explained. What a relief!\textsuperscript{12}

\textsuperscript{10} Although we translate a foreign complex expression using one of our own that involves synonymous words and imposes the same structure on them – i.e. a structure that results from the same combinatorial procedure – the resulting order of synonymous words need not be the same. For the basic elements of a language include schemata. Therefore word order will partially derive from where the ‘slots’ in these schemata are located. Thus an identity of combinatorial procedure is quite consistent with a difference of word order. For example, our “blue sky” interprets the French “ciel bleu” because (i) “bleu” means “blue”, (ii) “ciel” means “sky”, and (iii) the former complex expression results from applying “blue” to “sky”, and the latter results from applying “bleu” to “ciel”.

\textsuperscript{11} Wittgenstein’s \textit{Tractatus} (1922) offers a truth-theoretic view (which might well have inspired Davidson) of how the meanings of logically complex sentences are engendered by the meanings of logically simple (“elementary”) sentences. However, the \textit{simple} sentences were said to possess their meanings, not in virtue of their truth conditions being logically entailed by the referents of their words, but rather in virtue of the structural identity of the representing fact (concerning words) and the represented possible fact (concerning the referents of those words).

\textsuperscript{12} This last section, and the next one, are revised versions of material in “Deflating Compositionality”, Chapter 8 of my \textit{Reflections on Meaning} (2005).
6 Davidson’s Pre-emptive Objections

But will this deflationary approach do? Davidson himself was always aware of it. So it is worth our while to examine his various reasons for rejecting the idea and to consider how persuasive they are.

In the first place, he argues that a manual of translation does not itself fully specify meanings, and so cannot constitute an interpretation. For one can know (e.g. on the basis of testimony) that two expressions should be intertranslated – i.e. that they have the same meaning as each other – without having any understanding of either one of them.

However, the problem he set was to specify the assumptions we might make that would enable us to interpret a foreign language. And one good answer is that a correct manual of translation into our own language will do the trick. Granted, the information it provides will suffice only given an as-yet-unexplained further fact – namely that we understand our own language. But the explanation of that further fact is not that we make additional explicit assumptions – assumptions, this time, about our own words. For, if that were what understanding one’s own language amounted to, the question would arise as to how the terms in which the additional assumptions are articulated could themselves be understood – and we would be on the verge of a vicious regress. Rather, the understanding we have of our own terms must be seen as a species of know-how, consisting in our propensities for when and where to deploy them. So it’s wrong to insist, given a translation of some foreign expression, that further explicit assumptions about our own expressions are needed.

But, for the sake of argument, let us bow to this objection of Davidson’s – at least to the letter of it – and take up his challenge to specify what knowledge (whether explicit or implicit) would suffice to understand a language (whether someone else’s or one’s own). This creates no particular difficulty for the deflationary strategy, because it is a simple matter to implement it in terms of meaning-facts rather than translation-facts. We can begin by adopting the convention that each capitalised English expression is to refer (in English) to the meaning of the original lower-case expression: thus “Mars”

\[\text{See his “Radical Interpretation”, reprinted in Inquiries into Truth and Interpretation – esp. pp. 129–130.}\]

\[\text{It’s important to note (i) that Davidson himself takes a person’s understanding of her own language to consist in her dispositions to hold-true sentences as a function of her evidential circumstances; and (ii) that his own theses of sentences-interpretation do not take the form, “u is true iff p”, but take the form,}\]

\[\text{u is correctly interpreted by my ‘p’ just in case “u is true ↔ p” is a theorem of the}\]

\[\text{most charitable truth theory for the language of u}\]

\[\text{which (just like my naïve alternative) deliver nothing more than a mapping from expressions of}\]

\[\text{the target language to expression of the interpreter’s language.}\]
means MARS, “_rotates” means _ROTATES, “Mars rotates” means MARS ROTATES, and so on.¹⁵ Then we can invoke the above-mentioned deflationary principle of compositionality, but in a slightly different form due to Frege: namely, that the result of applying one term to others (thereby producing a complex expression) means the result of applying the meaning of the first term to the meanings of the others.¹⁶ From these premises, together with assumptions about how “Mars rotates” and “Marte gira” are composed from their component words, we can infer that they both mean the result of applying the function _ROTATES to the argument MARS – i.e. that they both mean MARS ROTATES. Thus we can derive (and hence explain), for each sentence of a language (including our own), a fact concerning what it means.

At the beginning of his “Truth and Meaning”, Davidson dismisses precisely that way of trying to give a

… useful account of how the meanings of sentences depend upon the meanings of the words (or other structural features) that compose them. Ask, for example, for the meaning of ‘Theaetetus flies’. A Fregean answer might go something like this: given the meaning of ‘Theaetetus’ as argument, the meaning of ‘flies’ yields the meaning of ‘Theaetetus flies’ as value. The vacuity of this argument is obvious. We wanted to know what the meaning of ‘Theaetetus flies’ is; it is no progress to be told that it is the meaning of ‘Theaetetus flies’. This much we knew before any theory was in sight. In the bogus account just given, talk of the structure of the sentence and of the meanings of words was idle, for it played no role in producing the given description of the meaning of the sentence.¹⁷ [emphasis added]

But the Fregean answer does not merely apply the logical law of identity to the meaning of “Theaetetus flies” – which would indeed have no semantic importance. Rather, it incorporates the principle of compositionality by maintaining that the meaning of the result of applying a function-expression to certain argument-expressions equals the result of applying the meaning of the function to the meanings of its arguments. Thus it characterises the meaning of “Theaetetus flies” by describing it as the result of applying the meaning of “flies” to the meaning of “Theaetetus”. Therefore, contrary to what Davidson

¹⁵ Note that although the term “MARS” is introduced to name the meaning of “Mars”, it is not synonymous with “the meaning of “Mars”. It is a substantive, contingent fact that “Mars” means what it does – i.e. that “Mars” means MARS.

¹⁶ Frege provided an especially compelling articulation of the deflationary view of compositionality. In his account: (i) each primitive term has a certain functional character (names are objects, predicates are first-level functions, quantifiers are second-level function, etc.); (ii) complex expressions result from applying terms to each other in a way that is consistent with their functional character (e.g. first-level functions can be applied only to objects); (iii) the referent and the meaning (i.e. sense) of a term each has the same functional character as that of the term itself; and (iv) the referent and meaning of each complex are determined by a sequence of functional application that exactly parallels the way that the complex expression itself is determined. See his “Concept and Object” and “Function and Concept” in Translations from the Philosophical Writings of Gottlob Frege, eds. P. Geach and M. Black (1952).

¹⁷ p. 20 of Inquiries into Truth and Interpretation.
says, our assumptions about the structure of a sentence, and about the meanings of its words, play an essential role in our characterisation of the meaning of that sentence.

It must be conceded that we have not yet identified (in some peculiarly direct sense) the meaning of “Theaetetus flies”. We have specified it merely via the construction description, “the result of applying _FLIES to THEAETETUS”; but we have not said what that result is, what the description describes.\(^\text{18}\)

However, no such identification is needed as far as interpretation is concerned. Our deflationary strategy for arriving at an understanding of foreign sentences works completely independently of any assumptions on that score.

Nevertheless, it’s not too hard to come up with a natural answer. If we apply _FLIES to THEAETETUS, what we get is a further meaning-property – namely, THEAETETUS FLIES. This meaning-property is possessed by certain complex expressions – by whichever sentences mean the same as “Theaetetus flies”. And it is constituted by the construction-property

being the result of applying a word meaning _FLIES to a word meaning THEAETETUS.

Going deeper: assuming that the pair of meanings, _FLIES and THEAETETUS, may be identified with specific non-semantic properties, U1 and U2, of the English words, “flies” and “Theaetetus” (e.g. perhaps their basic use properties, à la Wittgenstein), then what comes from applying the first of these meanings to the second will be the meaning that is constituted at a more profound level by

being the result of applying a word that has U1 to a word that has U2.

Thus the meanings of complex expressions may, in the end, be quite definitely identified.\(^\text{19}\)

A little later in “Truth and Meaning” Davidson makes a separate criticism of the deflationary Fregean approach:

This is the place to scotch another hopeful thought. Suppose we have a satisfactory theory of syntax for our language, consisting of an effective method of telling, for an arbitrary expression, whether or not it is independently meaningful (i.e. a sentence), and assume as usual that this involves viewing each

\(^{18}\) For a sympathetic elaboration of Davidson’s complaint that the deflationary approach fails to specify directly what the meanings of complexes are, see James Higginbotham’s “A Perspective on Truth and Meaning”, in The Philosophy of Donald Davidson, edited by Lewis Hahn (1999).

\(^{19}\) For further discussion and defence of the deflationary view of compositionality, see my “The Composition of Meanings” (1997, pp. 503–31), reprinted as Chapter 7 of my Meaning (1998).

Notice that a trivial derivation of compositionality – one might consider it an ‘explanation’ in some very weak sense – results from our identification of the meaning-property of a complex expression with the construction-property of that expression (i.e. its property of the form, ‘e results from applying such-and-such combinatorial procedure to words with so-and-so meanings’).
sentence as composed, in allowable ways, out of elements drawn from a fixed finite stock of atomic syntactic elements (roughly, words). The hopeful thought is that syntax, so conceived, will yield semantics when a dictionary giving the meaning of each syntactic atom is added. Hopes will be dashed, however, if semantics is to comprise a theory of meaning in our sense, for **knowledge of the structural characteristics that make for meaningfulness in a sentence, plus knowledge of the meanings of the ultimate parts, does not add up to knowledge of what a sentence means**. The point is easily illustrated by belief sentences. Their syntax is relatively unproblematic. Yet, adding a dictionary does not touch the standard semantic problem, which is that we cannot account for even as much as the truth conditions of such sentences on the basis of what we know of the meanings of the words in them.20 [emphasis added]

The central contention here is that knowledge of the syntax of a sentence – for example, a belief attribution – plus knowledge of what its words mean, will not enable us to infer the sentence’s truth condition. But I can find no construal of this claim in which it constitutes a good objection to the naïve deflationary strategy under consideration.

Does it mean that the imagined knowledge about a sentence, u, cannot yield any conclusion of the form ‘u is true if and only if p’? If so the claim is mistaken. Once we have determined, via the naive method described and illustrated above, that a sentence means the same as our “John believes that dogs bark”, we may straightforwardly conclude that it is true if and only if John believes that dogs bark. We simply invoke the schema: ‘s means the same as our “p” → (u is true ↔ p)’.

So, perhaps Davidson’s intention is to complain that the proposed account does not yield a *compositional* account of truth conditions – a Tarski-style deduction of them from premises about the referential properties of words. In that case, the obvious response is that part of the point of the approach I’m recommending is to undercut the need for such an account. For we can interpret foreign speakers perfectly well without it, merely on the basis of the unexplained principle that meaning is compositional.21

I conclude that Davidson’s resistance to the deflationary view of compositionality was always unjustified and remains so. In order to determine what

---

20 p. 21 of *Inquiries into Truth and Interpretation*.

21 Note that the quoted passage from “Truth and Meaning” appears before Davidson’s presentation of his own solution to the problem of how interpretation of an entire language is possible; it occurs in the context of critical discussions of various initial attempts to solve it. His arguments that these attempts all fail are intended to give support to the truth-theoretic alternative solution that he goes on to articulate. But in that case these arguments cannot legitimately presuppose that we already accept his solution.

Thus Davidson’s proposal to deploy truth-theoretic notions instead of meaning-theoretic notions in the derivations of sentence-meanings is not motivated – appearances to the contrary – by any antecedent argument that the latter notions are intolerably obscure (or are unsuitable for some other reason), but rests merely on the allegation (which I am attempting to refute) that purely meaning-theoretic derivations cannot be given.
interpretation of some sentence, u, of a foreign language would be dictated by a conjectured set of interpretations of its words, it suffices to assume that meaning is compositional. It suffices to find a sentence of our own that possesses the same syntactic structure as u, but where that structure is instead imposed on the words of our own that are dictated by the conjectured interpretations of the foreign words.

We don’t have to deduce anything of the form “u is true iff p” from those conjectures (or from their referential implications). We don’t have to worry about not being able to reformulate u in logic, so that such a deduction is even conceivable. And, if we do happen to be able to deduce a truth condition for u’s logical reformulation, we don’t have to worry about the fact (discussed in Section 4 above) that many other truth conditions – yielding different interpretations of u – will also be deducible.

Moreover, there is no need to explain the compositionality of meaning – either by analysing sentence-meanings in terms of truth conditions, or in any other substantive way. Indeed, there is no reason to suppose that the “explanation” (if we can all it that) is anything more than this: that what constitutes a non-idiomatic complex expression’s meaning is its property of being the result of imposing such-and-such structure on words with such-and-such meanings.

Finally, we can easily reject Davidson’s account of compositionality whilst respecting his principle of charity. For the idea that a good interpretation of foreign language, L, should optimize agreement between L-speakers and ourselves coheres perfectly with the deflationary method of deriving the interpretations of complex expressions from the interpretations of their words.

I think we can conclude that the truth-theoretic picture has nothing to recommend it. In fact there’s quite a lot to be said against it.

7 Truth

We have been comparing Davidson and Wittgenstein on meaning and interpretation. I’ve suggested: (i) that the main difference between them concerns compositionality, and (ii) that once Davidson’s mistaken account of this phenomenon is replaced with the common-sense one, then his methodology of interpretation becomes pretty much the same as Wittgenstein’s. In further support of these conclusions, let’s now turn to what these philosophers say about concept of truth.

Here again the received view is that they are on opposite sides: that Wittgenstein, the deflationist, explains truth in terms of meaning via a trivial equivalence schema (“The proposition that p is true ≡ p”), whereas Davidson, the inventor of truth-theoretic semantics, must (it would seem) deploy an explanatorily potent, inflationary concept of truth.
Here again this common view of the difference is actively promoted by Davidson himself. He stresses that whilst Tarski defines “true-in-L” on the basis of assumptions about what the expressions of \(L\) mean, he (Davidson) advocates the opposite explanatory order whereby the notion of meaning is theorized in terms of a primitive notion of truth.\(^{22}\) Accordingly, deflationism is subjected to a variety of criticisms and adamantly rejected.\(^{23}\)

And I would like to suggest that, here again, Davidson misunderstands his own position – failing to see that, despite his heavy reliance on truth to articulate his views, there’s nothing inflationary in any of those deployments.

I’ll now try briefly to substantiate this claim: that each of the points at which truth is brought into Davidson’s picture is quite consistent with its being merely a device of generalization, fully governed by the principle that each attribution of truth to a given proposition is equivalent to the proposition itself.

First: let’s return to his dictum that a good interpretation must be charitable – must result in the foreigner’s deepest beliefs and belief-forming practices coming out as true and truth-promoting. As we saw in section 2, this particular use of “true” is redundant. For the talk of “truth” can be wholly eliminated in favour of “agreement with us”. The intended requirement (roughly speaking) is that a good manual of interpretation should yield the result that what speakers of the foreign language would believe in given evidential circumstances is the same as what we would believe in those circumstances.

Second: consider the way that “true” is deployed in Davidsonian derivations of the truth conditions of sentences from the referents of their component words. As Michael Williams has observed, the validity of those deductions is entirely consistent with deflationism.\(^{24}\) Conditions on deflationary-truth logically follow, à la Tarski, from conditions on deflationary-reference and deflationary-satisfaction.

Third: what about Davidson’s thesis about how to establish that a certain sentence, “p”, of ours provides the correct interpretation of a given foreign utterance, \(u\)? We can do so, he says, by (and only by) discovering that the optimally agreement-promoting theory of truth for the language of \(u\) entails that \(u\) is true iff \(p\). But the plausibility of this thesis in no way presupposes that the operative notions of reference and truth be inflationary. On the contrary, it’s easy to see that whenever our naïve method of interpretation yields the result that

Since our names and predicates “\(n_1\)”, “\(n_2\)” … “\(f_5\)” … “\(f_9\)”, interpret the foreign words \(w_1\), …, \(w_9\), it follows that our “\(p\)” interprets the foreign sentence, \(u\).

\(^{22}\) See Introduction to Inquiries into Truth and Interpretation, p. xiv: “… while Tarski [explained] truth in terms of meaning. I have in mind the reverse”.

\(^{23}\) See Davidson’s “The Folly of Trying to Define Truth” (1996, pp. 263–78).

\(^{24}\) See Michael J. Williams’ “Meaning and Deflationary Truth” (2005).
then any adequate, optimal truth-theory for u’s language will have to be such that “u is true ↔ p” follows from the axioms “(x)(w₁ refers to x ↔ x=n₁)” and “(x)(w₄ is true of x ↔ f₅x)”, etc. – where the semantic concepts are deflationary.

Fourth and finally: we shouldn’t forget Davidson’s penchant for speaking of the ultimate grounds for deciding between competing manuals of interpretation as being “the sentences people hold true, and the circumstances in which they do so”. Pretty obviously he doesn’t mean that in holding true (for example) “dogs bark” one is explicitly thinking, “‘dogs bark’ is a true sentence”. No, one is simply thinking, “Dogs bark”. Thus the “true” in Davidson’s “holds true” is redundant. He could just as well have said that his basic data include, for example, that S holds “dogs bark” (or that S accepts “Dogs bark”) – where someone’s holding (or accepting) a sentence is explained in terms of its causal relation to what S says and in its role in S’s practical and theoretical reasoning. Evidently, his deployment of truth in “holds true” can provide him with no good reason to think that an inflationary account of the concept is required.²⁵

8 Conclusion

I’ve been reacting against the idea that Davidson’s view of meaning, given its explanation of how our understanding of sentences depends on our understanding of its component words, is a substantial improvement on Wittgenstein’s position. As we’ve seen, Davidson’s truth conditional account of compositionality is actually a step backwards. It is not needed (since the naïve deflationary picture is perfectly adequate), it’s plagued with difficulties, and it dictates a contorted formulation of his methodology of interpretation – a formulation that obscures the Wittgensteinian use-theoretic heart of it. Another mistake on Davidson’s part is his opposition to Wittgenstein’s deflationism with respect to truth. Here again, he embraces an implausibly inflated position, wrongly believing it to be demanded by his own more fundamental commitments.

It’s only because these point have remained obscure that so many philosophers could place Wittgenstein and Davidson on opposite sides of the big fence in semantics – indeed, as paradigmatic representatives of the two main camps. But I see Davidson as sitting on the fence, and rather uncomfortably. It’s

²⁵ Another place at which Davidson thinks he needs an inflationary concept of truth is in articulating his view that, in the case of certain sentences (for example, “That is red”), the condition for our being justified in accepting the sentence is that it be true. (See his “The Folly of Trying to Define Truth”). But even if that view were correct, the acceptability condition for each such sentence could perfectly well be stated without any mention of truth. One might say, for example, that “That is red” is acceptable only if the demonstrated object is red. The only advantage of bringing in the concept of truth is that we can generalize over all such sentences: “None is acceptable unless true”. But this is precisely the function of truth that’s emphasized by deflationists and explained by the deflationary equivalence schema.
somewhat ironic that the genuine (i.e. thoroughgoing) truth-theoretic semanticists were inspired by someone who was not really one of them.26

In sum: Davidson’s two-pronged account of meaning combines the good, the bad, and the ugly. The good is the ‘meaning as use’ picture of what makes for a plausible interpretation – a picture that’s visible (but with some difficulty) in his principle of charity. The bad is his truth conditional account of the compositionality of meaning. And the ugly is the ungainly result of sticking them together.27

References

Davidson, Donald. *Inquiries into Truth and Interpretation*. Oxford University Press, 1986.


---

26 Amongst philosophers, I’m thinking of Ernest Lepore, David Lewis, Kirk Ludwig, John MacFarlane, Jason Stanley, Timothy Williamson, and Seth Yalcin, amongst others. The thorough-going truth-theoretic approaches to semantics that they advocate strike me as in even worse shape than Davidson’s. For they contain his objectionable view of what’s needed to account for compositionality. But, in contrast with Davidson’s broader position, they reject any sort of use-based conception of meaning; so they lack the resources needed either to contribute to an adequate empirical account of linguistic activity, or to explain the standard methodology of translation.

27 I am most grateful to Claudine Verheggen for encouraging me to write this paper, for giving me helpful reactions to an earlier draft of it, and for inviting me to her “Wittgenstein and Davidson” workshop in Toronto (under the auspices of York University). I presented my ideas on the topic at that meeting (in March 2015) and received valuable feedback from the participants, especially from Bill Child, Kathrin Gläuer, Bernard Katz, Åsa Wikforss, and Meredith Williams. For further insightful comments I should also like to thank Stephen Schiffer and Michael Williams.
In *Meaning and Truth* (1970) P.F. Strawson famously contrasts two approaches to the question of what it is for words to have meaning: That of communication-intention theorists (represented by Grice, Austin, and the later Wittgenstein) and that of formal semantics theorists (represented by Chomsky, Davidson, Frege and the earlier Wittgenstein). The conflict between communication-intention theories and formal semantics, Strawson argues, is a conflict concerning what role the notion of communication is to play in our philosophical account of meaning: Theories of the former sort place the notion of communication at the center, whereas formal semantics theories fail to do so. While Strawson holds that both approaches have their merits, his aim is to show that the communication-intention theorist is closer to the truth. Only a theory that takes *conventions* to play a central role in the account of meaning will be able to secure the essential link between linguistic meaning and communication.

According to Strawson, therefore, the later Wittgenstein and Davidson end up on opposite sides in this struggle. Several contemporary Wittgenstein scholars agree, among them Hans-Johann Glock and Meredith Williams. According to them Wittgenstein puts forth an essentially social picture of language, with the shared conventions at the center, while Davidson defends an individualistic picture that ultimately fails to account for the public nature of language. I shall argue that this description is importantly mistaken. Naturally, there are many significant differences between Davidson and Wittgenstein. For instance, Davidson believed in the project of formal semantics – the project of trying to formulate a systematic theory describing the semantics of a language – and Wittgenstein did not. However, when it comes to the foundational or metasemantic question, Davidson and Wittgenstein are close: They both subscribe to the idea that meaning is determined by use, rather than by conventions, and they both take meaning to be essentially public and tied to its role in

1 Reprinted in Strawson 1971.
communication. There is no struggle here. Strawson is simply wrong to suggest that securing the public nature of meaning requires appealing to conventions.

The chapter is divided into three sections. The first sets the stage by presenting Davidson’s view on conventions and communication. In the second I sketch what I take to be the proper picture of the development of Wittgenstein’s view of meaning and rules leading up to the rule-following remarks in the *Investigations*. I argue that these remarks in fact are directed against the idea that language is an essentially rule-guided activity and I criticize the alternative, conventionalist reading of Wittgenstein defended by Glock. In the third section I discuss whether, nevertheless, there is a sense in which Wittgenstein takes the shared social practice to be essential to meaning, focusing on the arguments of Glock and Williams. In particular, I discuss whether Wittgenstein’s appeal to agreement shows him to be a conventionalist about meaning.

1 Davidson on Conventions

It is commonly taken for granted that securing the public nature of language requires an appeal to shared conventions. This assumption is quite clear in Strawson’s reasoning. On a truth conditional account of meaning, he says, the relevant semantic rules are not rules for communicating since they simply determine the truth conditions of the sentences of the language. This implies, Strawson argues, that someone may “understand a language competently”, have perfect linguistic competence, without using language as a means of communication (Strawson 1971, 172). It would follow that it is a contingent truth that the rules that determine the meanings of the sentences of a language are social or conventional, and that there is no principled reason why it could not be the case that “every individual might have his own language which only he understands” (Strawson 1971, 187). In order to avoid this, according to Strawson, the semantic rules must be understood in terms of the speaker’s communication intentions – as conventionalized ways of using language with the purpose of letting the audience know that one has a certain belief.

However, the claim that a truth conditional semantics implies that a speaker could have a language without being able to use it in communication, reflects a failure to keep the different semantic projects apart. Securing the public nature of meaning need not go via the semantics, construing meaning in terms of ‘rules for communicating’, but could go via the metasemantics. In particular, it is quite possible to combine a truth conditional theory of

---

2 For the distinction between semantics (sometimes called descriptive semantics) and metasemantics, see Stalnaker 1997. Semantic theories assign semantic values to the expressions used whereas metasemantic theories provide an account of the facts in virtue of which the expressions have certain semantic values.
meaning with a metasemantic theory of meaning that puts communication at the center. No one provides a better illustration of this than Davidson.

At the foundation of Davidson’s philosophy of language, famously, lie two questions: What could we know that would enable us to interpret a speaker’s words? And how could we know it? (Davidson 1973b, 125). These two questions correspond to the project of semantics on the one hand, and to the metasemantic project on the other. Davidson’s answer to the first question involves an appeal to a recursive Tarskian truth theory, and his answer to the second an appeal to radical interpretation. That is to say, Davidson holds that by providing an account of how the radical interpreter could reach an interpretation of a speaker’s words ‘from scratch’, without having any prior knowledge of the language spoken, we can reach an answer to the question of what determines meaning. By putting the interpreter at center, Davidson brings out his commitment to the idea that meaning is public: There is nothing more to meaning than what a radical interpreter can discern (once all the evidence is in). The publicity of meaning therefore lies at the foundation of Davidson’s project and is built into his metasemantics through the device of the radical interpreter.

The basic evidence for the interpreter is the speaker’s holding uninterpreted sentences true (and false) in various contexts. However, what a speaker holds true is the result both of what she believes and of the meaning of her words. The interpreter therefore needs to solve what Davidson calls the problem of the interdependence of meaning and belief. For instance, if Kurt utters ‘Es regnet’ when it is snowing there are two possibilities: He is mistaken about the weather (and his sentence is true if and only if it is raining) or his sentence is true if and only if it is snowing. How can the interpreter, on the basis simply of Kurt’s assent, determine which is the correct interpretation? Davidson proposes that the principle of charity can solve the problem by making assumptions about the nature of belief: On the assumption that there is large overlap in beliefs between speaker and interpreter, such that the speaker (by and large) holds a sentence true when it is true (by the interpreter’s lights), and that the speaker is (by and large) rational, the belief component can be held ‘steady’, allowing the interpreter to take the speaker’s holding a sentence true as evidence that it is true.3

The principle of charity therefore functions as a principle of meaning determination, mapping facts about use on to meanings in such a way that the speaker comes out as (by and large) rational and holding true beliefs.4 It cannot be applied piecemeal, however, and is not meant to eliminate error. Taking Kurt’s overall use into account, it may be that the interpretation that provides the best fit with the evidence, minimizing inexplicable error, entails that Kurt’s

---

3 For a discussion of the problem of meaning and belief and the principle of charity see for instance Davidson 1973a, 1973b, 1974a and 1974b.
4 Cf. Kathrin Glüer’s contribution to this volume, Chapter 4.
utterance is mistaken. The aim, Davidson writes, is not the absurd one of making disagreement and error disappear:

The point is rather that widespread agreement is the only possible background against which disputes and mistakes can be interpreted. Making sense of the utterances and behavior of others, even their most aberrant behavior, requires us to find a great deal of reason and truth in them. (Davidson 1974b, 153)

While Davidson takes meaning to be essentially public, he denies that this implies that conventions are essential to meaning.5 The attempt to account for meaning in terms of conventions, Davidson says, is an expression of the view that there is a connection between linguistic meaning and human attitudes described in non-linguistic terms. The hope is that by appealing to the speaker’s ulterior purposes (purposes that can be characterized without relying on the notion of literal meaning) we can get an account of literal meaning (Davidson 1982, 266). This, as we have seen, is Strawson’s strategy. Davidson is happy to grant that there are important connections between what a speaker’s words mean and his non-linguistic intentions and beliefs. However, he does not think these connections can be used to provide an account of meaning. One reason for this derives from considerations having to do with the public nature of meaning.6 Discussing the strategy of explaining linguistic meaning on the basis of non-linguistic intentions (which, as it happens, he ascribes to Wittgenstein), Davidson argues that this fails since we cannot make detailed sense of a person’s intentions and beliefs independently of making sense of his utterance (Davidson 1974b, 143). He puts the point in terms of radical interpretation, arguing that when it comes to radical interpretation there is no hope of appealing to the speaker’s intentions and beliefs as part of the evidence: “If this is so, then an inventory of sophisticated beliefs and intentions cannot be the evidence for the truth of a theory for interpreting his speech behavior” (Davidson 1974b, 144).

It is natural to object that, nevertheless, there are linguistic regularities and these are essential to communication. For this to be an interesting claim, Davidson stresses, it cannot simply amount to the truism that successful communication implies that speaker and hearer have understood the speaker’s words the same way. Rather, the claim must be that communication requires that speaker and hearer assign the same meaning to their words in advance, prior to their linguistic interaction. And this, Davidson argues, is

5 Davidson raises a range of objections to this thesis, especially in “Communication and Convention” (1982). I shall focus on the objections that concern the public nature of meaning and its role in communication.

6 Another reason concerns what Davidson calls ‘the autonomy of meaning’; the idea that meaning need be independent of ulterior purposes since a sentence can be uttered with any number of such purposes without its meaning changing (Davidson 1982).
simply not the case. Shared conventions are neither sufficient, nor necessary for communication. That shared conventions are not sufficient, Davidson suggests, is illustrated by the fact that there is a multitude of situations where a speaker uses her words in a non-standard way and yet manages to communicate, such as cases of malapropisms, neologisms, slips of the tongue, and simple misunderstandings. In these cases communication requires going beyond the standard meaning, and we normally do this without trouble. For instance, when television’s sitcom character Archie Bunker utters ‘Let’s have some laughter to break up the monogamy’ we have no trouble understanding that he is suggesting that we should have some laughter to break up the monotony. That shared conventions are not necessary is shown by the possibility of radical interpretation, where the interpreter manages to understand the speaker without any shared conventions being in place. Naturally, for the speaker to be interpretable he needs to use his words in fairly regular and stable ways, but this does not mean that there need be any shared regularities involved.

Davidson’s rejection of the appeal to conventions, therefore, is not a reflection of his not taking the role of meaning in communication seriously. On the contrary, it is precisely because he takes the central function of meaning to be that of explaining successful communication that he denies that conventions are essential to meaning: Since there can be successful communication without shared conventions, it follows that linguistic meaning cannot be identified with conventional meaning. Davidson does not thereby deny that language is a social art, or that people adjust their language use to cohere with that of others. He merely holds that this does not tell us anything essential about the nature of language or linguistic communication (Davidson 1982, 278). Nor does he deny that shared regularities make for ease of communication and therefore are of great practical importance. Common conditioning ensures, up to a point, that the same method of interpretation that we use for others will work for a new a speaker. As Davidson puts it, knowledge of shared regularities is a ‘crutch’ to interpretation – a crutch we could in theory have done without from the start (Davidson 1982, 279).

While Davidson is squarely in the anti-conventionalist camp, therefore, he nevertheless holds meaning to be essentially public. Strawson’s picture of the struggle is misleading since it presupposes that rejecting the importance of conventions is equivalent to giving up on the essential link between meaning and communication. Next, let us see which camp Wittgenstein belongs to.

---

7 See especially Davidson 1986.
8 Cf. Glüer 2012. Discussing Davidson’s view on conventions she writes: “It is because linguistic meaning is essentially public that neither convention nor any other form of shared regularity in the use of linguistic expressions is necessary for successful communication” (2012, 339).
To do this, I shall start by critically examining Hans-Johan Glock’s reading of the later Wittgenstein as a convention theorist.

2 Wittgenstein on Conventions

Part and parcel of Glock’s account of the later Wittgenstein as a convention theorist is a certain interpretation of the development of Wittgenstein’s views on rules from the *Tractatus* on (Wittgenstein 1922). Here, in brief, is Glock’s story (1996, 2010). In the middle period, linguistic norms come to play a central role in Wittgenstein’s philosophy. For example, in texts from the early 1930s Wittgenstein stresses that we do not regard language from the perspective of a mechanism, but from that of a calculus, a normative perspective. The big difference between the *Tractatus* and Wittgenstein’s middle period is that Wittgenstein comes to reject the idea of a complex set of inexorable norms hidden behind the surface of natural languages, moving towards a picture of the relevant norms as conventions guiding our linguistic use. Wittgenstein did therefore not abandon the *Tractatus* appeal to linguistic rules, but modified it and insisted that rules cannot guide linguistic behavior without being known to us. It is for this reason that Wittgenstein suggests the comparison with language and games: this analogy was meant to bring out that linguistic understanding involves mastery of techniques concerning the application of rules, stressing the social dimension of using language. The *Tractatus* conception of rules was therefore replaced in the middle-period texts with a down-to-earth conception of rules, rejecting the intellectualist view of rule-following as involving the consultation and interpretation of rule formulations. This culminates in Wittgenstein’s rule-following considerations in the *Philosophical Investigations* 201, and the notorious paradox that “no course of action could be determined by a rule, since every course of action could be made out to accord with the rule”. To avoid the paradox we just have to give up the idea that being guided by a rule involves interpretation, and to see that it involves a commitment to a shared practice.

There is much that is wrong with this picture of the development of Wittgenstein’s thoughts. In a joint paper, Kathrin Glüer and I have tried to spell this out (2010). Here are the essentials of the alternative picture we argued for.

First, it is not clear that norms play any role at all in the *Tractatus*. While Wittgenstein appeals to the rules of logical syntax these do not function as norms or guiding rules: Wittgenstein does not distinguish between rules and laws and he suggests that the question of how we follow these rules is philosophically irrelevant (see for instance *Tractatus* 4.0141 and 4.002). When rules come to the center in the middle period texts, this is not because he now has a different conception of rule-following, but because he is starting...
to question logical atomism, in particular the idea that all elementary propositions are independent of one another. He argues that there needs to be rules spelling out the internal connections between propositions and he proposes that meaning is to be understood in terms of a set of grammatical rules, a ‘calculus’: “I want to say the place of a word in grammar is its meaning” (Wittgenstein 1974, § 23).

Second, the picture of meaning and rules that emerges in the middle period stands in contrast not only to the *Tractatus* but also to the later texts. Glock is simply wrong to suggest that there is a smooth transition from the middle period view on rules to that of the *Investigations*. And the change concerns precisely the idea that meaning can be understood in terms of guidance by rules. In the *Tractatus*, Wittgenstein could dismiss the question of rule-guidance as philosophically irrelevant, since the rules of logical syntax were said to reflect the intrinsic nature of the world and were not of our own making. Once a sign was projected on to an object, the system of rules would simply kick in, whether the speaker was aware of it or not. By contrast, the rules of grammar in the middle period texts are said to be conventional or arbitrary, like the rules of a game. They are ‘autonomous’, not accountable to reality, since they themselves determine the meaning of the expression used. If so, however, it is essential to say something about the facts in virtue of which certain conventions hold rather than some other, equally arbitrary, conventions. Wittgenstein’s proposal is that they do so in virtue of the speaker’s following these rules. My use of signs, he says, is in accordance with any number of rules, but what makes representation possible is the fact that I am *guided* by a particular rule, that a particular rule is *involved* in my use. Only then can there be a distinction between acting correctly (in accordance with the rule) and acting incorrectly. As a result, it becomes essential to Wittgenstein to say something about how meaning rules guide speakers. And this is what gets him into trouble.

In order to account for rule-guidance, Wittgenstein appeals to the central role of speaker intentions. While the speaker’s use of signs may be in accordance with any number of rules, Wittgenstein suggests, the rule that he follows is the rule he *intends* to follow. The intention contains the general rule determining whether the application is correct or incorrect (Wittgenstein 1974, §58). “If you exclude the element of intention from language”, Wittgenstein writes, “its whole function then collapses” (Wittgenstein 1975, §20). The trouble is that at the same time he is arguing that following a rule cannot be accounted for in terms of mental events and processes. He argues that we do not have rules running through our heads and, moreover, that even if we did this would

---

11 See also Wittgenstein 1974, §95–99.
involve rule expressions, such as charts, and these can always be interpreted in different ways.\textsuperscript{12} Wittgenstein here anticipates the rule-following discussion in the *Investigations*, arguing that appealing to rules in the mind of the speaker would lead to an infinite regress of one rule expression interpreting the next. How, then, could an appeal to intentions help?\textsuperscript{13}

It is against this background, Kathrin Glüer and I have suggested, that we need to read the famous remarks on rule-following in the *Investigations*. We think Wittgenstein is quite right to suggest that intentions are essential to rule-guidance. The difference between acting in accordance with a rule and being guided by it, we have argued, consists in the rule playing a role in the speaker’s reasoning, in her intending to do what the rule requires – what we have called the ‘intention-reason view’ of rule-following (or the IR-view) (Glüer and Wikforss 2010, 157–158). However, our claim is, precisely because of this we should question the idea that rule-guidance underlies the very basis of all human activity, thought and language. Although much of human life is rule-guided, that which makes rule-guidance possible in the first place, is not itself based on guidance by rules. And we have proposed that this is precisely the conclusion that Wittgenstein came to in his famous rule-following remarks in the *Investigations*. The very point of these remarks is that speaking a language is not an essentially rule-guided activity: what determines meaning is not that speakers are guided by certain rules – that would lead to a regress of interpretations of the very sort that Wittgenstein struggles with in the middle period texts. Instead, what determines meaning is use, the practice of applying terms. This, we have suggested, is the real significance of his much discussed remark in *Investigations* §198 that interpretations by themselves do not determine meaning, and it is the real import of his dictum that meaning is use.\textsuperscript{14}

By contrast, Glock’s claim that there is a smooth transition from Wittgenstein’s middle period view on meaning and rules to the *Investigations* faces some serious textual challenges. If indeed the later Wittgenstein thought that conventions were essential to meaning one would


\textsuperscript{13} Wittgenstein’s comment in *Investigations*, p. 205 can be read in the light of this, as referring to a view he once held himself: “But it is just the queer thing about intention, about the mental process, that the existence of a custom, a technique, is not necessary to it”.

\textsuperscript{14} Horwich 2012 suggests a rather different interpretation according to which Wittgenstein’s dictum should be understood as a perfectly obvious definition, as trivial as the synonymy of ‘bachelor’ and ‘unmarried man’ (2012, 105). This is implausible for several reasons. First, one would be hard pressed to find a dictionary which takes ‘meaning’ to be synonymous with ‘use’. Second, Horwich at the same time ascribes to Wittgenstein the thesis that meaning is determined by use (2012, 107, 109), where the determination relation is understood as an ‘in virtue’ relation. This is in line with what I am suggesting here but it is not, of course, a mere trivial definition but a substantial metasemantic theory.
have thought that he would mention that or, at least, that the notion of conventions would play a central role in the *Investigations* (as it does in the middle period texts) but conventions are barely mentioned.\(^\text{15}\) Not surprisingly, when Glock cites textual evidence in support of his claim that the later Wittgenstein takes conventions to be essential to meaning, it is all coming from the middle period texts, such as *Waismann Dictations* (early 1930s), *Philosophical Grammar* (1933), Wittgenstein’s Cambridge Lectures 1932–35, and *The Big Typescript* (1933).

Moreover, the claim that Wittgenstein held on to the idea that meaning involves guidance by rules, is in tension with other central ideas in the later writings. First, Glock accepts the IR-conception of rule-following – i.e., that if an agent follows a rule in doing A, the rule must be part of his reason for doing A – and suggests that this was Wittgenstein’s view. The rule, Glock writes, must be part of the agent’s reason for acting: “He must intend to follow the rule” (Glock 1996, 325). The question is how Glock thinks it can be combined with another central theme in the *Investigations* – Wittgenstein’s rejection of the Augustinian idea that thought is prior to language (Wittgenstein 1953, §32). If linguistic meaning is determined by the speaker following certain rules, and rule-following requires intentions, then thought must be prior to, and independent of, language. Second, Glock even suggests that following a rule involves being able to cite the rule as a reason, justifying his action (Glock 1996, 325). However, it is a central idea in Wittgenstein’s later philosophy that we are not able to cite the rules of grammar. Indeed, this is precisely why we get into trouble philosophically: “Our grammar is lacking in this sort of perspicuity” (Wittgenstein 1953, §122). Something has to give. And, Kathrin Glüer and I argue, what has to give is the assumption that meaning is determined by the speaker’s being guided by certain rules.\(^\text{16}\)

These tensions remain in Glock’s more recent papers on Wittgenstein. Glock 2010 defends a conventionalist view of meaning (ascribing it to Wittgenstein) against an individualist view (which he ascribes to Davidson). Glock defines a convention as a shared, arbitrary rule: It is a *rule* because it provides reasons

\(^{15}\) One of the few places where ‘convention’ occurs in the *Investigations* is paragraph 355 where Wittgenstein writes of the language of sense-impressions that “this language like any other is founded on convention”. If one looks to the German, however, the word used is ‘Übereinkunft’ which simply means agreement. Glock notes that the term employed is ‘Übereinkunft’ and suggests that “it is even more blatant in its intersubjective connotations than *Konvention*” (2010, 101). This may well be, but intersubjectivity is distinct from rulishness. Below I shall return to the role of agreement in Wittgenstein’s later writings.

\(^{16}\) Although Wittgenstein holds on to the idea of ‘rules of grammar’, and develops in up until *On Certainty*, these emerge not as guiding rules that speakers follow, as conventions, but are construed as very basic judgments that play a special role in our language insofar as they cannot be given up without a change in meaning. As philosophers we fail to notice the special role of these judgments and this leads to conceptual confusions.
for regular behavior either directly or because of sanctions, it is *shared* because deviations are subject to normative reactions, and it is *arbitrary* because a different rule might have been adopted. The appeal to sanctions suggests lowering the bar for rule-following, since it suggests that the rule may provide a reason ‘indirectly’, and so need not be involved in the subject’s reasoning in any high-profile sense. And Glock argues that although applying sanctions does presuppose having some primitive intentional attitudes towards the behavior of others, sanctions can be understood without reference to intentions, as mere adverse reactions to a deviation from a regularity. The idea, thus, is that there can be *implicit* rules that guide behavior, without having been consciously consulted (Glock 2010, 94–96).\(^\text{17}\)

However, a subject may perform an action because she is afraid of sanctions if she does not do so, even if there is no rule involved at all.\(^\text{18}\) If indeed Glock wishes to hold on to the important distinction between acting in accordance with a rule and being guided by it, then the rule has to enter into the explanation of what the subject did, as a reason – just as his IR-view suggests. And as Kathrin Glüer and I have argued elsewhere, the appeal to implicit rules does not help (Glüer & Wikforss 2009, 59–63). No doubt, that a rule functions as a reason does not require that it is very explicitly formulated. However, as Quine stressed, if linguistic conventions determine meaning they could not even be formulated prior to their adoption and this undermines their explanatory force. Things are even worse if one, like Wittgenstein (and Davidson), denies that thought is prior to language. Then the conventions are such that the agent cannot even *intend* to follow them prior to having a language.\(^\text{19}\) What, then, remains of the crucial difference between being guided by a rule and merely acting in accordance with one?

Not only, therefore, is the textual evidence in support of Glock’s conventionalist reading of the later Wittgenstein scant. This reading is also in tension with ideas of Wittgenstein’s that there *is* textual evidence for. Much more plausible, I submit, is the conclusion that although Wittgenstein once held that meaning is determined by the speaker’s being guided by rules, he later came to reject this very idea.

\(^{17}\) Cf. Horwich 2012, 118–122.

\(^{18}\) Moreover, the appeal to sanctions does not solve the regress problem (assuming that is Glock’s intention) since the question remains whether a particular sanctioning was done correctly. See Glüer & Wikforss 2009a, 62–63.

\(^{19}\) Glock also suggests that we should reject Davidson’s claim that beliefs and intentions presuppose language: “nonlinguistic creatures can have not just beliefs and intentions, but also beliefs and intentions concerning the beliefs and intentions of others” (Glock 2010, 95). This of course eliminates all worries concerning how conventions determine meaning. However, it also means that Wittgenstein’s regress argument is stopped short. It is presumably because this is an idea at odds with Wittgenstein that Glock falls back on the appeal to implicit rule-following.
3 Meaning and Agreement

3.1 The Argument from Normativity

In the discussion following the paradox of paragraph 201, Wittgenstein famously suggests that human agreement is essential to language. For example, in paragraph 242 he writes: “If language is to be a means of communication there must be agreement not only in definitions but also (queer as this may sound) in judgments”. This has been taken by some of Wittgenstein’s interpreters to show that on his view meaning is essentially social. Wittgenstein’s idea, it is suggested, is that what is required to secure the distinction between correctness and incorrectness is precisely an appeal to human agreement – that is, to the larger social practice of using a language. Even, therefore, if it is right that Wittgenstein does not hold that speaking a language requires being guided by rules, it is nevertheless the case that he takes the shared, social practice to be essential to linguistic meaning. Coming back to the struggle at hand, if this reading is correct, there is still a fundamental conflict between the outlook of Davidson and that of Wittgenstein: while Wittgenstein takes the shared social practice to be essential to meaning, Davidson denies this.

This is the line defended by Meredith Williams (1999, 2000). Williams acknowledges that Davidson pays attention to the social aspect of language, but she argues that his conception of the sociality of language is tied to a very different idea than Wittgenstein’s conception of this. Whereas Wittgenstein endorses the “practice view”, requiring shared conventions and conformity of behavior among the participants, Davidson endorses the “interpretive view”, requiring merely mutual interpretability. Davidson appeals to the public character of language but only in the sense that language use provides evidence for the interpreter, while Wittgenstein takes language use to constitute meaning and understanding. This, Williams suggests, reflects profound differences in their conceptions of language: “Davidson’s account of language remains highly intellectualized and individualist, wedded to the view that the defining mark of language is its inherent systematicity and infinite productivity. Wittgenstein’s picture, on the other hand, emphasizes the nonrational basis of language, and our shared practical ‘conventions’” (Williams 2000, 301). And, she argues, Wittgenstein was right – language is essentially social.

Williams’s central argument in support of this turns on considerations having to do with how genuine normativity arises. An isolated individual could behave in all sorts of complex ways, she argues, but there would be no room for normativity unless her actions could be measured against the larger practice of the community. There can be norms, standards, only against the background of group harmony. That is the point Wittgenstein is making when appealing to the importance of agreement. Wittgenstein’s central claim, according to
Williams, is that the very idea of normativity, the structure within which the distinction between correct and incorrect can be drawn, cannot get a foothold unless the practice is a social one: “conformity of behavior is required for a normative contrast between correct and incorrect to get a foothold” (Williams 2000, 312).

A similar idea can be found in Glock 2010. Like Williams, Glock suggests that there is a fundamental conflict between Davidson and Wittgenstein and that this conflict turns on the role of the social practice. And like Williams, Glock stresses the essentially normative nature of meaning, distinguishing between two normativity theses: bare normativity of meaning (BNM) and rule-based normativity of meaning (RNM). According to (BNM) meaning is normative since there is an essential connection between meaning and correctness conditions – if a word is meaningful there must be conditions for its correct application. Thesis (RNM) is stronger, and requires that correctness conditions are determined by rules guiding the speaker’s use of terms. According to Glock, again, Wittgenstein subscribes to both theses, and it is in order to account for rule-guidance that he appeals to the essential role of the social practice (Glock 2010, 90). The point is not that speaking correctly can be identified with speaking as the majority does, but that “literal meaning is bound up with the notion of a linguistic mistake as a deviation from a shared norm, and hence with the notion of a convention” (Glock 2010, 111).

The distinction between (BNM) and (RNM) corresponds to the distinction drawn in Glüer and Wikforss 2009, between (ME)-normativism and (MD)-normativism. The first thesis says that facts about meaning (in themselves and essentially) have implications for how a speaker should use her terms – meaning engenders certain norms. The second thesis is a meta-semantic thesis, according to which meaning is determined by the speaker’s following certain rules. I have already argued against the suggestion that Wittgenstein subscribes to MD-normativism, i.e., to Glock’s thesis (RNM). The question now is whether he subscribes to ME-normativism (thesis BNM) and whether this shows that he takes the social practice to be essential to meaning.

Leaving Wittgenstein for a moment, let us consider the argument as such – i.e., the claim that because there must be semantic correctness conditions,

---

20 See also Williams 1999, 174–75.
21 It is not easy to see precisely what thesis Glock wishes to ascribe to Wittgenstein. Unlike Williams, he rejects the community reading of Wittgenstein according to which Wittgenstein held that meaning is necessarily social. At the same time he suggests that conventions are essential to meaning and that Wittgenstein regarded rule-following and language as typically social (Glock 2010, 101). However, that language is typically social is an empirical claim – a claim that no one will dispute and that does not tell us anything about the nature of meaning.
22 Glock does not spell this out in terms of oughts but in terms of permissions. Nevertheless, (BNM) qualifies as a type of ME-normativism since the claim is precisely that meaning in itself, and essentially, has normative implications (see footnote 25 below).
meaning is normative in a sense that makes it essentially dependent on the larger social practice. It seems plausible enough that there is a certain type of normativity that can only arise in a community context. Consider Robinson Crusoe, only a wilder one who has never lived in a human society. When he eats he fails to live up to all human conventions of eating and eats like an animal. Is his eating incorrect? Does he violate any conventions of eating? Hardly. His way of eating is neither correct, nor incorrect unless we think of him as being part of a social practice. The question is, can a similar reasoning be applied to the issue of semantic correctness conditions: Is it the case that the distinction between correct and incorrect usage requires an appeal to the wider social practice?23

The decisive question, it should be clear, is whether there is a connection between semantic correctness and normativity, as Glock’s thesis (BNM) presupposes. What is beyond dispute is that there is an essential connection between linguistic meaning and semantic correctness conditions. This is just a way of capturing the basic semantic relation between word and world: If an expression is meaningful, then there is a difference between the conditions under which the term does apply and the conditions under which it does not. As Wittgenstein puts it, there must be a difference between what is right (semantically correct) and what merely seems right to the speaker. However, Kathrin Glüer and I have argued, the fact that an expression has semantic correctness conditions does not have any normative implications.24 Thus, we have questioned the move from a statement of correctness conditions, such as (C), to a normative statement, such as (N):

(C) For any speaker S, and any time t: if ‘green’ means green for S at t, then it is correct for S to apply ‘green’ to an object x iff x is green at t.

(N) For any speaker S, and any time t: if ‘green’ means green for S at t, then S ought to apply ‘green’ to an object x iff x is green at t.25

We are happy to grant that there are some construals of “correct” that are normative. The important point, however, is that the relevant notion of correctness in this context is that of semantic correctness and that is not a normative notion. How semantic correctness conditions are construed depends on the choice of the basic semantic concept, such as truth or warranted assertibility. Either way, we have argued, the notion of semantic correctness is non-normative in precisely the sense that no statements about what we ought

23 There is of course one notion of linguistic correctness that does require appeal to the social practice: that of using a word like others do. However, this is distinct from the notion of semantic correctness that is our concern here.


25 There are alternative formulations of (N), for instance in terms of permissions rather than oughts, and Glock follows this (Glock 2010, 98–99). Notice, however, that the anti-normativist rejects these versions of (N) as well, insisting that semantic correctness conditions do not have normative implications of any sort (see below).
(not) to or may (not) do with \(w\) directly follow from (C). Assuming that ‘green’ is true of green objects only, it does not follow that the subject ought to apply ‘green’ in certain ways. If she applies ‘green’ to a red object, she has made a false statement but she has not violated any semantic rules.

Glock shows awareness of this objection, but argues that BNM “marks a straightforward normative dimension of meaning” (Glock 2010, 96–97). The anti-normativists, he suggests, have conflated empirical propositions such as “In English, it is correct to apply ‘drake’ to an object x \(\text{iff}\) x is a male duck” with the norm proposition “It is correct to apply ‘drake’ to an object x \(\text{iff}\) x is a male duck”. Since the latter functions as an explanation it has a normative function, laying down conditions under which ‘drake’ can be meaningfully applied to an object x (Glock 2010, 97). These lexical norms are to be understood as a type of constitutive, permissive norms, giving the speaker the semantic entitlement to apply ‘drake’ to a male duck, and licensing her to draw certain inferences (from “This creature is a male duck” to “This creature is a drake” and vice versa). Someone who applies ‘drake’ to a goat violates this rule, but she may still be said to be speaking English even though the mistake itself does not count as English (Glock 2010, 99).

There is much that is problematic here. What (C) lays down is not conditions under which applying ‘green’ is meaningful, but merely semantic correctness conditions in the sense above (such as truth conditions). Of course, (C) does have something to do with meaningfulness, insofar as it says that if an expression has a certain meaning then it also has certain correctness conditions. That is, without correctness conditions no meaning. But that is very different from saying that the conditions themselves concern the meaningfulness of an application. Because Glock fails to see this, he ends up having to say, awkwardly, that whenever someone makes an empirical error (mistaking a goose for a duck) the mistake does not count as English.

Moreover, by suggesting that correctness conditions entail permissions to apply the term to certain objects and draw certain inferences, Glock simply ignores the anti-normativist argument that no normative consequences (whether in terms of oughts or permissions) follow from statements of correctness conditions, such as (C).\(^{26}\) It does not follow from (C) that \(S\) is permitted to apply ‘green’ in certain ways or to draw certain inferences – it merely follows that if \(S\) applies the term to non-green objects the application is mistaken (and, similarly, that certain inferences drawn by \(S\) will be truth-preserving, others not). In the end, Glock’s argument reduces to the claim that “it is difficult to

\(^{26}\) Notice, also, that in the formulation of (C) there is no conflation of the sort suggested by Glock with the empirical proposition “In English it is correct to apply ‘green’ only to green objects”. What is stated is merely that there is an implication from an expression being meaningful to its having certain correctness conditions – this is not an empirical statement about what a word means in a particular language.
deny that ‘correct’ and ‘incorrect’ are normative terms, terms of appraisal” (Glock 2010, 97). However, as we have noted elsewhere, standard dictionary definitions of ‘correct’ list both normative and non-normative uses (Glüer & Wikforss 2015, 72). More importantly, as stressed above, the decisive question is not whether ordinary usage construes ‘correct’ as a normative notion, but whether the relevant notion of semantic correctness need be construed normatively for the purposes of semantics – and this, we have argued, is not the case.

Now, if one takes this on board and denies that the need for semantic correctness conditions itself shows meaning to be normative, then the appeal to the community practice comes in a different light. One can then grant that there is a certain type of normativity that only arises within the context of the community (such as the normativity of conventions) but deny that this shows anything about the need for a communal practice in order to secure semantic correctness conditions. If the direct argument from correctness conditions to normativity fails, so does the direct argument from correctness conditions to the need for the community. For instance, William’s claim that “the very idea of normativity… cannot get a foothold unless the practice is a social one” (Williams 1999, 175) may well be right but it does not follow that there can be no semantic correctness conditions without the social practice.

Let us return to Wittgenstein. That the direct argument from semantic correctness to the community fails, naturally, does not show that Wittgenstein did not subscribe to this argument: Even great philosophers subscribe to mistaken doctrines. The important question, then, is whether there is some textual evidence suggesting that Wittgenstein in fact held that an appeal to the community practice was essential to account for semantic correctness conditions, i.e., for meaning.

3.2 Wittgenstein on Agreement

Williams’s main evidence derives from Wittgenstein’s claim that there is a way of following a rule which is not an interpretation and his appeal to the role of agreement. She makes her point by drawing a contrast between Wittgenstein and Davidson. Since Davidson construes meaning in terms of interpretation, she argues, his view entails that there is interpretation ‘all the way down’. He therefore fails to recognize that interpretation must be the exception and is only possible against the background of a shared language. Although Davidson does appeal to the importance of agreement in beliefs, Williams argues, he does not see that what is needed is a much more pervasive agreement: there need be shared techniques and practices of language. Davidson’s claim that speaker and interpreter need not share any linguistic norms or conventions therefore undercuts the possibility of interpretation and communication: “Shared beliefs require shared practices and techniques, practices and techniques that
Wittgenstein identifies as crucial to the language game and so part of language” (Williams 2000, 308). This, according to Williams, is the significance of Wittgenstein’s famous appeal to agreement. It is only because our reactions to a given training and to certain stimulus are in agreement, that ‘going on as before’ is possible and thereby the distinction between correct and incorrect applications. At the foundation of language and communication lies not interpretation (that would lead to a regress) but shared bedrock judgments of sameness.

It is clear that Williams misconstrues the role of interpretation in Davidson’s account. His claim is not that meaning is determined by the speaker’s applying a certain interpretation to her words – or by the interpreter doing so.27 On the contrary, as we have seen above, Davidson holds that a plausible theory of meaning determination must proceed without assuming mental content. This is why he takes meaning to be determined by speaker use, her pattern of assent and dissent in a context, and the device of the radical interpreter is meant to illustrate precisely how use determines meaning. The regress argument does therefore not apply to Davidson’s metasemantics. Nor would Davidson object to the idea that Williams ascribes to Wittgenstein, that meaning is constituted by use. It is of course correct that Davidson construes particular applications of terms as evidence for the semantic theory, but the sharp contrast drawn between constitution and evidence is problematic in the context of Davidson precisely because he takes meaning to be evidence constituted.28 Although any particular use may be mistaken, the overall use of the speaker determines the meaning of her words. This means that the function from use to meaning is many-one, and that not every difference in use is a difference in meaning. However, this is perfectly compatible with the thesis that meaning is constituted by use – indeed, anyone defending such a thesis, including Wittgenstein, would have to give a many-one construal of the function from use to meaning for the claim to be at all plausible.29

The question is whether Williams gets Wittgenstein right. As suggested above, there is no reason to think that the conclusion of Wittgenstein’s regress

27 Nor is his claim that learning a first language involves the child applying an interpretation on the expressions taught. See Verheggen 2006 who notes that in the case of a child learning a first language Davidson can call her judgments ‘blind’ just as much as the communitarian can (Verheggen 2006, 215). Davidson also stresses that he is not “concerned with the infinitely difficult problem of how a first language is learned” (Davidson 1986, 100).

28 As Kathrin Glüer argues in her contribution to this volume (Chapter 4), in Davidson’s philosophy evidence has an "epistemico-metaphysical double significance”.

29 Criticizing Davidson’s idea that two individuals rarely speak the same language, Glock suggests that Davidson makes the mistake of assuming that “every divergence in linguistic use amounts to a difference in idiolects” (Glock 2010, 102). Williams makes a similar accusation, suggesting that “any difference in belief or use of an expression marks a different language” (Williams 2000, 305). This, however, is precisely to misunderstand the role of the principle of charity and to miss that Davidson takes the function from use to meaning to be many-one.
argument is that shared conventions are essential to meaning. On the contrary, the argument shows how problematic the idea is that meaning is determined by speakers’ following certain rules. But what role does agreement play then? Does it support the claim that Wittgenstein takes the shared linguistic practice to be essential to meaning?

This issue has been much debated in Wittgenstein scholarship. Scholars have typically proceeded along the lines suggested by Williams, by providing an independent argument in support of the idea that meaning requires a social practice, then suggesting that Wittgenstein must have subscribed to this argument. Looking at the texts, however, there is little evidence in support of this reading of Wittgenstein. Instead, I believe, Wittgenstein stresses the role of agreement for very different reasons: Not because he takes meaning to be determined by the community practice, but because he is investigating the very preconditions of having such a practice. In this, interestingly, his appeal to agreement is quite close to that of Davidson.30

As suggested above, Davidson appeals to the principle of charity in order to solve the problem of disentangling meaning and belief. This problem, again, concerns precisely how use determines meaning. Davidson’s proposal is that by assuming by and large agreement in belief (and by and large rationality), the interpreter can use the speaker’s pattern of assent and dissent to reach an interpretation of the speaker’s words, spelling out the semantic correctness conditions of her expressions. The problem of meaning and belief, it should be clear, is closely connected to the problem of rule-following that occupies Wittgenstein in paragraphs 143–242 of the Investigations: How can the individual’s application of a sign determine that she uses it in accordance with meaning M₁ (‘following rule R₁’) rather than in accordance with meaning M₂ (‘following rule R₂’)? How can it be that use (which is limited) determines correctness conditions (which extend beyond the actual use)? Unlike Davidson, of course, Wittgenstein frames these points in terms of rules and rule-following. The reason for this, Kathrin Glüer and I have suggested, is that Wittgenstein held that there is a certain analogy between rules and meaning: The meaning of a term determines its correct application just as a rule determines a set of actions being in accordance with it, and both meanings and rules are ‘grasped’ by speakers and rule-followers.31 However, that Wittgenstein takes there to be such an analogy should not mislead anyone into thinking that he takes guidance by rules to be essential to meaning, and it should not obscure the close

---

30 This is stressed in Glüer 2000.
31 Glüer & Wikforss 2010. As we note, the analogy between meaning and rules is quite explicit in On Certainty. The meaning of a word, Wittgenstein writes, is a kind of employment of it and that is “why there is an analogy between the concepts ‘meaning’ and ‘rule’” (Wittgenstein 1969, §62).
connection between Wittgenstein’s problem of rule-following and Davidson’s problem of meaning and belief.32

It is in the light of this, I believe, that we should see Wittgenstein’s appeal to agreement, such as his claim in paragraph 242, that if language is to be a means of communication there must be agreement not only in definitions but also in judgments, i.e., in beliefs. Like Davidson, Wittgenstein suggests that communication, agreement in meaning, requires agreement in belief. It is not enough, for instance, that we both agree on an ostensive definition such as ‘this is green’; the possibility of communication (the public nature of meaning) also requires that we agree on the truth value of a number of sentences involving ‘green’, that there is an agreement in empirical judgments. In Remarks of the Foundations of Mathematics, we find the precursor of paragraph 242, where this idea is quite explicit. The phenomenon of language, Wittgenstein writes, is based on regularity, on agreement in action, and it is of the greatest importance that the enormous majority of us agree on certain things:

We say that in order to communicate, people must agree with one another about the meanings of words. But the criterion for this agreement is not just agreement with reference to definitions, e.g. ostensive definitions… but also an agreement in judgments. It is essential for communication that we agree in a large number of judgments (Wittgenstein 1956 VI, 39).33

This suggests that Wittgenstein relies on a principle of meaning determination that is very similar to Davidson’s principle of charity, mapping facts about use on to semantic correctness conditions in such a way that the speaker comes out as (by and large) sharing our beliefs and our way of reasoning. For instance, we can determine whether someone uses ‘green’ with the same meaning as we do on the assumption that she (by and large) agrees with us in her judgments. If she consistently applies ‘green’ to non-green objects it can be concluded (ceteris

---

32 In his later writings, when discussing Kripke’s Wittgenstein Davidson also tends to frame his discussion in terms of rule-following (see for instance Davidson 1990 and 1992). This is not because he has suddenly accepted that speaking a language involves guidance by rules (cf. Davidson 1992, 113) but because he too sees that there is an analogy between the problem of meaning and the problem of rule-following. In these texts Davidson also discusses the notion of ‘going on as before’, much like Wittgenstein, and argues that it requires interaction with a second person – an argument that goes beyond his early account of interpretation. According to some, a similar argument can be found in Wittgenstein’s rule-following discussion and in his remarks on the possibility of a private language (see Verheggen 2006 and 2007). I will not discuss Davidson’s second person argument, or Wittgenstein’s private language argument, since my main concern is with the idea that meaning is social in the sense that conventions determine meaning.

33 At points Wittgenstein’s formulations are very similar to those of Davidson, in particular in On Certainty. For example: “The truth of my statements is the test of my understanding of these statements. That is to say: if I make certain false statements, it becomes uncertain whether I understand them” (Wittgenstein 1969, §80–81).
paribus) that she means something different by ‘green’.34 This agreement is an agreement “on which the working of our language is based” (Wittgenstein 1953, §240), a precondition of the very possibility of a shared language. Naturally, in the case of language instruction (which Wittgenstein often considers) the agreement in belief will lead to an agreement in language (if the instruction is successful), but the two types of agreement must not be conflated. On the contrary, it is against the background of agreement in belief that it can be determined whether there is agreement in language.

One difference between Wittgenstein and Davidson is that whereas Davidson investigates the preconditions of communication by focusing on what goes on in successful communication, Wittgenstein tends to consider cases where communication breaks down – as when someone responds randomly or suddenly goes on in an unexpected way (such as the ‘bent’ rule-follower who is taught the series +2 but starts adding 4 beyond 1000):

Following a rule is analogous to obeying an order. We are trained to do so; we react to an order in a particular way. But what if one person reacts in one way and another in another to the order of training? Which one is right? (Wittgenstein 1953, §206)

By considering cases such as these, Wittgenstein brings out how language and the possibility of communication rest on contingent facts about us as human beings. This leads him to emphasize a further, more primitive type of agreement: an agreement in how we react to a given training. For instance, when given an ostensive definition we agree in what it is to go on in the same way, in what things count as being the same as the one ostended. Paragraph 206 is accordingly followed by a discussion of the notion of sameness, stressing the extent to which it rests on our shared primitive reactions, culminating in his famous remarks in paragraph 219: “This is how it strikes me. When I obey a rule, I do not choose. I obey the rule blindly.” Wittgenstein’s point is not that meaning is determined by rule-guidance and that there is a way of being guided by a rule ‘blindly’, but that language depends on there being a primitive agreement in how we project from a finite number of cases, in things striking us in a certain way.

Williams is therefore right to suggest that Wittgenstein stresses a further type of agreement, one that is simply taken for granted by Davidson.35 Wittgenstein

34 Cf. Investigations §143 where Wittgenstein discusses the difference between a subject making mistakes and making ‘systematic mistakes’: In this case, Wittgenstein says, “we shall be tempted to say that he has understood wrong”, i.e. that he follows a different rule (means something different).

35 However, as noted by Glüer (this volume, Chapter 4), it is plausible that Davidson simply built this type of agreement into the principle of charity, since he speaks of the principle as involving the speaker to be responding to the same features of the world that the interpreter would respond to (Davidson 1991). This is to be expected since there is an intimate connection between this type of agreement and agreement in belief: we could not agree on a range of objects that they are
sometimes describes this deeper-lying type of agreement not as an agreement in belief (‘opinion’) but as an agreement in ‘forms of life’, presumably in order to stress that it is a more primitive type of agreement, functioning as a presupposition of agreement in concepts and belief (Wittgenstein 1953, §241). However, again, that there need be this further type of agreement does not show that there need be agreement in language, shared conventions – on the contrary, the agreement he talks about is a preconditions of having a shared language in the first place. And if this precondition fails, if we cannot “find a great deal of reason and truth” in the subject (Davidson 1974a, 153), then the conclusion is not that we have encountered a language that we do not understand, but that the subject is not a language user. Here, Wittgenstein and Davidson give the same answer: there is not enough detectable regularity for us to call it a language (Wittgenstein 1953, §27). This is part and parcel of the idea, central to both of them, that meaning is essentially public – accessible from our point of view.

4 Concluding Remarks

I have argued that there are important similarities between Davidson’s and Wittgenstein’s approaches to metasemantics: Both take meaning to be determined by use rather than rule-guidance, both take meaning to be essentially tied to communication, and both investigate the conditions under which communication is possible arguing that communication requires agreement in judgments. Strawson is wrong to suggest that Wittgenstein and Davidson are on the opposite sides of his Homeric struggle. He is wrong about this, in part, because he assumes that in order for meaning to be public, meaning need be determined by the shared conventions. But he is also wrong when he claims, like Glock and Williams, that Wittgenstein takes meaning to be essentially tied to conventions.36

Instead of Strawson’s struggle, however, another one can be detected, one that does seem to divide the field of philosophy of language into two opposing camps: the struggle between making meanings perfectly public and making them perfectly objective. As Kathrin Glüer argues in her contribution to this

---

36 It should be noted that Strawson’s text is an inaugural lecture from 1969, and for that reason he is only discussing Davidson’s early paper “Truth and Meaning” from 1967. In this paper the basic ideas concerning meaning and radical interpretation are laid down, but the development of these ideas, in particular Davidson’s view on conventions and communication, comes only in the later papers discussed above. This may explain why Strawson takes Davidson to belong firmly to the formal semantics camp. Nevertheless, leaving Davidson aside, Strawson simply assumes that the project of formal semantics is incompatible with the idea that meaning is essentially tied to communication.
volume (Chapter 4), Davidson and Wittgenstein both present metasemantic accounts that are epistemic in nature, relating to how speakers such as ourselves could come to know the meaning of another’s expressions. The appeal to agreement brings this out since the upshot is that the properties picked out by our terms will be sensitive to our way of responding to stimuli and categorizing objects. This means, Glüer argues, that meanings won’t be perfectly objective, and she suggests we have a dilemma here: the dilemma of objectivity and sensitivity (this volume, Chapter 4).

Here, I think, we have a struggle that quite plausibly could be described in Homeric terms: the struggle between theories that take meanings to be perfectly objective (a contemporary representative would be Kripke) and theories that take meanings to be essentially public and tied to communication. Indeed, this is precisely one of the central fault lines between the early and the later Wittgenstein. On Strawson’s way of setting things up, the difference between early and late Wittgenstein is that the early Wittgenstein belongs to formal semantics theories whereas the later Wittgenstein belongs to communication-intention theories. This is implausible not only because the later Wittgenstein is not a communication-intention theorist but also because it does not seem quite right to describe the Tractatus as a text in formal semantics. More illuminating is to see the change from early to late Wittgenstein in terms of a move from an objective construal of meanings to an epistemic construal. In the Tractatus, meaning is understood thoroughly non-epistemically: not in terms of the discriminatory capacities and sensitivities of speakers but in terms of objectively given properties in the world. Once the sign is projected on to the object its role in language is fully determined. As argued above, it was Wittgenstein’s skepticism about this idea that first led him to appeal to the role of conventions in meaning determination. Eventually, it was his skepticism about the appeal to conventions that led him to argue that language presupposes agreement in judgments and in techniques – landing him on the same side of this particular struggle as that of Davidson.

References


37 I have in mind, here, the Kripke of Naming and Necessity 1980, and his account of natural kind terms. However, it should be clear that what drives the skeptical argument in Kripke’s 1982 book on Wittgenstein is precisely the idea that meanings should be perfectly objective, available sub specie aeternis as it were (see Glüer this volume, Chapter 4).
1 Introduction

Both Ludwig Wittgenstein and Donald Davidson were deeply concerned with language. Most deeply, they cared about foundational, or meaning theoretic questions, questions of the kind: What is meaning? What is it for words to have meaning? And how do we know what words mean? Not only did they care about the same questions, they also shared some very fundamental ideas about their answers. The opening of Quine’s *Word and Object*—“Language is a social art” (Quine 1960, viii)—would in fact be an excellent motto for the meaning theoretic thought of all three of them. Whether investigating rule-following, radical translation, or radical interpretation, these thinkers take a distinctive perspective on the metaphysics of meaning: They recognize the constitutive role of its epistemology. In Davidson’s words:

As Ludwig Wittgenstein, not to mention Dewey, G.H Mead, Quine, and many others have insisted, language is intrinsically social. This does not entail that truth and meaning can be defined in terms of observable behavior, or that it is ‘nothing but’ observable behavior; but it does imply that meaning is entirely determined by observable behavior, even readily observable behavior. That meanings are decipherable is not a matter of luck; public availability is a constitutive aspect of language. (Davidson 2005, 56)

But while it has become comparatively standard, not least due to the immense influence of Kripke’s *Wittgenstein* (1982), to read the rule-following considerations as concerned with meaning determination, it is less common to also see their epistemological strands in this light.¹ It is in precisely this light, however, that I think the deep affinity between Wittgenstein and Davidson on meaning

¹ Wright is an exception to this; cf. Wright 1980; 1989; 2007; 2012. There also is a more recent tendency in the literature on Kripke’s *Wittgenstein* to play up epistemological demands on a “straight solution”, but these do not concern the publicness of meaning. Rather, Kripkenstein picks up on another epistemological aspect of the rule-following considerations: the idea that meanings must be such that they justify the speaker’s use of linguistic expressions (cf. e.g. PI 217). Cf. a. o. Jackman 2003; Kusch 2006; Guardo 2010; Merino-Rajme 2015.
determination is best appreciated. Moreover, while it has become fairly standard to structure the debate in terms of the contrast between reductivism and non-reductivism, this very landscape will come to look somewhat different from the perspective suggested here.

In this chapter, I shall assume that the rule-following considerations are — in a sense to be made more specific, however — about meaning determination. My question will be whether the principle of meaning determination used in the early Davidson’s account of meaning determination — the principle of charity — provides an answer to what I will call “Wittgenstein’s paradox.” I shall proceed as follows. In Section 2, I shall work out the connection between the rule-following considerations and meaning determination and isolate the aspect of Wittgenstein’s paradox I want to explore: the “problem of objectivity.” Then, I shall run us through the basics of the radical interpretation account of meaning determination in Section 3. In Section 4, I shall argue that the principle of charity does seem to fall prey to the problem of objectivity, a verdict I shall ultimately endorse after unsuccessfully trying to rescue objectivity by means of Lewisian natural properties in Section 5.

2 Rule-Following and Meaning Determination

The rule-following considerations (roughly, Philosophical Investigations 138–242 and Remarks on the Foundations of Mathematics, section VI) culminate in what I shall call “Wittgenstein’s paradox”:

This was our paradox: no course of action could be determined by a rule, because every course of action can be made out to accord with the rule. (PI 201)

Two very basic questions concerning this well-known passage are the following: First, how is this about meaning? And second, how did Wittgenstein get into this “paradox”?³

As Åsa Wikforss and I have argued at length elsewhere (Glüer and Wikforss 2010), it is wrong to think that the rule-following considerations are about meaning because Wittgenstein conceives of speaking a language as an essentially rule-guided activity. While this might still be the received view, it is in fact an idea very much at odds with central tenets of the later Wittgenstein — such as the idea that meaning is, or is determined by, use. The rule-following considerations thus do not present a deep and subtle defense of the idea that

² On this, I take Verheggen 2003 and myself to be in agreement — even though we disagree on how, precisely, the affinity is best developed (cf. Verheggen 2006; 2007; 2013, Glüer 2006b, Glüer and Pagin 2003).

³ There is also the question in what sense, if any, the problem Wittgenstein formulates here is a paradox. I shall not pursue this question; rather, I shall simply use “Wittgenstein’s paradox” as a name for the problem.
language is a rule-guided activity against Wittgenstein’s own paradox. Rather, they provide some deep and subtle reasons for rejecting this idea.4

By the time of the Investigations, Wittgenstein is no longer thinking of language as a calculus of grammatical rules. Instead, he is investigating meaning by means of the game analogy, the analogy between speaking a language and playing a game, between meaning and rules. And the game analogy is precisely that: an analogy, and a rich and rewarding one at that. Its most important aspect in this context is this: the meaning of an expression determines its correct application in much the same way that a rule determines a set of actions as being in accordance with it. Also, both meanings and rules are things speakers and rule-followers are said to “grasp”. It is probably only in On Certainty that Wittgenstein fully and explicitly brings these fundamental elements of his late thought together, but there, he clearly sums up the intimate relation between the game analogy and the idea that meaning is, or is determined by, use:

A meaning of a word is a kind of employment of it. For it is what we learn when the word is incorporated into our language. That is why there is an analogy between the concepts ‘meaning’ and ‘rule’. (OC 662)

But as Verheggen neatly brings out when formulating what she calls the “determination problem”, trouble is brewing once we put these things together:

Wittgenstein has reason to wonder how a sign that is meaningful can determine its applications, for he has been developing the view that all there is to the meaning of a word is its use in a language. (Verheggen 2003, 289)

So, the rough answer to the two questions above is this: Wittgenstein’s paradox is about meaning because in the rule-following considerations he is investigating meaning by exploring the analogy between meaning and rules. And he gets himself into the paradox by thinking of the meaning of an expression as “a kind of employment of it”, as use.

It will be useful to distinguish between two aspects of the determination problem, however. Both of these concern what we can call the “determination target”, i.e. the second relatum of the determination relation in question. According to the game analogy, the meaning of an expression is like a rule in determining the correct applications of that expression. This determination target has several worrisome aspects. First of all, it consists of a potential infinity of correct applications. The question then is: How can meaning possibly determine a potential infinity of applications in advance of these applications actually being made? This question is quite intriguing in itself, but it becomes rather perplexing once combined with the idea that the meaning of an

4 We are, of course, by no means claiming that this is all they do.
expression is, or is determined by, its use, i.e. its actual applications. How could
its use, which is finite, determine an infinity of correct applications for an
expression?\(^5\) We can call this aspect of the determination problem the “infinity
in advance problem.”

Second, what we can call the “objectivity problem” concerns a different
property of the target applications of an expression: their correctness. In many
cases, these applications need to be such that their correctness is an objective
matter if there is to be any correctness at all. The question then is: How can
meaning possibly be such that the correctness of the applications of meaningful
expressions is an objective matter? And again, the question becomes rather
perplexing once combined with the idea that the meaning of an expression is, or
is determined by, its use. Wouldn’t whatever we do have to count as correct,
thus undermining the very idea of correctness?\(^6\)

Distinguishing these problems provides us with a somewhat more precise
idea of how the idea that meaning is, or is determined by, use gets Wittgenstein
into his paradox. If meaning is use, it is very hard to see how meaning can
possibly determine what it is supposed to determine: a potential infinity of
objectively correct applications.

It has become customary, I said above, to read the rule-following considera-
tions as about meaning determination. In the light of the problems just spelled
out, this might appear questionable or even false. But it isn’t. The rule-
following considerations are about meaning determination – just not in
a totally direct way. To see why, let me first explain why the claim might appear
to be false. Questions about meaning determination concern what determines
meaning, not what gets determined by meaning. In meaning determination, we
might say, meaning is the determination target. In Verheggen’s “determination
problem”, by contrast, meaning is the not the target, but the determination base.
Here, meaning does the determining.\(^7\)

Keeping these determination relations distinct, we get a picture like the
following:

\[(\text{DMD}) \quad \text{Use} \rightarrow_{D_1} \text{Meaning} \rightarrow_{D_2} \text{Application}\]

Ultimately these relations are supposed to compose, of course: Whatever
determines meaning thereby determines correct application. In what follows
I shall therefore both speak of meaning and of use determining correct applica-
tion. Intuitively, both \(D_1\) and \(D_2\) appear to be relations of the same kind or

\(^5\) Cf. Verheggen 2003, 289, Glüer and Wikforss 2010, 156. See also Pagin 2002.
\(^6\) Cf. Verheggen 2003, 289f. I take it to be obvious that both the infinity in advance and the
objectivity problem are amongst those most exercising Wittgenstein in the rule-following
considerations.
\(^7\) Essentially the same distinction is drawn in Pagin 2002, 156f.
“strength”; both are metaphorical determination relations and their holding appears to be a matter of at least metaphysical necessity.\(^8\) Wittgenstein’s paradox is that this seems to be an illusion: Nothing metaphysically determines correct applications for an expression.

Wittgenstein himself, identifying meaning with, or at least very closely tying it to, use, does not seem to blame the problem on \(D_1\). The culprit, in his opinion, is \(D_2\). Thus, his problem is how there could be meaning at all even though “no course of action could be determined by [it]” (PI 201). In Kripkenstein’s version of the paradox, on the other hand, the problem does not seem to be that meaning does not determine correct application, but that there is nothing that determines meaning. According to Kripke’s skeptic, what we need to give up is \(D_1\).\(^9\)

This, then, is the sceptical paradox. When I respond in one way rather than another to such a problem as ‘68 + 57’, I can have no justification for one response rather than another. Since the sceptic who supposes that I meant quus cannot be answered, there is no fact about me that distinguishes between my meaning plus and my meaning quus. Indeed, there is no fact about me that distinguishes between my meaning a definite function by ‘plus’ (which determines my responses in new cases) and my meaning nothing at all. (Kripke 1982, 21)

But as I said, insisting on this distinction by no means amounts to saying that the rule-following considerations are not about meaning determination. They are, just in less direct and more complicated, more subtle ways. For one thing, Wittgenstein gets into his paradox because of the idea that meaning is, or is determined by, use. It is meaning-as-(determined-by)-use that threatens to fall short of delivering infinity in advance and objectivity. If this result can be made to stick, it tells us something quite stunning about meaning determination: If meaning is determined by use, infinity in advance and objectivity go by the board.

But this is not all. At this point, we might well feel like asking: Why not construe this result as a reductio ad absurdum of the idea that meaning is use? Why not go looking for a better idea about meaning determination? The answer suggested by the rule-following considerations seems to be: Because it wouldn’t help. Wittgenstein’s paradox would still be with us. This even more dramatic conclusion can be seen as motivated in terms of the development of Wittgenstein’s thought through the middle and later periods.\(^10\) But the upshot of this development also finds concentrated

\(^8\) Wittgenstein himself arguably thought of these relations as non-contingent and “internal.” The obtaining of an internal relation probably amounts to something stronger than metaphysical necessity.

\(^9\) I am following Pagin 2002, 157, fn. 9 here.

\(^10\) In Glüer and Wikforss 2010, we provide a more detailed reconstruction of this development as it concerns precisely rule-following and meaning.
expression in the rule-following considerations itself—thus the intense dis-
cussions of explanations of rules, of mental states (or processes) of under-
standing or mental states (or processes) of grasping rules, as well as of
calculating machines implementing rules.

Now, if we understand these discussions as proceeding by the method of exclusion, it seems reasonable to doubt that the list of potential “mean-
ing determiners” is exhaustive. But there is more to the method Wittgenstein employs: There is system. A recurring complaint—most
prominently concerning explanations and mental states—is that a vicious
regress results. If, for instance, we explain the meaning of a word in other
words, the explanation will consist of just more linguistic items—items
just as much in need of explanation as the one we started with. The basic
idea is that whatever item we provide as a candidate for meaning deter-
miner, this item will turn out to be itself in need of getting its meaning
determined, or of getting “interpreted.” Whatever the candidate, it will
just bring us back to Square One, back to the very same kind of problem
we started with.

To see the underlying mechanics, we do not need to consider all the partic-
cular candidates for meaning determiner. We can instead look at a general-
ization of our determination scheme (DMD):

\[
\text{(DMD)} \quad \text{Base} \rightarrow_{D_1} \text{Meaning} \rightarrow_{D_2} \text{Application},
\]

where ‘Base’ now is a variable ranging over (sets of) candidates for meaning
 determiner. The first observation then concerns determination relations quite in
general. It is this: Whatever you put into a determination base, it will not, just
by itself, determine your target. Basically, determination relations are (one-one
or many-one) functions from one domain or set of entities \(S_1\) to another domain
or set of entities \(S_2\). And if all I tell you is that something is a function of
a certain kind, or set, of items \(S_1\), there is a sense in which that doesn’t tell you
much at all: For any \(S_1, S_2\), there are sufficiently many such functions to make it
the case that for any pair of items \(s_1\) from \(S_1\) and \(s_2\) from \(S_2\), there is a function
mapping \(s_1\) onto \(s_2\). Translated into Wittgensteinian terms, this amounts to the
following claim: Whatever kind of item we put into Base, “every course of
action can be made out to accord with [it],” that is, every possible application
can be “made out” to be correct. Providing a determination base therefore is
only part of the answer to any determination question. We also need to know the

\[11\] Relations of metaphysical determination might, of course, have properties over and above being
such functions, properties such as having a direction (from the metaphysically more funda-
mental to the metaphysically less fundamental) or providing a special kind of explanation, but
these properties do not matter at this point.
relevant function or principle of correlation: We need to know by what principle the items in the determination target are determined by those in the determination base.

And the crux of the rule-following considerations is that this need for a principle itself poses a problem, a problem of the very same kind as the one we started with. As soon as there is more than one candidate for being the relevant principle of correlation, the worry is, we are back to Square One. For then, it needs to be determined which principle is the right one. As Pagin puts it with respect to states of understanding: “if meaning is to be determined by a state of understanding, then the state of understanding must also select the principle of correlation, and this brings us back to the first problem again” (Pagin 2002,160). As long as we do not provide justification for picking a particular such principle, our account of meaning determination is in trouble.

It is at this point that turning to Davidson seems so promising. For as we already began to see in the introduction, Davidson’s account of meaning determination aims at providing precisely the ingredients we need now: A determination base, a determination principle, and, crucially, justification for both the choice of base and the choice of principle. Moreover, there is that deep affinity between, on the one hand, these Davidsonian choices and their justifications and, on the other hand, the answers at least suggested by the later parts of the *Investigations*, answers turning on the social character of language and crucially involving agreement (in primitive reactions and judgments)

---

12 As Pagin points out, this holds equally well for dispositions. In the context of the rule-following considerations, the real problem with meaning determination by dispositions is neither their (supposed) finiteness nor that we are disposed to make mistakes. The problem is that, like any other items we might put into the determination base, dispositions by themselves do not “pick” a principle of correlation. Unless such a principle is specified, we can map any meaning/correct applications whatsoever onto any expression from any given set of dispositions regarding that expression. In Kripke’s discussion of dispositions, a rather simple such principle is in fact implicitly assumed: According to this principle, the uses a speaker is disposed to make of an expression simply are the correct uses. But why this principle rather than any other principle linking dispositions to correct applications? This question cannot be answered by merely complicating the principle of correlation. What is needed is justification for picking a particular principle instead of any of the others. Cf. Pagin 2002, 161.

13 This holds largely independently of the question of whether we think of an account of meaning determination as in the business of reduction or not. As long as what we put into the determination base does not, somehow, already select the principle of correlation, the problem recurs. Arguably, it recurs for both reductive and non-reductive, but informative accounts. The only kind of “account” that might remain unaffected is a sort of quietism about meaning determination insisting that meaning facts determine meaning facts. Note, however, that this is not because determination base and target are identical. There are many ways of mapping a domain onto itself. Rather, the thought would have to be that for a fact to be a meaning determining fact it needs to be such that it can obtain while the meaning fact it determines obtains. This would not be the case for a mapping from meaning facts to meaning facts which does not map every meaning fact onto itself.
as well as a common way of understanding, and explaining, action. In the next section, we shall therefore turn to Davidsonian meaning determination, more precisely, to the account of meaning determination that Davidson offers in a series of papers concerned with radical interpretation.

3 Radical Interpretation

In this section, I will run us through the basics of Davidsonian radical interpretation as an account of meaning determination. My focus will thus be on the “early” Davidson and a series of papers we can call “the radical interpretation papers.” My aim here is just to provide enough background for understanding the Davidsonian answers to the following questions: What is in the determination base for meaning? What is the principle of meaning determination? Why are these the base and principle of meaning determination— or: what justifies the Davidsonian choices of base and principle?

To fully appreciate the Davidsonian choices and their justification, we need to be clear about the basic perspective he takes on meaning. According to Davidson, meaning is a theoretical concept. Its main purpose is the explanation of successful communication by language. The same holds for concepts like those of reference, predicate, or sentence; their “main point... is to enable us to give a coherent description of the behavior of speakers, and of what speakers and their interpreters know that allows them to communicate” (Davidson 1992, 108f).

What speakers and their interpreters know when they successfully communicate by language are things like the following: what someone said, what the

---

14 In PI 206, Wittgenstein asks us to imagine coming “as an explorer into an unknown country with a language quite strange to [us].” And he refers to the idea of the “common behavior of mankind” (“die gemeinsame menschliche Handlungsweise”) as a “system of reference by means of which we interpret an unknown language.” Cf. also PI 243. One commentator who reads these passages in very Davidsonian terms, indeed, is Hopkins (cf. Hopkins 1999; 2012). According to Hopkins, Wittgenstein solves his paradox by construing meaning as constituted by ideal radical interpretation, i.e. by adopting a form of judgment-dependence about meaning. There are commentators ascribing this kind of “interpretivism” to Davidson, too (cf. for instance Byrne 1998 and, possibly, Williamson 2004, 137). As Wittgenstein is concerned with meaning and content determination, such an interpretivism would be rather obviously question-begging, however (cf. Gross 2015). And the same goes for Davidson. In both cases, exegetical charity rules these interpretations out. See Boghossian 1989, 546f for discussion of the more general idea (proposed in Wright 1989) of construing meaning or content as response-dependent.

15 It is, of course, by no means obvious what Wittgenstein’s own solution, if any, to his paradox is supposed to precisely consist in. I shall not enter into these discussions here. For a comparison between Davidson and Wittgenstein on the role of agreement, see Glüer 2000, however.

16 The most important of the radical interpretation papers are Davidson 1973; 1974; 1975; 1976. A useful overview can be found in Davidson 2005. By contrast, Verheggen (1995; 2000; 2006) draws on affinities between the later Wittgenstein and the “later” Davidson on meaning and triangulation. For more on triangulation, see also Pagin 2001; Glüer 2006b; Verheggen 2007; 2013.
uttered expressions mean, and how to express a certain thought in language. This knowledge is the output or result of what we can call our linguistic ability or competence. Concentrating on the interpretive side of this ability, Davidson suggests approaching the fundamental meaning theoretical question – the question “What is it for words to mean what they do?” – indirectly: by means of two others. Classically, these are formulated in the course of the opening paragraph of “Radical Interpretation”:

Kurt utters the words ‘Es regnet’ and under the right conditions we know that he has said that it is raining. Having identified his utterance as intentional and linguistic, we are able to go on to interpret his words: we can say what his words, on that occasion, meant. What could we know that would enable us to do this? How could we come to know it? (Davidson 1973, 125, emphasis added)

And some years later, in the introduction to the collection Inquiries into Truth and Interpretation, Davidson recapitulates his project:

What is it for words to mean what they do? … I explore the idea that we would have an answer to this question if we knew how to construct a theory satisfying two demands: it would provide an interpretation of all utterances, actual and potential, of a speaker or group of speakers; and it would be verifiable without knowledge of the detailed propositional attitudes of the speaker. (Davidson 1984, xiii)

As is well known, Davidson proposes that the theory we are after is a formal semantic theory for a natural language $L$. According to him, such a theory is compositional, and takes the form of a Tarskian truth-theory (t-theory). A t-theory for a language $L$ is supposed to give the meaning of each sentence of $L$ by specifying its truth-conditions. And according to Davidson, the meaning of an expression of $L$ is precisely its systematic contribution to the truth-conditions of the sentences it occurs in, a contribution spelled out by the correct t-theory for $L$. By using a t-theory as a formal semantic theory for a natural language $L$, Davidson submits, we can describe or model the linguistic competence that allows for interpreting utterances in $L$. The resulting knowledge is empirical knowledge. And a formal semantic theory for a natural language $L$ is an empirical theory – it is an empirical

---

17 The reader unfamiliar with this framework might for instance consult my (2011), where a t-theory for a fragment of English is provided in the Appendix.
18 There is no need, Davidson submits, to assign entities – such as propositions – as meanings to expressions: “My objection to meanings in the theory of meaning is not that they are abstract or that their identity conditions are obscure, but that they have no demonstrated use” (Davidson 1967, 21).
19 It is important to distinguish between two potential objects of knowledge here: According to Davidson, what speakers know is the output or result of their linguistic competence, i.e. what utterances mean. This does not mean that they know the theory by means of which we model that competence (cf. Davidson 1986, 96). Thus, the hypothetical form the question takes in “Radical Interpretation”: “What could we know that would enable us to do this?”
question whether any particular such theory is correct for \( L \), i.e., whether it gives the right meanings for utterances in that language. Such knowledge is based on evidence, justified by empirical data. What are the data supporting formal semantic theories for natural languages?

According to Davidson, there are two major restrictions on the foundationally interesting data for semantic theorizing. First, to learn something about what meaning is, we must be able to formulate these data in terms not presupposing meaning or any other semantic notions – “without essential use of such linguistic concepts as meaning, interpretation, synonymy, and the like” (Davidson 1973, 128; see also Davidson 1974, 142f). To learn something about what meaning is, we thus are after the “ultimate evidence” (Davidson 1973, 128) for any correct theory of interpretation. And second, since we are talking about knowledge that any competent speaker has, the data must be “plausibly available to a potential interpreter” (Davidson 1973, 128), where the interpreter is just another competent speaker. What kind of data is there that could do this job? This is the question that leads us directly to the idea of radical interpretation.

In radical interpretation, the interpreter is to construe a \( t \)-theory for a radically foreign language, a language she doesn’t know anything about at the start. The only evidence available to her consists of data about the behaviour of the speakers and the observable circumstances in which it occurs. Such data, Davidson holds, allow for the identification of a certain kind of attitude the speakers hold to uninterpreted sentences: the attitude of holding an uninterpreted sentence \( s \) true (at a time \( t \)). This is an intentional attitude, in fact, a belief, but it is an attitude of the kind that Davidson calls “nonindividuative” (Davidson 1991, 211): the interpreter can know that a speaker holds this attitude towards \( s \) at \( t \) without knowing what \( s \) means and, thus, without knowing which belief the speaker thereby has. Thus, no meaning (or content) theoretic questions are begged when using data about this attitude in the account of meaning determination.

The only thing special about the radical interpreter then is that, in contrast to an ordinary competent speaker, she has huge amounts of such data at her disposal: “We may as well suppose”, Davidson writes, “we have available all that could be known of such attitudes, past, present, and future” (Davidson 1974, 144).\(^{20}\)

The radical interpreter thus collects vast amounts of data like the following:

\(^{20}\) What this precisely means is not so easy to understand, however. On the one hand, Davidson is concerned with making sure that the available evidence can support sufficiently many of the differences in meaning it is pre-theoretically plausible to think we can detect. But on the other hand, we need to be careful not to stretch the limits of the evidence available to the radical interpreter beyond recognition. After all, her evidence is supposed to be accessible to an everyday interpreter.
Kurt belongs to the German speech community and Kurt holds true ‘Es regnet’ on Saturday at noon and it is raining near Kurt on Saturday at noon.

Sufficient numbers of observations like (E) then are supposed to support t-theories from which theorems like the following t-sentence can be derived:

\[(T) \text{ ‘Es regnet’ is true-in-German when spoken by } x \text{ at time } t \text{ iff it is raining near } x \text{ at } t.\]

Subscribing to Quinean confirmation holism, Davidson construes the relation between theory and evidence as a holistic one. It is whole t-theories that are supported by the data to varying degrees. “[T]he method,” Davidson explains, “is … one of getting a best fit” (Davidson 1973, 137).

But what is it for data like (E) to “fit” t-theories entailing (T)? The basic idea is to assign the conditions under which speakers hold sentences true as the truth conditions of those sentences. But speakers hold all sorts of things true under all sorts of circumstances. More precisely, whether a speaker holds a sentence true under given circumstances depends crucially on the (further) beliefs of the speaker. Take a speaker who (erroneously) believes that there is an elaborate system of sprinklers on the roof. Upon looking out of the window on a rainy Saturday at noon, such a speaker might not only fail to believe that it is raining, but even form the belief that someone must have turned those sprinklers on. This is just one example of the pervasive phenomenon Davidson calls the “interdependence of belief and meaning.” The interdependence is evident in this way: a speaker holds a sentence to be true because of what the sentence (in his language) means, and because of what he believes. Knowing that he holds the sentence to be true, and knowing the meaning, we can infer his belief; given enough information about his beliefs, we could perhaps infer the meaning. (Davidson 1973, 134f)

Just by themselves, observations like (E) thus do not provide any evidence whatsoever for a t-theory. As long as the interpreter can ascribe any old belief, be it ever so weird or absurd, all such observations can be squared with any old t-theory. It is here that the principle of charity is supposed to kick in. In one of its earliest formulations, Davidson formulates the principle as follows:

\[(PC) \text{ Assign truth-conditions to alien sentences that make native speakers right when plausibly possible. (Davidson 1973, 137)}\]

It provides a method for solving the problem of the interdependence of belief and meaning.

\[21\] For more detailed accounts of the method of radical interpretation, see Lepore and Ludwig 2005 and Glüer 2011.
by holding belief constant as far as possible while solving for meaning. This is accomplished by assigning truth conditions to alien sentences that make native speakers right when plausibly possible, according, of course, to our own view of what is right. (Davidson 1973, 137)

The radical interpreter, that is, tries to hold belief constant both between himself and the alien speaker, but also for the alien speaker over time. Now, we can see why Davidson proposes to “take the fact that speakers of a language hold a sentence to be true (under observed circumstances) as prima facie evidence that the sentence is true under those circumstances” (Davidson 1974, 152). If beliefs, and belief ascriptions, are restricted by the principle of charity, as Davidson argues they are, then a sentence’s being held true under certain circumstances does provide evidence that it is true under those circumstances.

The evidence is prima facie, however. That is, it can be overridden by other, stronger evidence. People do make mistakes, and some of the vast number of data that the radical interpreter collects will have to be considered as overridden by others. But which? Davidson: “The basic methodological precept is … that a good theory of interpretation maximizes agreement. Or, … a better word might be optimize” (Davidson 1975, 169).22

It is not only truth that is important here, however. Mistakes can also come in the form of incoherence, in the form of drawing the wrong inferences from what one believes. Beliefs, Davidson maintains, come in coherent clusters, if they come at all (cf. Davidson 1977, 200). Take Fido, the dog. While we might very well be in a situation where it is plausible to think that someone believes that Fido is a car, this very hypothesis is instantly made much less plausible if we also think that they at the same time fail to draw obvious inferences – such as that Fido is an artifact, not an animal – from it. This pressure towards coherence extends all the way to a subject’s actions – the hypothesis that you believe that Fido is a car comes under pressure, too, if we observe you carrying a leash while muttering “Fido needs to be taken out a bit.”23

22 Optimizing agreement involves weighting mistakes or disagreements and minimizing a theory’s overall score. This is better than simply counting the number of mistakes a theory ascribes (which might not even be possible given that the number of sentences in a language is infinite) because “some disagreements are more destructive of understanding than others” (Davidson 1975, 169). Usually, being wrong on simple observational matters such as whether it rains around one is more destructive than disagreement on highly theoretical matters. Being wrong about one’s own mental states or about how things look to one is worse than being wrong about other’s mental states or about how things are. The general idea is that a mistake is the more weighty the more epistemologically basic it is, the more basic, that is, to the totality of our knowledge: “The methodology of interpretation is, in this respect, nothing but epistemology seen in the mirror of meaning” (Davidson 1975, 169).

23 Here we have, then, those elements of meaning determination crucial also to the later, post-rule-following Wittgenstein: agreement in judgement (and primitive reactions) and a common way
The maxim to make the speaker right when plausibly possible encompasses both these elements: To make the speaker right when plausibly possible is to optimize the beliefs ascribed in such a way that they are, at least in basic cases, mostly true and coherent. Charitably interpreted speakers therefore always come out as persons of a certain, basic rationality. In later writings, Davidson sometimes explicitly separates the two components of charity – truth and coherence:

The process of separating meaning and opinion invokes two key principles which must be applicable if a speaker is interpretable: the Principle of Coherence and the Principle of Correspondence. The Principle of Coherence prompts the interpreter to discover a degree of logical consistency in the thought of the speaker; the Principle of Correspondence prompts the interpreter to take the speaker to be responding to the same features of the world that he (the interpreter) would be responding to under similar circumstances. Both principles can be (and have been) called principles of charity: one principle endows the speaker with a modicum of logic, the other endows him with a degree of what the interpreter takes to be true belief about the world. Successful interpretation necessarily invests the person interpreted with basic rationality. It follows from the nature of correct interpretation that an interpersonal standard of consistency and correspondence to the facts applies to both the speaker and the speaker’s interpreter, to their utterances and to their beliefs. (Davidson 1991, 211)

We shall have occasion to get back to this formulation of charity later. Here and now, we can summarize charity’s role in radical interpretation. Given the methodology of best fit and the interdependence of belief and meaning, the principle of charity is supposed to fulfill two essential functions: It

1. restricts belief ascription so that observations like (E) provide data for t-theories, and
2. ranks t-theories by fit to the totality of available data such that the best is/are correct.24

These, then, are in rough outline, the answers to the two questions posed at the beginning of “Radical Interpretation”: According to Davidson, Tarskian t-theories can be used as formal semantic theories for natural languages. And the data supporting them ultimately are data about the behaviour of speakers and its observable circumstances. To round off the picture, we need to connect two more dots: We need to connect the epistemology of meaning with its metaphysics. Why do we learn something about what meaning is from learning how to justify semantic knowledge? Why would the epistemology of meaning tell us something about its metaphysics?

of making action intelligible. Davidson in fact came to think that we ultimately needed “A Unified Theory of Thought, Meaning and Action” (Davidson 1980).

24 According to Davidson, it is pretty much inevitable that there will be more than one “best” theory.
Davidson’s Quine-inspired answer to this question is encapsulated in the following claim: “What a fully informed interpreter could learn about meaning is all there is to learn” (Davidson 1973, 148). There are no meaning facts beyond those that can be known on the basis of evidence available to the interpreter.

Quine revolutionized our understanding of verbal communication by taking seriously the fact, obvious enough in itself, that there can be no more to meaning than an adequately equipped person can learn and observe; the interpreter’s point of view is therefore the revealing one to bring to the subject. (Davidson 1990b, 62)

But intuitively, it is quite unusual for evidence to have such epistemico-metaphysical double significance; with respect to most objects or properties we do not think that the facts about them are exhausted by the evidence available to us. Why would meaning be different? Because meaning is essentially public:

What we should demand … is that the evidence for the theory be in principle publicly accessible…. The requirement that the evidence be publicly accessible is not due to an atavistic yearning for behavioristic or verificationist foundations, but to the fact that what is to be explained is a social phenomenon. … As Ludwig Wittgenstein, not to mention Dewey, G.H. Mead, Quine, and many others have insisted, language is intrinsically social. This does not entail that truth and meaning can be defined in terms of observable behavior, or that it is ‘nothing but’ observable behavior; but it does imply that meaning is entirely determined by observable behavior, even readily observable behavior. That meanings are decipherable is not a matter of luck; public availability is a constitutive aspect of language. (Davidson 2005, 55f, emphasis added)

Language is essentially social: Meanings are such that they can be understood. This, for Davidson, is the most fundamental thing about language. And, as I said right at the beginning of this section, in Davidson’s hands, meaning is nothing more than a theoretical notion used to explain linguistic communication. So, he argues, there cannot be more to meaning than what we can know about it. The data for the t-theory that the radical interpreter is after therefore have an epistemico-metaphysical double nature. They epistemically support the theory, but at the same time, Davidson tells us, they “entirely determine” the very thing the theory is a theory of. The data available to the radical interpreter thus form the determination base for meaning. Meaning is an evidence-constituted property.

By taking this perspective Davidson is able to suggest justifications for both his choice of base and his choice of principle for meaning determination. Strictly speaking, Davidsonian meaning determination is two step: First, observable behaviour in observable circumstances determines attitudes of holding true towards uninterpreted sentences. And in the second step, attitudes of holding true towards uninterpreted sentences determine, via the principle of
charity, meanings for those sentences (and, simultaneously, contents for beliefs (and other propositional attitudes)). Davidson focuses almost exclusively on the second step, and so shall we. The base we are concerned with thus contains attitudes of holding uninterpreted sentences true, and meanings are determined for those sentences by means of the principle of charity. According to Davidson, these choices are justified by what he argues is a “constitutive aspect of language”: the public nature of meaning.

That meanings are knowable is indeed a constitutive aspect of language, it seems to me. And that the semantic notions are theoretical notions whose purpose is the understanding or explanation of successful linguistic communication also seems extremely plausible. Moreover, these two insights go hand in hand, and once fully adopted, they do seem to unlock resources unavailable from the austere, “platonist” perspective of the rule-following considerations, resources for constraining both what is a reasonable candidate for meaning determiner and what is a reasonable candidate for meaning determining principle. It is thus rather natural to wonder whether adopting this Davidsonian point of view might not make answers available to the very questions the rule-following considerations are tempting us to despair of – answers congenial to Wittgenstein.

4 Casey, Alien, and the Problem of Objectivity

We have seen why it is tempting to think that charity might be an answer to the rule-following considerations. A thorough investigation of this idea would require looking into more aspects of charity than I have space for here. One question concerns the epistemic and modal status of charity. The rule-following considerations seem to be premised on the relevant determination relations being very strong modal relations. And while some commentators argue for Davidsonian charity being a conceptual necessity (and knowable a priori – cf. a.o. Lepore and Ludwig 2005), I find that both implausible and exegetically strained. It is better, or so Peter Pagin and I have argued elsewhere, to construe Davidsonian charity as an a posteriori necessity, most plausibly a nomological one.25 But in that case, how much of an answer to the rule-following considerations can we hope to get from charity?

Another question concerns the justification of charity. How, precisely, does the publicness of meaning justify the choice of charity as the meaning determining principle? A common assumption amongst commentators appears to be that this is relatively straightforward: Charity is justified because it makes meaning knowable (cf. a.o. Lepore and Ludwig 2005). But to avoid this being an instance of affirming the consequent, we would also need to argue

that no other principle would make meaning knowable. In radical interpreta-
tion, this means arguing that no other principle would make holding uninter-
preted sentences true into data for t-theories. But it seems fairly obvious to me
that any principle that assigns truth-conditions as a function of holding true
attitudes would do precisely that.\textsuperscript{26} We might, of course, feel that alternative
such principles would not result in what we intuitively think are the correct
meaning assignments for our actual languages, but if that feeling is justified, it is
not justified by such principles’ making meaning unknowable. The justification
for charity therefore must at least in part originate elsewhere. As I have argued
(Glüer 2006a; Glüer 2011, esp. Chapter 3), Davidson himself locates it in the
very nature of belief – belief, he argues in numerous places, “is in its nature
veridical” (Davidson 1982,146).\textsuperscript{27} If there are any beliefs at all, that is, they

\textsuperscript{26} Ludwig disagrees:

“Glüer objects that charity isn’t needed to succeed at interpretation. Any principle will do. Is that
right though? What about the principle that says that the interpreter should interpret the speaker
as massively wrong about his or her environment? How does that enable the interpreter to use
her evidence to gain access to a detailed picture of the speaker’s meanings and attitudes? It is no
guide whatsoever. Or suppose the principle is to assign beliefs about prime numbers on the basis
of the number of words in sentences held true. Does that relate the interpreter’s evidence to how
she interprets the speaker? Charity gives us a principle that shows us how the evidence
marshaled can be brought to bear systematically upon our interpretation of the other as
a speaker and an agent. It is not clear that there is another principle that can do the same job.”
(Ludwig 2014, 469)

The second principle Ludwig considers is quite irrelevant, as it is not a principle “determining
truth conditions on the basis of holding true attitudes” (as I said was required; Glüer 2011, 143).
But the first could be used just as much as charity, it seems to me. Charity makes holding true
into evidence because it allows us to “take the fact that speakers of a language hold a sentence to
be true (under observed circumstances) as prima facie evidence that the sentence is true under
those circumstances” (Davidson 1974,152). But if what we could call “anti-charity” holds, i.e. if
speakers are (to be construed as) massively wrong about their environment, we can take the fact
that they hold a sentence true under observed circumstances as prima facie evidence that the
sentence is \textit{false} under those circumstances. I see no reason to think that this method could not
be used to construct a “detailed picture of the speaker’s meanings and attitudes.” The resulting
picture might not be correct for actual speakers, but that is irrelevant here. The question is
whether meanings would be knowable on the basis of data about holding true, if meanings were
determined (on the basis of holding true) by a principle other than charity. Again, it seems to me
fairly obvious that the answer is yes for \textit{any} principle assigning meanings to sentences on the
basis (or as a function) of holding uninterpreted sentences true (under observable
circumstances).

\textsuperscript{27} Ludwig agrees, but thinks that this claim is too controversial for charity to rest upon. It needs to
be argued for itself, and according to him, the argument Davidson provides for belief’s veridical
nature is indirect. It goes precisely via radical interpretation and charity’s being the only
available principle allowing us to use our evidence (cf. Ludwig 2014, 468f). I quite agree that
belief’s veridical nature requires argument and that Davidson provides such arguments. Here’s
one that’s quite typical:

“A belief is identified by its location in a pattern of beliefs; it is this pattern that determines the
subject matter of the belief, what the belief is about. Before some object in, or aspect of, the
world can become part of the subject matter of a belief (true or false) there must be endless true
beliefs about the subject matter.” (Davidson 1975, 168)
come in largely true and coherent clusters. For Davidson, it is because of this essential fact about beliefs that the radical interpreter can use charity to break into the interdependence of belief and meaning. The choice of charity as the meaning determining principle is justified, according to him, because meaning is essentially public and belief is essentially veridical. But isn’t this latter claim way too philosophically ambitious, or way too contentious, for basing an answer to the rule-following considerations on?

While intriguing, these are not the questions I am going to focus on here. Rather, I’d like to go back to Wittgenstein’s paradox. More precisely, I want to go back to that aspect of the paradox that I called the “problem of objectivity” above. Meaningful expressions have conditions of correct application. And, in many cases at least, the correctness of an application has to be an objective matter if there is to be any correctness at all. The correctness of the use of our expressions must have a certain sort of independence from us – it must, for instance, be independent of our actual applications, judgments, and other relevant kinds of reactions. Assuming that the meaning of an expression is determined by its use, the question is: How could the use of an expression ever “ground” this kind of objectivity? Wouldn’t whatever an expression is applied to have to count as correct, thus undermining the very idea of correctness? What I am going to investigate in the remainder of this paper is how meaning as determined by charity fares with respect to the objectivity problem.

I am going to take for granted that charity does not preclude the ascription of a plausible amount of mistakes to normal human speakers. It is, for instance, perfectly possible for the radical interpreter to have sufficient evidence for interpreting a speaker’s predicate ‘spunk’ as expressing the concept beetle, thereby ascribing a mistaken belief when the speaker applies ‘spunk’ to the occasional spider. What I shall look at are rather situations in which the radical interpreter cannot for the life of her figure out what an alien speaker could possibly mean by what seems to be a predicate. For it at least appears to be possible to construe t-theories for such “speakers” that satisfy charity – but

But these arguments do not seem to go via radical interpretation; in fact, they seem quite direct to me. Here’s another, as far as I can tell clearly arguing from what is possible regarding belief to what it means to ascribe beliefs:

“Beliefs are identified and described only within a dense pattern of beliefs. I can believe a cloud is passing before the sun, but only because I believe there is a sun, that clouds are made of water vapor, that water can exist in liquid and gaseous form; and so on, without end. No particular list of further beliefs is required to give substance to my belief that a cloud is passing before the sun; but some appropriate set of related beliefs must be there. If I suppose that you believe a cloud is passing before the sun, I suppose you have the right sort of pattern of beliefs to support that one belief, and these beliefs I assume you to have must, to do their supporting work, be enough like my beliefs to justify the description of your belief as a belief that a cloud is passing before the sun.” (Davidson 1977, 200)
without allowing the interpreter to understand the speaker. Here is Pagin’s case of *Casey and Alien*:  

Assume that interpreter Casey from Earth embarks on the interpretation of apparent speaker Alien from Outer Space. Casey identifies a candidate predicate $\Phi$ that seems applied to some objects and withheld from others by Alien, but Casey sees no pattern in the usage. None of the property concepts Casey can come up with matches even approximately the pattern of Alien’s applications.

Casey then decides to learn from Alien, and starts defining a new predicate $F$ in his own language. It is defined by cases: true of objects that Alien applies $\Phi$ to, false of objects that Alien withholds $\Phi$ from, and for all objects $b$ *unconsidered* by Alien, $F$ is true of $b$ just in case $b$ is a rocket. Clearly, by interpreting $\Phi$ to mean $F$, and assuming Casey has identified atomic sentences with $\Phi$ as predicate and a demonstrative as subject term, Alien’s demonstrative $\Phi$ sentences all come out true.

Casey then goes on to do the same with other predicates, and also with what he identifies as grammatical particles, and sentence constructions. For each sentence held true at a time, on a case by case basis, an interpretation is given of the parts and the syntactic operations that makes the sentence come out true at that time. Some arbitrary interpretation is provided for all cases not considered by Alien. So Casey’s meaning theory is compositional and complete (with respect to the syntax he has identified), and results in only true beliefs being attributed to Alien. (Pagin 2013, 236)

Casey’s method – though ingenious – is no good as a method of interpretation. It is no good in two respects. First, despite appearances to the contrary, Alien might in fact not be speaking any language at all. Using Casey’s method, there is no way an interpreter could come to the conclusion that what initially appears to be speech behaviour in fact is not. But, second, even if we assume that Alien is speaking a language, Casey’s method is faulty: Casey’s theory clearly does not allow him to understand Alien.

Take $\Phi$. What Casey is lacking when it comes to $\Phi$ is an *independently possessed concept* that would allow him to subsume at least a (weighted) majority of the objects Alien appears to apply $\Phi$ to. As long as Casey is unable to come up with, or form, such a concept to interpret $\Phi$ as expressing, he will not understand Alien. And Casey will not be able to come up with, or form, such a concept as long as he cannot detect any similarity in the $F$ objects.

---

28 Pagin here makes use of the old idea that the radical interpreter, when faced with the task of interpreting speakers of a language with conceptual resources much more advanced than her own, can learn from the speakers and acquire their concepts in the process of radical interpretation, adding new terms for those concepts to her own, i.e. the meta-language in which she formulates her t-theory. Cf. Harman 2012, 17.

29 But isn’t the main problem with the theories Casey’s method delivers that they don’t make the right predictions? No, not really. First, it isn’t so clear whether Davidsonian t-theories are even supposed to be predictive. “Fit” is a merely retrospective notion, and, as we saw above, it is some supposed totality of data, “past, present, and future”, that according to Davidson determines meaning. Once you have that totality, there is nothing left to predict (cf. Pagin 2013, 237). Pagin (1999) emphasizes that the method of applying charity is a matter of getting a best fit to
Charity thus appears to allow meaning assignments that do not secure objectivity. An object (that Alien has considered) satisfies $\Phi$ precisely in case Alien has applied $\Phi$ to it. Objectivity couldn’t be much further away. What’s special about Casey is the combination of showing that, while this kind of “meaning determination” appears to be compatible with charity, it has the consequence that a predicate need not express anything recognizable as a meaning whatsoever. And the root of the problem here seems to be precisely that the “meanings” assigned by Casey’s method do not secure objectivity: there is no sense in which these meanings make the correctness of Alien’s applications of $\Phi$ an objective matter, independent of Alien’s actual applications. Let’s call meanings that do secure objectivity “objective meanings.”

We can then summarize Casey’s case as follows: Satisfying charity does not seem to guarantee that the meanings so determined are the genuine, “objective” article. Meanings that are both public and objective, or so it now seems, require detectable similarity.

We need to be careful here, however. If we look beyond its early formulations, there is a good question whether Casey’s method really satisfies charity. Take the following passage (again):

The process of separating meaning and opinion invokes two key principles which must be applicable if a speaker is interpretable: the Principle of Coherence and the Principle of Correspondence. The Principle of Coherence prompts the interpreter to discover a degree of logical consistency in the thought of the speaker; the Principle of Correspondence prompts the interpreter to take the speaker to be responding to the data and, thus, to achieve accommodation, as opposed to prediction. He argues that because of this, compositionality cannot be justified from charity.

And second, even if Casey had the relevant totality of data, and thus would not need to make any predictions, his theory would be useless. The main problem of Casey’s theory is not that according to it, unconsidered objects fall under $\Phi$ if they are rockets: Even if Casey had observed Alien consider every object, he wouldn’t understand Alien precisely because he doesn’t detect any similarity in the things that Alien does apply $\Phi$ to.

If Alien hasn’t considered an object, whether it satisfies $\Phi$ depends, not on Alien’s application, but on whether it’s a rocket or not. So that part is fine.

This is because the “meanings” Casey’s theory assigns do not reflect “real”, independently possessed, concepts of his. There are perfectly kosher concepts in the vicinity, of course: For $\Phi$, there is the concept of either being such that Alien has considered and applied $\Phi$ to one or such that Alien has not considered one and being a rocket. That, however, clearly is not the concept, if any, Alien expressed when using $\Phi$. Nor is it the concept expressed by $F$ as defined by Casey, even though these are co-extensive.

Once we recognize the need for using our own concepts in interpretation, considerations like these can be turned into an argument for semantic holism: Even if we require our semantic theory to be compositional, and its composition rules to be projectable, we would need to accept intuitively senseless interpretations if we didn’t construe the determination relation between data and meanings as one of global best fit. It is only if meaning is determined holistically, and by means of a many-to-one determination relation allowing for the ascription of false belief, that the possibility of interpretation in terms of our own concepts is secured. For an argument like this, see my (2001).
same features of the world that he (the interpreter) would be responding to under similar circumstances. Both principles can be (and have been) called principles of charity: one principle endows the speaker with a modicum of logic, the other endows him with a degree of what the interpreter takes to be true belief about the world. (Davidson 1991, 211, emphasis added)

Here, the principle of correspondence can plausibly be read as building detectable similarity right into charity: It requires the interpreter “to take the speaker to be responding to the same features of the world that he (the interpreter) would be responding to under similar circumstances.”

But if charity requires detectable similarity, it might seem as though we have hopped out of the frying pan – and right into the fire. This is because detectable similarity here is similarity detectable by the interpreter. This, it might seem, would take the objectionably “subjective” element from Casey’s method and replace it with another, potentially equally objectionable “subjective” element: the interpreter’s own sensitivity to the similarity in question. If charity requires detectable similarity, the worry is, the interpreter plays an essential role in meaning determination. She couldn’t be, as David Lewis once put it, “a way of dramatizing our problem – safe enough, so long as we can take it or leave it alone” (Lewis 1974, 334). And if the interpreter plays an essential role in meaning determination, we are stuck with the objectivity problem. Or so the worry goes.

5 Objectivity and Sensitivity

I think it is very plausible to think that Davidson did incorporate detectable similarity into charity.\(^\text{33}\) And ultimately, that does mean that charity falls short as an answer to the rule-following considerations. Meanings, as determined by charity, won’t be fully and pristinely objective. But it is important to

\(^\text{33}\) He more and more came to stress the causal (and social) externalist element in his account of meaning and content determination, construing the “features of the world” that both speaker and interpreter react to as the “common causes” of these reactions: “Communication begins where causes converge” (Davidson 1983, 151) could be the slogan of this development. This kind of externalism requires shared sensitivities, commonly detectable similarities across speakers and their interpreters:

“What makes these the relevant similarities? The answer again is obvious; it is we, because of the way we are constructed (evolution had something to do with this), who find these responses natural and easy to class together. If we did not, we would have no reason to claim that others were responding to the same objects and events (i.e. causes) that we are. It may be that not even plants could survive in our world if they did not to some extent react in ways we find similar to events and objects that we find similar. This clearly is true of animals; and of course it becomes more obvious the more like us the animal is.” (Davidson 1990a, 202)

Similarly, Quine came to speak of “a preestablished harmony, between your standards of perceptual similarity and mine” that is required for any “meeting of minds” (Quine 2000, 2; see also Quine 1995, 21f).
understand precisely what kind of contribution the interpreter makes to meaning determination.

Most significantly, what requiring detectable similarity amounts to is a restriction – not on the determination base, but on the “things” eligible to be meanings. Think of Casey again. What his theory assigns to $\Phi$ doesn’t intuitively qualify as a meaning – in the sense of something we can mean, i.e. something that can correctly be assigned to a predicate of natural language as its semantic value. This is not because it’s not a property – for all I care, it’s a property alright. Nevertheless, it does not seem to be the right kind of property – for there to be detectable similarity, a property needs to be such that the interpreter is sensitive to its being instantiated. To be eligible to be meant – i.e., such that a predicate is satisfied by an object iff the object has it – a property needs to be what I shall call “interpreter-sensible.”\(^{34}\) But that would seem to imply that we get “objective” meanings only if we restrict the realm of what can be meant to the properties we are sensitive to. That is, the correctness of the applications of our words can be “objective” only to the extent that we have already restricted what can be meant to those properties we can detect.

And this would mean that correctness of application ultimately is not fully objective, not totally independent from our judgments. This is a failure of objectivity in a more subtle sense than the most flagrant one. Objectivity fails in the most flagrant sense if, given that we mean something by an expression, correctness coincides with whatever we apply it to. It fails in a more subtle sense if what is eligible to be meant (at least in the basic cases) coincides with what we can detect, i.e. have an ability to judge correctly. Of course, an ability to detect the instantiation of a property does not require being correct in every single ascription of it. Nevertheless, if we restrict the eligible to the detectable, correctness of application will no longer be fully independent from our judgments – there wouldn’t be correctness without the ability to judge correctly.

Talking of eligibility here is of course meant to direct our thoughts to David Lewis. If what we are after is a restriction of the things eligible to be meant, we ought to consider whether Lewis doesn’t have a solution on offer, an impeccably objective solution moreover. The idea would be to use Lewisian natural properties to secure detectable similarity: Properties need to have a certain degree of naturalness to be eligible to be meant, and properties such as those assigned by Casey are too unnatural.

In his 1983 paper “New Work for a Theory of Universals”, Lewis at least seems to be suggesting such a constraint on meaning determination when he writes

\(^{34}\) This is not meant to suggest that it has to be what usually is called a “sensible property”, i.e. a property directly detectable by perception. The interpreter just needs to have some sort of ability to “detect” it, not necessarily a purely perceptual one.
that the saving constraint concerns the referent – not the referrer, and not the causal channels between the two. It takes two to make a reference, and we will not find the constraint if we look for it always on the wrong side of the relationship. Reference consists in part in what we do in language or thought when we refer, but in part it consists in eligibility of the referent. And this eligibility to be referred to is a matter of natural properties. (Lewis 1983, 371, see also Lewis 1984, 226ff)

That naturalness might help precisely with detectable similarity is further suggested when Lewis gives general job descriptions for the natural properties such as the following:

Natural properties would be the ones whose sharing makes for resemblance, and the ones relevant to causal powers. … Let us say that an adequate theory of properties is one that recognises an objective difference between natural and unnatural properties; preferably, a difference that admits of degree. (Lewis 1983, 347)

Eligibility then is supposed to be one of the measures by which to rank assignments of reference, meaning, or content. And for Lewis, just as for Davidson, the constraints or measures determining the ranking are constraints on whole theories (or assignments, interpretations), so we need a way of determining the degree of eligibility a whole theory has on the basis of the eligibility of the individual assignments of content or reference.35

It is, I take it, fairly clear that the account of meaning determination Lewis suggests to save by means of naturalness is not the one he endorses.36 He does, however, suggest introducing an analogous constraint into his own account of the determination of propositional attitude content (cf. Lewis 1983, 373ff). And, most intriguingly in our present context, he also suggests dealing with “Kripkenstein’s puzzle” by means of an appeal to natural properties:

[W]e must pay to regain our naiveté. Our theory of properties must have adequate resources to somehow ratify the judgement that instances of adding are all alike in a way that instances of quadding are not. The property of adding is not perfectly natural, of course, not on a par with unit charge or sphericality. And the property of quadding is not perfectly unnatural. But quadding is worse by a disjunction. (Lewis 1983, 376)

Thus it is fairly clear that Lewis considers it perfectly fine to use constraints using the naturalness of properties in accounts of meaning or content determination if need be. My question is whether a Davidsonian account of meaning determination can use naturalness to restrict the properties eligible to be meant. More precisely, my question is whether such an account can use naturalness to

---

35 When it comes to how precisely to do that, Lewis doesn’t say much more than the following: “Ceteris paribus, an eligible interpretation is one that maximises the eligibility of referents overall” (Lewis 1984, 227). Nothing here will hang on the details, but those interested in how this might work can look at Williams 2007. He suggests a method for determining eligibility values for whole interpretations, and provides discussion.

36 For relevant discussion, see Schwarz 2014; Weatherson 2013.
restrict the eligible properties in a way that does not essentially involve the interpreter.

I have investigated this question at greater length elsewhere (cf. Glüer 2016). Here, I shall only give the bare bones of that discussion. The crucial point is the following: In the context of Davidsonian meaning determination, constraints need to be if not motivated by, then at least compatible with, the public nature of meaning. Restricting the eligible meanings now appears to be motivated in precisely this way: What Casey teaches us is that for meaning to be public, it is not enough that the determination base be a certain way. It is also required that only certain meaning assignments are eligible. The question is whether restricting what can be meant to the *natural properties* can be motivated in this way – while keeping the interpreter out of the picture, so to speak.

I have argued that this is not the case. The attempt to do so faces a dilemma. What we are trying to do is to keep the interpreter out of determining the eligible properties while at the same time securing detectable similarity, i.e. making sure that the interpreter is sensitive to the eligible properties. As far as I can see, there are three ways in which Lewis characterized the natural properties. On two of them, detectable similarity can be secured, but the interpreter is very much involved in characterizing the properties as natural. On the third, the interpreter is out of it – but so is the guarantee for detectable similarity. Thus, it seems that it is either-or: We can either keep the interpreter out or secure detectable similarity. We can either have objectivity, that is, or sensitivity. This seems to be an instance of a more general dilemma. Let’s call it the “dilemma of objectivity and sensitivity.”

As I said, Lewis makes use of three ways of characterizing properties as more or less natural. One is by providing examples:

The mereological sum of the coffee in my cup, the ink in this sentence, a nearby sparrow, and my left shoe … is an eligible referent, but less eligible than some others. (I have just referred to it.) Likewise the metal things are less of an elite, eligible class than the silver things, and the green things are worse, and the grue things are worse still – but all these classes belong to the elite compared to the counted utterly miscellaneous classes of things that there are. (Lewis 1984, 227)

A second way of characterizing the distinction proceeds by reflecting on what is more or less rational to believe and desire (cf. e.g. Lewis 1983, 375). And third, there are passages where Lewis suggests that there are “perfectly natural properties” which it is up to physics to discover. Examples he gives are mass, charge, and quark colour and flavour (cf. Lewis 1984, 228). Other properties can then be defined in terms of the fundamental physical properties. And degrees of naturalness can be characterized as follows: the longer

---

37 In my (2016), I construe the verdict regarding Lewis as an instance of what I there call “the dilemma of purity and sensitivity.”
the chain of definability between a property and the perfectly natural properties, the less natural it is.

It should be clear that the first two methods of characterizing naturalness rely on ourselves and our reactions as a kind of black box. So understood, the natural properties would indeed seem to be properties an interpreter would be sensitive to, but by the same token our understanding of naturalness remains dependent on precisely those sensitivities. Not so on Lewis’s third way – no reference to ourselves and our sensitivities appears to be involved in characterizing naturalness in terms of the length of chains of definitions. But by the same token, there is no longer any guarantee for these properties being interpreter-sensible. Once we identify the perfectly natural properties as the fundamental physical properties, sensitivity to naturalness becomes a metaphysical coincidence.

The dilemma we face when trying to restrict eligibility in terms of naturalness can plausibly be expected to generalize: Whatever makes for eligibility needs to be such that detectable similarity does not become mere coincidence. This would seem to require characterizing eligibility in a way that essentially relates it to an interpreter’s sensitivities. And this, in turn, would mean that the problem of objectivity is with us to stay. It was precisely its epistemico-metaphysical nature that seemed to provide the radical interpretation account of meaning determination with the additional resources required for answering the rule-following considerations. And indeed, the account does have resources towards justifying choices of determination base and principle, choices that otherwise might seem totally arbitrary. Nevertheless, it is its very epistemico-metaphysical double-nature that ultimately prevents the account from escaping the problem of objectivity. Interestingly, it is not primarily the determination base – the form of “use” the account works with – that turns out to be problematic. Rather, objectivity remains elusive because publicness not only requires the data to be public. Publicness also turns out to require restrictions on what is eligible to be meant. Ironically, it turns out that we get “objective” meanings only if we restrict the realm of what can be meant to the properties we are sensitive to. Unless we find a way of preventing the dilemma of objectivity and sensitivity from generalizing, the problem of objectivity therefore cannot be solved by means of charity.38

38 While this means that charity ultimately does not answer the rule-following considerations, it doesn’t have to mean that a radical interpretation account of meaning determination is doomed. This depends on what precisely we expect such an account to be like. Importantly, this result also does not mean that the eligible properties have to be construed as relational properties, for instance as dispositions to elicit certain reactions in (suitable) interpreters (under certain circumstances). It does mean that we cannot just identify them with the categorical bases of these dispositions. But we can instead construe them as the properties that make objects disposed to elicit those responses – where what that description refers to, or is
Once we adopt the Davidsonian epistemico-metaphysical take on meaning, the problems generated by the rule-following considerations look somewhat different from what we have become used to. The main concern is meaning determination, but what appears threatened is not primarily the “naturalist” or “reductivist” character of the base – what looms large is rather the problem of objectivity. What the rule-following considerations teach us is precisely to appreciate the metaphysical consequences of meaning’s peculiarly epistemic nature. As Åsa Wikforss argues in her contribution to this volume, this is precisely one of the fault lines between the early and the later Wittgenstein; cf. this volume, Chapter 3.

We might need to learn to live with meanings not being fully and totally objective – in the “platonist” sense the rule-following considerations are premised upon. Wittgenstein himself, of course, went further than anything suggested here – he seems to have concluded that meaning requires the “common behaviour of mankind” as a frame of reference (PI 206), quipping

If a lion could talk, we could not understand him. (PI II, X, 223)

Nothing I have argued here forces so drastic – or so paradoxical – a conclusion. What I have argued, however, is that a radical interpretation account of meaning determination does require a background of shared sensitivity, of commonly detectable similarity.

References


satisfied by, varies across possible worlds, depending on which property is the categorical base of the relevant disposition in the world in question. For details, see Glüer 2016. There is consequently no sense in which meaning – whether in the sense of “having a meaning” or in the sense of “being a property meant” – has to be understood as response-dependent (in Wright’s sense; cf. Wright 1988, see also Johnston 1992) on a Davidsonian account. See above, fn. 14.

As Åsa Wikforss argues in her contribution to this volume, this is precisely one of the fault lines between the early and the later Wittgenstein; cf. this volume, Chapter 3.

Some commentators take Wittgenstein’s main point to be the more radical one that the very “idea of determination is incoherent” (Goldfarb 2012, 78) and that, thus, no account of meaning whatsoever is possible. I am with Verheggen 2003 here in thinking that Wittgenstein rather points to the need to gain a better understanding of what kind of account of meaning is possible.

For most helpful discussion and comments, I would like to thank Claudine Verheggen, Peter Pagin, Åsa Wikforss, Bill Child, and audiences in St. Andrews, Stockholm, Barcelona, and Toronto.


5 Davidson’s Treatment of Wittgenstein’s Rule-Following Paradox

Claudine Verheggen

1 Introduction

There is large agreement among commentators on Wittgenstein and Davidson that they both see an intimate connection between the ways in which we use our words and what we mean by them. More controversial already is the question whether they also conceive of the use that is essential to meaning in the same way, and whether they should thereby be understood as advocating non-reductionism, the rejection of any account of meaning that does not invoke semantic notions. Even more controversial is the question whether, assuming non-reductionism is accepted, some constructive philosophical work on meaning can still be accomplished. Wittgenstein is often considered as a quietist, a label Davidson would certainly not embrace. My aim in this chapter is first to show that they both fiercely argue for non-reductionism in similar ways, and that consequently they conceive of the use that is essential to meaning in a similar way. In a nutshell, they both think that a full account of meaning requires us to consider this use within a semantic context, so that we cannot say what speakers mean by their words, and what words mean, without saying what speakers use their words to mean, and we cannot answer the question what makes it possible for someone to have a language without thinking of her as already having one. However, whereas Wittgenstein makes only very general remarks about the kind of use he takes to be essential to meaning, Davidson has much more to say about the topic and, as a result, provides a significantly richer and more constructive way to address the paradox about meaning and rule-following developed by Wittgenstein.

I believe that the rule-following paradox is the centrepiece of Wittgenstein’s *Philosophical Investigations* and, for reasons that should become clear, I believe that considering this paradox remains the best way to start investigating meaning. Indeed, ignoring it cannot be an option for anyone reflecting philosophically about meaning. Not surprisingly, therefore, there has been much debate, especially since Kripke (1982), as to what kind of treatment Wittgenstein’s paradox should receive. Should the sceptical problem it
embodies be dissolved by showing that it rests on untenable assumptions? Could the sceptical problem be given a straight solution, the kind of solution the sceptic is seeking? Or, in so far as the problem calls for a solution, could the solution be only sceptical, conceding that the sceptic’s demand cannot be satisfied – the kind of account she is seeking cannot be obtained – and offering an alternative account of meaning? I believe the problem had better be dissolved, resting, as it does, on an erroneous, indeed, incoherent, conception of the kind of account of meaning we could secure. However, contra many commentators on Wittgenstein, I also believe that dissolution does not entail quietism. This, as we shall see, is on a par with the claim that non-reductionism does not entail quietism. In fact, I take it that Davidson would agree that the problem is better dissolved. Yet, what can be construed as Davidson’s treatment of the paradox is, to repeat, much more elaborate and instructive than Wittgenstein’s, especially than what many commentators on Wittgenstein take it to be.

I shall start by reviewing the considerations Wittgenstein advances in support of the paradox, and I shall argue that they overlap with those that motivate Davidson’s account of meaning. Wittgenstein’s are however more encompassing than Davidson’s and, as a result, can be understood as solidifying Davidson’s approach. I shall then articulate Davidson’s response to a problem that is left standing after the paradox itself has been repudiated.

2 Wittgenstein’s Paradox

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 everything 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. (Wittgenstein 1953, §201)

Wittgenstein talked about rules in part because he was much concerned with mathematical rules, not to mention games, but he did intend the paradox to apply to any rule-governed activity (in the minimal sense of ‘rule-governed’ to be elucidated below) and, in particular, to the use of language. Thus, applied to linguistic meaning, the paradox becomes:

No application of a linguistic expression could be determined by the meaning of the expression, because every application of an expression can be brought into accord with its meaning. The answer was: if every application can be brought into accord with the meaning of the expression, then it can also be brought into conflict with it. And so there would be neither accord nor conflict here.

On the face of it, though the paradox is straightforward, it is also rather puzzling. It is, on the one hand, straightforward because a meaningful
expression that is always applicable correctly, or not, as the case may be, is not a meaningful expression. If an expression is meaningful, it is subject to conditions of correct application – we might say that its applications are governed by these conditions, hence my talk of rule-governed activity above. For instance, if ‘green’ means blue for a speaker, then the speaker applies ‘green’ correctly when, and only when, she applies it to blue things. I give this odd-sounding example on purpose, to stress the sense of “correctness” which is at play here, which has to do with the semantic relations between words and extra-linguistic reality, and not with the question whether the speaker is using words in accord with the conventions of her community. A speaker who used ‘green’ to mean blue would be incorrect according to the standards of English, but not incorrect according to the standards governing her use of the word, when she says, “the sky is green”, if the sky is indeed blue when and where she says it. What needs to be stressed here, and what I take Wittgenstein to be stressing, is that the meaning of expressions and their conditions of correct application, understood strictly as stated above, are inextricably connected. But then, on the other hand, the paradox is rather puzzling. For how could a meaningful expression fail to determine which of its applications are correct or not?

Wittgenstein gives us a clue as to what may have led to the paradox when he says, in the paragraph following its statement, that “what we thereby show is that there is a way of grasping a rule which is not an interpretation” (Wittgenstein 1953, §201). This indicates that Wittgenstein has been investigating various pictures or theories of meaning, all of which, he now maintains, make grasping the meaning of an expression dependent on interpreting it in a certain way. This in turn suggests that, according to those pictures or theories, in order to grasp the meaning of an expression, I must grasp something that confers on the expression its meaning. And it is that something that requires interpretation if it is to do its meaning-providing job. To put it slightly differently, this suggests that, according to those pictures or theories, expressions have meaning in virtue of being associated with something else, which, again, must be interpreted in order to do its meaning-providing job. Thus the rule or meaning referred to in the statement of the paradox is better understood as the expression of the rule or meaning, whose meaning-provider we must grasp in order to regard it as a rule or a meaningful expression. The trouble, according to Wittgenstein, is that all the meaning-providers we can think of can always be interpreted in such a way that all applications can be deemed to be correct, or incorrect, as the case may be, making it impossible for expressions to be governed by conditions of correct application, and thus making it impossible for expressions to be meaningful after all. Even if the meaning of a word can

1 This has been widely recognized as a platitude. See, e.g., Blackburn (1984), Wright (1986), Boghossian (1989), Hattiangadi (2006) and Glüer and Wikforss (2009).
sometimes be given by interpreting it, that is, by explaining it in other words, as Wittgenstein says, “Interpretations by themselves do not determine meaning” (Wittgenstein 1953, §198), for interpretations themselves need to be provided with meaning. And thus eventually something other than interpretations will have to be appealed to in order to explain what it is that determines the meaning of expressions.2

What we need to investigate at this stage are the pictures or theories that, according to Wittgenstein, are committed, presumably unbeknownst to their advocates, to the claim that grasping the meaning of an expression depends on interpreting it in a certain way, and which are thus subject to the paradox. Just what are these theories maintaining, which commit them to what turns out to be a devastating claim? This is what Wittgenstein spends the first 200 sections of Philosophical Investigations unravelling.

3 The Sources of the Paradox

To repeat, I take the main point of the paradox to be that some theories of meaning simply cannot make room for the distinction between correct and incorrect applications of linguistic expressions and that they therefore fail lamentably as theories of meaning.3 I take it that this necessary condition on meaningfulness has been on Wittgenstein’s mind from the beginning of his investigations, even though he does not focus on it until section 138 of Philosophical Investigations. I also take his basic starting point to be the question made explicit by Davidson: “what is it for words to mean what they do?” (Davidson 1984, §xv) That this is also Wittgenstein’s starting point, is, I think, supported by the first “picture” of meaning he examines, which is part of Augustine’s answer to the question, and a crude version of his own early Tractarian view. According to this “particular picture of the essence of human language”, as Wittgenstein calls it, “[e]very word has a meaning. The meaning is correlated with the word. It is the object for which the word stands” (Wittgenstein 1953, §1). This picture, prima facie at least, is pretty commonsensical. As Wittgenstein remarks, though, initially what we think of as instances of words whose meanings are the objects they stand for are mostly “nouns like ‘table’, ‘chair’, ‘bread’, … people’s names, and … the names of certain actions and properties” (Wittgenstein 1953, §1). In short, what we think of are mostly words for the familiar features of the world surrounding us. This

2 Note that throughout I shall be using ‘determine meaning’ and ‘provide with meaning’ interchangeably.

3 They cannot account for a trivial kind of normativity, which is not to be confused with other, more robust kinds of normativity, which have been the subject of much debate in the past fifteen years. See, e.g., Wikforss (2001), Hattiangadi (2006), Glüer and Wikforss (2009), Verheggen (2011) and Whiting (2013).
referentialist picture\(^4\) is however beset with difficulties, including, as Wittgenstein notices, the fact that it cannot apply to all kinds of words which do not stand for anything in the physical world or even for anything at all. But what is more serious is that the picture cannot be true of any kind of word in the end. It cannot be true of any word that its meaning is determined simply by the thing for which it stands or with which it is correlated. For the immediate question is, what is that thing, how is the correlation established?

The first inclination is to reply: by pointing. One points at the feature of one’s environment that the word is supposed to stand for. But the question is, what feature is that? Let us say that, as in Wittgenstein’s example, someone points at a group of nuts and says, “two”. What is ‘two’ supposed to stand for? The number two? This particular group of nuts? Or someone points at an object and says “sepia”. Does ‘sepia’ stands for the colour of the object? Its material? Its shape? Or someone says “Henry”, pointing at a person. Is this the name of a person? A colour? A race? A point of the compass? As Wittgenstein concludes, “an ostensive definition can be variously interpreted in any case” (Wittgenstein 1953, §28). One way to settle on an interpretation is by adding, “this number is called two, this colour sepia, this person Henry”. But what about these new words, are their meanings also provided by what they stand for or are correlated with? What is this? How are these new correlations to be established? Obviously, as we will have to settle on interpretations of these new words, more words yet will have to be introduced, words which will in turn have as their meanings the things they stand for. Which things? All of this is to say that standing for, or being correlated with, some feature of the world around us cannot be all there is to a word having the meaning it has, even for those words that can be said to stand for things in the world around us.

Note that there are two problems here. There is first an indeterminacy problem or that of picking out the relevant aspect of the thing pointed at. If there is no principled way to do this, then one is always free to pick out an aspect of the thing pointed at that makes any application of the word with which it is connected correct, or incorrect, as the case may be, so that no conditions of correct application can be established. On the other hand, if there is a principled way to pick out the relevant aspect, this way is presumably linguistic. But then we have a regress problem or that of accounting for the meanings of the words used in solving the indeterminacy problem. The indeterminacy problem is, however, the fundamental one,\(^5\) as eventually we will be left with no words to

---

\(^4\) As Fogelin (1996, 37) first called it.

\(^5\) Which problem is the fundamental one, and how the problems are related, are, as all things Wittgensteinian, a matter of dispute. Thus George Wilson, contra McDowell (1992), thinks that the central problem is an indeterminacy problem (Wilson 1998, 111, fn12). Meredith Williams distinguishes between the “Infinite Regress Argument” and the “Paradox of Interpretation” and argues that the Regress “shows that in the end meaning must be something other than an act of
explain the meaning of other ones, and we will just have “to emit an inarticulate sound” (Wittgenstein 1953, §261). In short, the moral of the discussion so far is that, at bottom, there is no way that mere connections between words and things in the world around us can be what determine the conditions under which words are correctly applied and so what determine their meaning.

It might be thought, prima facie at least, that there are several ways to address the foregoing predicament. A little reflection shows that none will do, however. Thus consider three possible replies. First, it might be said that the problem lies in there being only one single act of pointing. Several such acts would do to isolate the relevant aspect, for it would be the aspect all the things pointed at have in common. But how is one to determine this common aspect, as any one thing shares many aspects with others? It cannot be retorted that the relevant aspect is the same as the one isolated by a previous pointing, for the question arises how the aspect was isolated on that previous occasion. And, of course, it cannot be said that the aspect is the same as the one isolated on the very first pointing, since this isolating precisely did not occur. A second reply is to say that the problem lies in the act of pointing being individual. If the pointing was performed simultaneously by several people, the relevant aspect could be isolated – it would be the aspect they are all pointing at. But this reply pretty obviously will not do. No matter how many people are pointing together, the aspect of the thing pointed at will remain as ambiguous as it is with a single person pointing. Are they together pointing at the colour of the object, or its material, or its shape? And having the group of people pointing together repeatedly will incur the same objections as those incurred with having a single person pointing repeatedly. A final reply has it that the relevant aspect of the thing pointed at is in fact easy to isolate: it is the aspect one person, or a group of people, has in mind while pointing. But what is it for someone to have a particular aspect in mind? Determining this is just as problematic as determining what aspect of a thing one is pointing at, if one has no language yet with which to do it.6

The above considerations will have been very familiar to many philosophers and commentators on Wittgenstein in particular. Old news though they may be, however, I do not think that they have been properly assimilated and recognized as containing the seeds of Wittgenstein’s non-reductionism. Before I say more about this, I shall continue reviewing the other potential determinants of meaning Wittgenstein examines and rejects, for they all lead to the paradox, and they do so fundamentally for the same kind of reason.

interpretation or that which requires interpretation. The Paradox builds on this to establish a more radical conclusion, namely, that the very distinction between correct and incorrect collapses from within the Classical View: We have no standard for correctness at all” (Williams 1999, 159). I have myself changed my mind on which problem is the fundamental one – see Verheggen (2003).

If familiar objects and features of the world surrounding us will not do as determinants of meaning, what about essences or definitions of those things? What about something, whatever exactly it is, that captures what all the members in the extension of a given general term have in common, something, we may say, that provides the necessary and sufficient conditions for its applications? After all, we may think, there must be such a thing which is present when the applications of the term are correct and absent when they are incorrect.

Wittgenstein argues that such determinants of meaning – essences or universals – are neither needed nor useful. He does this in a myriad of ways, by means of examples that show that the applications of words are rarely sharply circumscribed, yet this in no way jeopardizes our understanding and our ability to communicate by means of them. The part of the argument I wish to focus on is that which shows not just that meanings are not determined by essences or universals but that they could not be determined in that way. On the contrary, the claim that the meanings of words are provided by essences leads to the paradox and hence to the denial of meaning.

For the question is, how are we supposed to select those essences? Which are to be associated with which words? Wittgenstein compares grasping them to having “general” samples in one’s mind, like a sample “of what is common to all shades of green”, say “a sample of pure green” (Wittgenstein 1953, §73). The trouble is that any such sample could be interpreted either as a sample of all that is greenish or as a sample of pure green. Moreover, what is to prevent us from interpreting the sample as a sample of the particular shape it has, or as a sample of everything that has a shape for that matter? A sample is not a sample of any particular kind of thing until it is used, that is, understood or regarded, as such. For any sample, it could be interpreted in such a way that the applications of the word it is associated with are deemed to be correct, or incorrect – the indeterminacy problem. And if one thinks that one can settle on an interpretation, there is the regress problem looming again and the question with which samples the words used in the interpretation are to be associated, that is, the indeterminacy problem again.7

The discussion of samples allegedly determining the meanings of general terms and guiding their applications leads Wittgenstein to the related talk of “the rule according to which [we] proceed” (§82). There is no doubt that in many cases we can appeal to some rule to explain why we have applied a word in a certain way; for instance, we apply ‘chair’ to some objects made for sitting, though that very rule may also be useless in some cases, e.g., for “disappearing”.

7 As forcefully argued by Goldfarb (1997), the problem does not affect just terms that are family resemblance terms, contrary to what some commentators on Wittgenstein have thought. Rather, it affects all terms – no essences could determine their meaning.
More to the point, though, the question is, in what way does appeal to rules help us to give an account of what determines meaning in the first place? Here Wittgenstein compares rules to signposts or tables, which tell us which way to go or to look. However, he asks, “is there only one way of interpreting [a signpost]” (§85)? Is a schema not needed to explain the use of a table, and can we not imagine further rules to explain any particular explanatory schema (§86)? Not that a signpost usually leaves room for doubt or that a table is incomplete without a schema. The point again is that appeal to a rule is fine in so far as its interpretation is fixed – but if it is not, it can always be interpreted in such a way that the applications of the word it is supposed to be guiding are always correct, or not. But not all interpretations of rules can themselves be explained in terms of further rules. As before we have two problems: an indeterminacy problem and a regress problem, which is bound eventually to leave us with the indeterminacy problem.

I believe that, by section 89 of Philosophical Investigations, the argument against the idea that we can ever account for the meaning of words simply by invoking entities, whatever exactly these are, for which words stand or with which they are correlated or associated is in effect completed. But Wittgenstein has another go at it in response to an objection the interlocutor has made to the view Wittgenstein has been suggesting as an alternative to the various accounts of meaning he has considered, the view, loosely put, that “the meaning [of a word] is the use we make of the word” (§138). The interlocutor objects: “But we understand the meaning of a word when we hear or say it, we grasp the meaning at a stroke, and what we grasp in this way is surely something different from the ‘use’ which is extended in time” (§138)! “When someone says the word ‘cube’ to me, for example, I know what it means. But can the whole use of the word come before my mind, when I understand it in this way” (§139)? Well, Wittgenstein answers, “but on the other hand isn’t the meaning of the word also determined by this use? And can these ways of determining meaning conflict? Can what we grasp at a stroke agree with a use, fit or fail to fit it? And how can what is present to us in an instant, what comes before our mind in an instant, fit a use” (§139)?

Initially, it may not be altogether clear what the objection is here. Of course, when I understand the meaning of a word that I hear or speak, I do not have all of its uses in my mind – meaning cannot simply be use. Yet it may be determined by use, as Wittgenstein suggests, and so I need not have all of its uses in my mind when I understand it. What may be bothering Wittgenstein is that there is something puzzling to the idea that meaning is determined by use if it is the meaning of a word that determines, at least in part, which uses of the word are correct and which incorrect. After all, he has rejected all kinds of theories because the entities they postulate as determinants of meaning could not be invoked to explain why some applications of words rather than others are
correct, since these entities could always be regarded in such a way that any application could be deemed to be correct, or not. But now it looks as if the idea that meaning is determined by use may fare no better. Could use not be thought of in such a way that all the applications it determines are correct, or not? Thus the interlocutor’s worries prompt Wittgenstein to examine anew some traditional determinants of meaning, this time in a slightly more systematic way. (Interestingly, familiar features of the world are not reconsidered, presumably because they did not figure in any theory of meaning at the time, but also because of Wittgenstein’s focus on mathematical terms.)

I shall be brief with these, for now we are in even more familiar territory, as we are focusing on what has been regarded by most commentators as the core of the rule-following considerations. The first candidates for determinants of meaning Wittgenstein examines here are mental pictures, such as the drawing of a cube coming before one’s mind when one hears ‘cube’ (§139). The problem with these is that, not only are such pictures not necessary to determine the meaning of words like ‘cube’, they are also not sufficient. The same picture could dictate different applications (§140), all of which could be said to be correct – the indeterminacy problem again. And being provided with a schema that shows how to project the picture would not help, for different applications of the schema could also be imagined (§141) – the regress problem generating the indeterminacy problem again.

Turning to a different kind of example, viz., mathematical expressions, Wittgenstein addresses the claim that the correct application of an expression, say, the continuation of a series, is determined by an algebraic formula one has in one’s mind. Wittgenstein replies: “But this is where we were before. We can indeed think of more than one application of an algebraic formula; and while every mode of application can in turn be formulated algebraically, this, of course, does not get us any further. – The application is still a criterion of understanding” (§146). The interlocutor will not give up, however. He objects: How could what I have done so far determine what I mean? After all, my applications are finite and the series infinite (§147). Indeed, we may want to add, what I have done so far is compatible with my going on now in different ways, all of which depend on how I understand what I have done so far, and so all of which can be deemed to be correct. As already hinted, this suggests that use is in no better position than other traditional determinants of meaning and may face the same difficulty. So it is not as if we no longer have to explain how expressions can be governed by conditions of correct application once we have abandoned traditional theories and replaced them by some kind of use theory. Or is it? Before I answer this question, let me sum up the considerations that have led to the paradox.

Wittgenstein has considered several candidates as determinants of meaning and so as what might be appealed to in order to explain what determines the
correct applications of an expression, that is, as what these applications have to conform to in order to be correct. These were familiar features of the world around us, essences or ideal samples of those things, rules, mental pictures or formulas, and actual uses of expressions.\(^8\) There is no suggestion that these items have much in common. Indeed, there is no suggestion that none of them plays any role whatsoever in determining the meaning of expressions – as we shall see both external features, in the first instance, and uses, properly conceived of, do play such a role. The problem here has not much to do with the items themselves (which is not to say that different items may not face different problems in addition); rather, it has to do with our considering them by themselves, that is, independently of what we take them for and how we employ them in our linguistic practices. Thus the problem has to do with a certain conception of how they are supposed to provide words with meaning. It is the claim that they could do this independently of our saying how we apprehend them or what we take them to be, in effect, independently of specifying what meaning they are providing which words with. But we need a language in order to specify what items provide our expressions with meaning. That is, we need a language in order to be able to identify what is supposed to provide us with one to begin with.\(^9\) But then the idea of meaning resting solely on the association of words with extra-linguistic entities is in fact incoherent.\(^10\)

It should be stressed that the problem Wittgenstein has articulated is both very general and far-reaching. That the relevant aspect of a meaning-providing item cannot be isolated or identified without a language means, in effect, that any attempt to give a non-reductive account of meaning is bound to fail. This is so not simply because we evidently need a language in order to describe or characterize anything but because we need to say what meaning the item is providing if it is to be understood as providing any meaning. No description of what speakers mean by their words, and no explanation of what makes it possible for them to mean anything by their words, can be given outside of a semantic context, that is, without saying what speakers mean by their words and that they do mean something by them. In short, there is no answering the

\(^8\) It might be thought that one potential candidate, viz., dispositions, is conspicuously missing. I have left it out here because Wittgenstein himself did, but I have argued elsewhere that dispositions face the same kind of interpretation problem as other candidates – see Myers and Verheggen (2016, Chapter 2). For a focus on dispositions and the problems they generate, see Kripke (1982, 22–37).

\(^9\) This is true even if we think of the potential candidates as universals or essences, i.e., in effect, meanings. We still need a language to connect them with the expressions they are supposed to provide meaning with. This is emphasized in Wilson (1998).

\(^10\) The quietist may diagnose the problem differently, as having to do, not with our conception of how meaning-providers provide words with meaning, but rather with the very idea of a meaning-provider – see, e.g., Goldfarb (2012).
question what it is for words to mean what they do simply by appealing to things conceived of as meaningless.\textsuperscript{11,12}

Where does this all leave us? Is there anything interesting we can still say about meaning when it looks, in effect, as if we have been driven to the vacuous conclusion that all we can say is that what makes it possible for us to have a language is our having a language? No wonder Wittgenstein has been regarded as a quietist. However, as we shall see with Davidson, there is still quite a bit we can say to shed light on meaning. In fact, as we shall see first, even Wittgenstein has something to say.

4 Wittgenstein’s Treatment of the Paradox

Here is Wittgenstein’s immediate reaction to the paradox he has stated, this time in full: “what we thereby show is that there 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’” (Wittgenstein 1953, §201). Following a rule, he continues, is “a practice” (§202). Then presumably speaking meaningfully is a linguistic practice. Wittgenstein has thus not discarded the deep connection he sees between use and meaning. But there is a significant difference between his previous understanding of use and the present one. At least some of Wittgenstein’s talk of use, prior to his statement of the paradox, can be understood as talk of the production of signs that are not yet meaningful, as what needs to be added to “dead” signs, in order for them to become alive and meaningful, as he puts it in The Blue Book (1958, 4–5). It would be hard to understand his talk of use determining meaning (Wittgenstein 1953, §139) if use were to be conceived of there as meaningful – the talk would simply be empty. But, as we saw, uses conceived of as additions to signs in order to make them meaningful are just as subject to the sceptical problem as any other entities we may think of adding. The paradox has revealed that this conception of meaningfulness is incoherent. How exactly then has the sceptical problem embodied in the paradox been addressed?

\textsuperscript{11} Cf. Stroud’s discussion of Wittgenstein’s idea of “phenomenon” (Stroud 2012, 23).

\textsuperscript{12} The breadth of the problem is not always recognized. Indeed, the focus, following Kripke, on the so-called rule-following considerations has led many philosophers to argue that it is a certain conception of the entities that are alleged to determine meaning that generates the paradox. Thus, for instance, for Wilson, it is classical realism or what others have called semantic Platonism (see, among others, Pears (1988)). Alternatively, for Tim Thornton, for instance, the central guilty conception is what he calls representationalism, the idea that the meanings of terms are to be explained by appeal to mental states whose contents are provided by inner mental representations (Thornton 1998, 1–2). Now, there is no denying that these two views may have been in the foreground of Wittgenstein’s mind, especially given the options put forward in his days, indeed Platonism and traditional empiricism. But this should not make us lose sight of the fact that these are not the only culprits.
The paradox has it that whatever applications one may make of an expression can be deemed to be correct, or incorrect, as the case may be. Thus the paradox has it that no expression can ever be meaningful, since no conditions of correctness can govern its applications. As Kripke put it, following the paradox, “It seems that the entire idea of meaning vanishes into thin air” (Kripke 1982, 22). Of course, it does so vanish if we think that all it takes for an expression to have meaning is to be associated with something that provides it with meaning. These meaning-providers, conceived of in this way, are what the sceptic is seeking and not finding, hence the sceptical problem – how can we mean anything by any expressions? Now, Wittgenstein has argued that the account of meaning the sceptic recommends cannot be obtained; indeed, Wittgenstein has shown that the sceptic’s demand is incoherent. This, I think, is what distinguishes Wittgenstein from the sceptic, who despairs of finding the relevant meaning-providing items, but keeps regarding her quest as perfectly intelligible, and so keeps thinking that there is a sceptical problem that needs to be solved. Wittgenstein, on the other hand, rejects the sceptical problem by denouncing its main premise, that is, to repeat, by insisting that the idea of meaning resting solely on connections between words and extra-linguistic items is incoherent. It is not, as we shall see, that meaning determinants cannot be found. Rather we have to revise our conception of how they play a role in determining meaning. Thus the sceptical problem, resting as it does on an erroneous conception of how this is to be done, is dissolved rather than solved. But where does this leave us, philosophically? Is there still a problem left over about meaning that needs to be solved?13

If there is, it cannot be the problem to which the paradox was an answer, that is, it cannot be the question of how a rule or expression determines its correct applications. Once it has become clear that the determinants of meaning cannot be conceived of in themselves, independently of our interpretation of them, and thus that the rule or expression are to be thought of as being used meaningfully, this question no longer has any purchase.14 It had some only because we were considering them as not yet meaningful and had the wrong conception of what would make them meaningful. Now that we have seen that we cannot say anything about what makes them meaningful except that they are used meaningfully, there is no problem left; it is in virtue of what they mean, in virtue of their conditions of correct application, that expressions determine which of

13 ‘Sceptical problem and solution’ are Kripke’s words, but I think it makes perfectly good sense to apply them to Wittgenstein, at least in light of the interpretation I have given of what leads to the paradox. Note also that I am not claiming that my interpretation exactly coincides with Kripke’s. In many respects, pretty obviously, it does not.

14 It might be thought that there is still the question how correct applications can be determined objectively, but this question in turn has to do with how we get conditions of correct application to begin with.
their applications are correct, and which incorrect.\textsuperscript{15} Matters seem to have become too easy, however, and quite magical. Of course if an expression has conditions of correct application, there is no problem with saying how it determines its applications to be correct, or incorrect. The real question here should be what determines its conditions of correct application in the first place, what makes it possible for it to have any and to have the ones it does have. Even though the answer to this question will eventually have to involve meaningful use, perhaps something can be said about use, independently of its being meaningful, that sheds light on the topic. I take it that the following remarks of Wittgenstein’s are an attempt to do just that.

As already noted, Wittgenstein thinks of the use that is essential to meaning as a practice. But he also thinks of it as a “custom” and as a “technique” (Wittgenstein, 1953, §199). Thus, he writes:

[A] person goes by a sign-post [follows a rule] only in so far as there exists an established [regular] use of sign-posts, a custom. (§198, my emphasis)

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).

To understand a sentence means to understand a language. To understand a language means to have mastered a technique. (§199, first two emphases mine)

I take it that by “technique” Wittgenstein has in mind, at least in part, the idea that to speak meaningfully is to be engaged in a rule-governed activity in the sense that there is such a thing as doing it right or doing it wrong – expressions may be applied correctly or incorrectly. As he writes: “[W]e lay down rules, a technique, for playing a game” (§125). He also talks of the “technique of applying words” (§262, §557). As I have argued, the main point of the paradox is to condemn certain pictures or theories of meaning as incapable of making room for that distinction. But the notion of custom does shed further light on the way words must be used if they are to be meaningful: some of them at least must be used repeatedly. Though Wittgenstein does not say so at this stage of Philosophical Investigations, this seems to be a development of his earlier claim that an ostensive definition cannot by itself endow a word with meaning. Of course, an ostensive definition can do this if it is clear what role the word being defined plays in the language. That is to say, an ostensive definition can be used to endow a word with meaning once a language is already in place. But then not all words could be endowed with meaning by being ostensively

\textsuperscript{15} Thus, to this extent, I now agree with McDowell (1992, 40); I did not appreciate the ambiguity of Wittgenstein’s talk of use in Verheggen (2003).
defined – at least some of them must have been used on multiple occasions in order to be meaningful.  

I am of course not suggesting here, on Wittgenstein’s behalf, that what determines the meaning of at least some expressions is simply their having been used repeatedly in certain ways. This would be belying what I have been at pains so far to establish, which is that no reductive explanation of meaning is forthcoming and that using words in certain ways, in so far as these uses are thought of as not yet meaningful, cannot be all there is to speaking meaningfully. What I take Wittgenstein to be doing here is laying out a necessary condition for there to be meaningful expressions: at least some of them must have been used repeatedly. I think this reading is warranted by Wittgenstein’s talk of ‘only’ and ‘possible’ to formulate his claims, as well as his insistence that his explanation of how I react in a particular way to a sign – I do so because I have been trained to do so – is not causal but constitutive – this is what “following-the-sign really consists in” (Wittgenstein 1953, §198). The notions of training and teaching figure prominently in Wittgenstein’s positive remarks about what it is to follow a rule or mean something by an expression. “One is trained to do so” – i.e., to follow a rule or obey an order (§206). This indicates that he also thinks that our interactions with others and the world we share play an important role in determining what we mean by our expressions and thus in our having a language. This in turn indicates that features of our environment have themselves a role to play. I do not wish to comment on these remarks, however, as they are very sketchy and the subject of much debate. Rather, I want to turn to Davidson, who has forcefully argued for the necessity of people interacting with one another and the world they share if they are to have a language at all. Davidson’s account is prompted by considerations similar to Wittgenstein’s; indeed, Davidson can be understood as motivating the sceptical problem anew and then going on to address it in a more robust way.

5 Davidson’s Sceptical Problem

Throughout his life, Davidson advocated what he called perceptual externalism, according to which part of the answer to the question what it is for words to mean what they do is that their meaning is based, in the first instance, on connections between speakers’ utterances and the objects and events in their environment that typically cause them to produce those utterances. This view

---

16 Cf. Wittgenstein’s suggestion that, even though I can invent a game that is never played by anyone, someone cannot invent such a game, if “mankind has never played any games” (Wittgenstein 1953, §204).
17 See also Wittgenstein (1953, §§189, 190 and 197). Admittedly, Wittgenstein does not have a whole lot positive to say.
19 See, e.g., Davidson (1991a, 200–1).
was initially a consequence of his reflecting on radical interpretation, the interpretation from scratch of an alien speaker’s utterances. Davidson thought that reflecting on radical interpretation would provide an answer to the question what it is for words to mean what they do and that it would do so in a non-question-begging way, since the idea is to ascertain what it takes to understand another person without the help of translators or bilingual dictionaries. Given the assumption that what people mean by their words is necessarily public,²⁰ Davidson thought that whatever we learn from how we could interpret an alien speaker has consequences regarding the answer to the question what it is that determines or constitutes what we mean by our words. Since, to put it very generally, regular connections between speakers’ utterances and features of their environment have to be attended to in order to interpret their utterances, it follows, again to put it very generally, that these features are essential to determining the meaning of their utterances. But there is a problem Davidson did not consider when reflecting on radical interpretation and which became the focus of his reflections on triangulation, the idea of two creatures responding simultaneously to features of their environment and to one another’s responses.

Here is the problem: if the typical causes of speakers’ basic utterances (such as “there is a table”, or “there goes a rabbit”) determine, in part, what they mean by these utterances, the question then is, what are these typical causes? Davidson argues that this is not a question that could be answered if we think of a speaker, or even a group of speakers, merely producing utterances in response to their environment. We, and they, for that matter, if that is all they are doing, could not say what the causes of their utterances are and so what determine their meaning, because for any given utterance what caused it is indeterminate or ambiguous. And this is true no matter how many utterances of the same type are produced and how regular the connections between utterances and their external causes seem to be. Indeed, this is true no matter how many speakers are producing the utterances. According to Davidson, considered in themselves, the causes of our utterances are “doubly indeterminate: with respect to width and with respect to distance. The first ambiguity concerns how much [what ‘part or aspect’] of the total cause of [an utterance] … is relevant to [meaning] … The second problem has to do with the ambiguity of the relevant stimulus, whether it is proximal (at the skin, say) or distal” (Davidson 1999, 129–30).

The distance problem is specifically Davidson’s and, according to him, it cannot be solved by a solitary person, a person who has always been socially isolated. For, he contends, a solitary person is incapable of determining whether the cause of her utterance is located in the world around her or at the surface of her skin or, if in the world around, whether it is one of the distal causes we

²⁰ See, e.g., Davidson (1979, 235) and Davidson (1990b, 62).
would normally consider as typical or something farther away or, for that matter, closer to her.\textsuperscript{21} The width problem is akin to the very first problem we found with Wittgenstein. It is the problem of determining what aspect of the distal cause is to be the relevant one, whether it is, say, the surface of an object, or its shape, or its colour, or, as Davidson suggests, it is the problem of determining how big a chunk of the world is the relevant distal cause; is it, say, the table or the set of the table and the chairs around it? And it is a problem that may subsist even if the other one is solved. Indeed, according to Davidson, the distance problem is solved, at least in part, by having at least two persons responding simultaneously to their environment and to each other, that is, by having at least two persons engaging in “primitive” triangulation. This is the kind of triangulation that even non-linguistic animals can engage in – Davidson gives the example of two lionesses trying to catch a gazelle and coordinating their behaviour by watching each other and the gazelle and reacting to each other’s reactions (Davidson 2001b, 7). Primitive triangulation, Davidson maintains, helps to make progress in isolating the cause, at least in isolating a distal cause, for it is the common cause of the two interlocutors’ responses, whatever is located at the intersection of the two lines one might draw from the participants in the triangular interaction. Primitive triangulation, however, “does not require intensional attitudes” (Davidson 2001c, 292). Though it now makes sense to describe interlocutors as reacting to distal, rather than proximal, causes, the relevant aspects or parts of the causes have not yet been isolated, let alone conceived of as the causes they are by the interlocutors.

I have my doubts about the need to triangulate in order to solve the distance problem.\textsuperscript{22} Either way, however, even if the distance problem is solved once people’s responses are shared, the width or aspect problem is left standing. For, to repeat, the mere fact that responses are shared does not make them responses to one rather than another aspect of the environment. This is something Davidson himself appreciates, as he writes: “The difficulty [of saying what is being meant] remains no matter how many people follow the same course of action” (Davidson 2001b, 3). And note that the difficulty of saying what is being meant is that of saying which cause is the same as which, which response is correct and which is incorrect. The problem might be expressed as I earlier expressed Wittgenstein’s: if the cause of one’s utterance can always be taken or interpreted in some way or another, then it can always be taken or interpreted in such a way that, no matter what response someone gives, it is correct, or incorrect, as the case may be, so that no conditions of correctness govern her utterances and hence these have no meaning. Thus there is no way that appealing merely to regular connections between people’s responses and

\textsuperscript{21} See, e.g., Davidson (1992, 118) and Davidson (2001b, 4).

\textsuperscript{22} For more on this, see Myers and Verheggen (2016, Chapter 1, Section 2.1).
features of their environment that cause these responses can explain what determine the conditions under which expressions are correctly applied and so what determine their meaning.

Clearly the problem Davidson emphasizes here is closely connected to the indeterminacy problem we recognized earlier with Wittgenstein. Clearly, if nothing more could be said about how the causes are identified as typical, we would be faced with the paradox. But Davidson cannot be content simply to reject the problem as being based on an incoherent assumption, the assumption that meaning is determined on the basis of mere connections between utterances and their typical causes, even though he would denounce this assumption just as much as Wittgenstein does. In fact, Davidson precisely does not make this assumption, as he is always careful to say the history of causal interactions only “partly” determine the meaning of our utterances. The difference between Wittgenstein and Davidson is the latter’s conviction that perceptual externalism is true – the typical causes of our utterances are what, in the first instance, determine, in part, the meaning of our utterances. Something after all has to do the work – meanings cannot come out of nowhere, nor are they ready-made. Even if we think of meaningful words as words being used meaningfully, meaningful uses do not occur in a vacuum. Moreover, as mentioned above, Wittgenstein would presumably agree that the world around us plays a crucial role, given his insistence on ostensive teaching in particular. But how can perceptual externalism be true? How are typical causes fixed? How do features of the world around us come to be seen as contributing to providing (at least some of) our words with meaning? Triangulation, and the kind of social externalism that flows from it, is Davidson’s answer to that question and thus, we might say, Davidson’s treatment of, i.e., full answer to, the paradox.

Before I turn to this, which I regard as Davidson’s improvement on Wittgenstein, let me emphasize that Wittgenstein, on the other hand, has much to contribute to strengthen Davidson’s externalism. It might be argued that the difficulty with fixing the causes of people’s utterances, and so with fixing meanings, indicates that semantic externalism cannot be true and that we should opt for internalism instead. For then, instead of appealing to external causes, we could appeal to features of the mind that are independent of the world surrounding it, such as, to cite those Wittgenstein discusses, mental pictures or abstract entities grasped by the mind. However, as demonstrated at length by Wittgenstein, the internalist is in the exact same kind of predicament as the externalist. The predicament for the internalist is really a version of the aspect problem. As we saw earlier, mental items are no better off than the external objects and events causing our utterances, at least as long as these are considered by themselves. (Needless to say, if these mental items are already meaningful, there is no indeterminacy problem; but
there is then the question what constituted their meaning what they do in the
first place.)

6 Davidson’s Solution to the Sceptical Problem

The question before us is, how are the typical causes of people’s utterances, and hence their meanings, fixed? Of course, if people have a language, they can use it to fix the cause of a particular utterance, just as they can define a colour-word by pointing at the relevant colour. But in the absence of any language, how can this be done? Davidson’s reply, strictly speaking, is that there is no answer to that question, that is, there is no explicating someone’s first entry into language – this is the point of the non-reductionism. However, this does not entail that we cannot spell out some conditions that made fixing causes possible. What needs to be stressed, to begin with, is that fixing causes, and hence meanings, is tantamount to fixing the conditions of correctness that govern the application of expressions. It is to distinguish between correct and incorrect responses one gives to the environment. It looks as if, though, since this distinction cannot emerge magically from mere connections between responses and their causes, those responding must themselves draw the distinction. They must somehow determine which causes are the same as which, which responses are correct and which incorrect. But this seems to lead to non-reductionism with a circular vengeance. For they cannot settle on the conditions of correctness that govern the applications of their expressions without using meaningful expressions which, in order to be meaningful, must have conditions of correct application. Is this not admitting defeat and forcing quietism upon us? This might indeed explain the reading of Wittgenstein’s positive remarks according to which all we can do to illuminate the notion of meaning is describe the linguistic practices people are trained and engaged in when they have a language. All we can say is that the conditions of correctness governing the applications of their words are determined by those at play in their linguistic community. But Davidson has something to add to this.

Once it is recognized that language users themselves need to contribute to establishing the conditions of correct application of their words, it can be asked what puts them in a position to draw the distinction between correct and incorrect applications. Here it is important to stress that, even though it is speakers who draw the distinction, that is, it is speakers who first determine, at least in part, what the conditions of correctness governing the applications of their words are, whether these conditions are met or not is, once they are

23 Note that the rejection of internalism helps to vindicate Davidson’s assumption that meaning is necessarily public. I spell this out in Myers and Verheggen (2016, Chapter 1).

24 See, e.g., Davidson (1999, 127) and Davidson (2001c, 293).
established, an objective matter. Once the distinction between what seems to be the same cause or the correct response and what is in fact the same cause or the correct response is drawn, and thus the meaning of the response is fixed, whether a given application is correct or not does not depend on the speakers’ say on the matter. If whatever application seemed to be correct could in fact be correct, then no conditions of correct application could be established – we would in effect be back to the paradox. But then our question becomes what puts people in a position to draw the distinction between what seems to be the case and what is in fact the case; what does put them in a position to have, for short, the concept of objectivity? Davidson’s answer is that only those who have triangulated linguistically can have it, that is, only those who have given linguistic responses simultaneously to one another’s linguistic responses and to features of their shared environment. This claim must of course not be understood temporally. It is not as if first people somehow get the idea of there being a distinction between correct and incorrect responses to their environment and then use the distinction to fix the causes of their responses and endow them with meaning. The correctness or incorrectness of responses could not be recognized unless the responses themselves were subject to conditions of correctness. Rather, the claim is that it is only of people who have triangulated linguistically that we can make sense of the idea that they have the concept of objectivity and that they thus can draw the relevant distinction. Why is this?

Interestingly, Davidson often invokes Wittgenstein to answer this question, though there is again much debate as to whether Wittgenstein endorsed the claim that only a socially situated individual could distinguish between following a rule and thinking that one is following a rule. Davidson’s claim, in a nutshell, is that only someone who has triangulated linguistically could do so because she has been involved in disagreements as to what the causes of her and her interlocutor’s utterances are, and she has had to settle on resolutions of their disagreements. The distinction has thereby been “forced” upon her (Davidson 1975, 170). Let me unpack these claims.

To begin with, it must be stressed that it is linguistic triangulation that is needed here, that is, a special kind of interpersonal communication. It is the kind of triangulation the radical interpreter engages in since, in order to interpret an alien speaker, she has to figure out what the causes of her inter-prettee’s basic utterances are. It is the kind of triangulation we engage in when ostension is needed to figure out the cause of a speaker’s utterance. And it is the kind of triangulation the child has to go through to be thought of as having acquired a first language. Though Davidson insists that there is no explaining

---

26 See Davidson (2001c, 294). Note that this indicates the continuity of Davidson’s thought from radical interpretation to triangulation. I defend this continuity claim in Myers and Verheggen (2016), especially Chapter 3.
the mechanisms of one’s first entry into language, that is, there is no explaining how the child bridges the gap between the extensional realm and the intensional realm, he also finds it useful to describe some of the steps a child has to go through in order to acquire a language. Some necessary steps can be described, but no non-circular sufficient ones can. Thus, let us work through his example of a child learning the meaning of ‘table’.

Recall first that, according to Davidson, if we consider the child in isolation, the stimulus that causes her to mouth ‘table’ is indeterminate in the distance way: it could be said to be the stimulation of her nerve endings, just as much as a table in her vicinity, among many other options. If, however, the child triangulates with her teacher, then it becomes possible to start answering the question what the cause of her uttering ‘table’ is. For

Involved in our picture there are … three similarity patterns. The child finds tables similar; we find tables similar; and we find the child’s responses in the presence of tables similar. It now makes sense for us to call responses of the child responses to tables. Given these three patterns of response we can assign a location to the stimuli that elicit the child’s responses. The relevant stimuli are the objects or events we naturally find similar (tables) which are correlated with responses of the child we find similar. It is a form of triangulation … Where the lines from child to table and us to table converge, ‘the’ stimulus is located. (Davidson 1992, 119)

This, however, is only the primitive kind of triangulation, which, according to Davidson, makes it possible to solve the distance problem, that is, to “give a meaning to the idea that the stimulus has an objective location in a common place” (Davidson 1992, 119). But engaging in primitive triangulation is not sufficient for anyone to have the idea. For this to happen, “[t]he interaction must be available to the interacting creatures” (Davidson 1992, 120).

To have the concept of a table … is to recognize the existence of a triangle, one apex of which is oneself, the second apex another creature similar to oneself, and the third a … table … located in a space thus made common.

The only way of knowing that the second apex of the triangle – the second creature or person – is reacting to the same object as oneself is to know that the other person must also know that the first person constitutes an apex of the same triangle another apex of which the second person occupies. For two people to know of each other that they are so related, that their thoughts are so related, requires that they be in communication. Each of them must speak to the other and be understood by the other. They don’t … have to mean the same thing by the same words, but they must each be an interpreter of the other. (Davidson 1992, 121)

What does it take to be the interpreter of another?

It is useful to answer this question by focusing on radical interpretation, which makes conspicuous elements that go into the constitution of meaning that usually need not be present for communication between linguistic peers. The radical interpreter’s goal is to construct for a speaker a theory of meaning
that would enable her to understand any sentence the speaker may utter. Davidson’s idea was that a Tarski-style theory of truth would do the trick, that is, a theory that would give the truth-conditions of any sentence the speaker may utter. A properly constructed theory could do this in such a way that the truth-conditions obtained would indeed be meaning-giving. The details do not matter for our purposes.\footnote{See Davidson (1973, 1974, 1990a), among others.} What is important is to see how, in broad outline, that theory could be constructed, for it should also shed light on the linguistic triangulating process and so on how causes and conditions of correct application get to be fixed. Indeed, as Davidson himself remarked, reflecting on radical interpretation is bound to “throw indirect light on what makes possible a first entry into language” (Davidson 1991b, 210).

Recall that the radical interpreter is supposed to interpret a speaker’s utterances without assuming any knowledge of their meaning nor any detailed knowledge of the speaker’s propositional attitudes (since knowing these in any detail would require understanding her utterances). Thus the only way for the interpreter initially to proceed is by assuming that many of the utterances the speaker holds true express by and large the beliefs the interpreter also has in the circumstances in which they are held true. She will then match utterances of the speaker held true in specific circumstances with utterances of her own that express beliefs she herself has in the circumstances. (Davidson thinks that held true sentences can be detected without knowledge of the beliefs they express.) For instance, she will match a speaker’s holding true an utterance of ‘Gavagai’ when a rabbit is in the vicinity of the speaker with her belief that there is a rabbit nearby. Eventually, she will have accumulated enough evidence of this sort to start giving the truth-conditions of the sentences the speaker holds true. And from these so-called T-sentences, she will be able to extract primitive expressions – expressions whose meaning does not have to be explained by reference to other ones – which she will then match with primitive expressions of her own. Knowledge of these and of their modes of combination in turn will enable her to say how they contribute to the truth-conditions of the sentences in which they occur and thus to understand any sentence the speaker may utter.

Leaving the details aside, it is important to emphasize two related aspects of the enterprise. First, some extended history of causal interactions between the speaker and the interpreter and with their shared environment must be available. This is in part because, second, primitive expressions must occur in different utterances produced in different circumstances. This is what will enable the interpreter to nail down the aspects of the causes of the speaker’s utterances and hence their meaning. Thus on the first occasion someone utters ‘There is a table’, there is no reason to interpret her as talking about tables rather than brownness, or flatness, or stylishness, or table and chairs, etc. It is
only when these words, and many others, have been uttered in many different circumstances, and in many different combinations, that we may be able to nail down tables as the aspect or part of the world shared by speaker and interpreter that contributes to determining the meaning of the speaker’s ‘table’. But for that shared aspect so to contribute, it must be recognized as the shared aspect. And this is something that can be done if speakers can distinguish between being the same and seeming to be the same, which again can be achieved if they have had the opportunity to disagree on some matters and to settle the disagreements, even if by agreeing to disagree. As Davidson writes, “the concept of error … appears when there is a divergence in normally similar reactions” (Davidson 2001b, 12). If the responses of triangulating interlocutors diverge, the distinction is “forced” upon them because they have no choice but to negotiate the resolution of the divergence. For a solitary person it would be entirely up to her how to settle the divergence. She would have the only say, and thus the last say, on the matter, which is to say that she can have no say. Moreover, if the divergence of responses is to put someone in a position to have the concept of objectivity, the matter must be settled linguistically. Someone who interacted with a non-linguistic creature would be no better off than the solitary person, having again the only say on the matter and thus no say at all.

Needless to say, there have been numerous objections to Davidson’s triangulation argument. I do not intend to address them here, with one exception, however. For this is the objection that has often been made to those (or by those) who, like Davidson, interpret Wittgenstein as making the claim that a solitary person cannot distinguish between what seems correct to her and what is correct. The objection, in a nutshell, is that triangulating speakers are in the exact same position: what seems correct to them is correct. Whatever they say goes. But reflecting on the triangulating process precisely affords Davidson an answer to that objection. What needs to be stressed is that this process, which includes both agreement and disagreement, and negotiation of disagreements, is what enables speakers to fix the causes, and hence the meanings, of their responses to features of their environment – the fixing of conditions of correct application and the very ability to distinguish between correct and incorrect applications go hand in hand. Now, to be sure, once the two disputants have settled their dispute, once they have agreed or agreed to disagree, that is, once they have fixed the meanings of their responses, whatever they say the conditions of correct application, and so the meanings of her words, are does go. But this does not entail that whatever application they say is correct on a given occasion is the correct application. As Wittgenstein suggested, the fact that they “agree in their language” does not entail that the

28 I have done so extensively in Myers and Verheggen (2016, Chapter 1, Section 3).
truth-value of what they say also depends on their agreeing on this truth-value. In short, in order to appreciate the difference between the solitary person and the triangulating person, it is important to focus not on the end product of the triangulating process, but on the process itself which necessarily involves a back and forth of the sort that the solitary person cannot engage in.

Let us at last take stock. What is it, according to Davidson, for words to mean what they do? It is to have been used in linguistic triangulating situations. The meanings of words are fixed, in the first instance, through regular connections between triangulating interlocutors’ utterances and features of the world around them that have been understood as the particular features they are by both interlocutors, that is, features which both interlocutors have agreed are the features relevant to determining the meanings of the speaker’s words. For, as Davidson observes, they do not have to mean the same thing by the same words, though, given the significance of their surroundings in determining what their words mean, many of their meanings are bound to overlap. Thus, it is not just that, as Wittgenstein put it, “it is not possible that there should have been only one occasion on which only one person followed a rule” (Wittgenstein 1953, §198). But it is not possible that there should have been only one occasion on which more than one person followed a rule, and it is not possible that there should be only one person who followed a rule, or several rules, for that matter, albeit on several occasions. Of course, as noted before, not every word has to be multiply applied and to be applied by multiple persons in order to be meaningful – the meaning of some words can be explained by means of other words, and someone could in that way introduce a new word and never use it. It is also not true for every word that refers to some feature of the world around us that one must have triangulated on that feature in order to mean something by it. But “there must be a causal history of that person that traces back, directly or indirectly, to triangular experiences” (Davidson 2001c, 293).

To end, though I have deliberately avoided discussing the rule-following paradox as developed by Kripke, I cannot resist sketching what Davidson’s answer to it would be like. Consider then Kripke’s sceptic asking what makes me so sure that by ‘table’ I mean table and not tabair, “where a ‘tabair’ is anything that is a table not found at the base of the Eiffel tower, or a chair found there” (Kripke 1982, 19). Both Wittgenstein and Davidson have shown that there is nothing I can appeal to in order to justify my claim that it is table, and not tabair, that I mean, short of saying that it is what I mean by ‘table’. Anything else, including past uses and dispositions non-semantically described, would be compatible with my meaning tabair instead of table. Kripke himself took the appeal to states of meaning to be a “desperate” move (Kripke 1982, 51). But there is nothing “mysterious” about them once the story of their coming into

---

30 See Wittgenstein (1953, §241).
being is fleshed out in the way Davidson recommends. My meaning table rather than tabair by ‘table’ has been determined by my triangulating linguistically on tables and my having been understood as meaning table. I have asked people to sit at my table, to admire my table, to clean the table, to set the table, and I have done all of this in the presence of tables. I have not asked people to have dinner at my chair, though I also have used ‘chair’ in many triangular situations and have been understood as meaning chair by ‘chair’. In short, unless I wish to puzzle my audience, there is no reason for me to apply ‘table’ to a chair, even if it is found at the base of the Eiffel Tower. (This is not to say that the triangulating evidence we have at our disposal is always sufficient to interpret someone correctly. But then the problem becomes an induction problem, as Davidson once pointed out.\footnote{Strangely, Davidson himself saw it exclusively as an instance of the problem of induction (Davidson 1992, 111). I discuss this further in Myers and Verheggen (2016, Chapter 3).} It is no longer the problem we have been concerned with, which was that of saying what makes it possible for us to mean anything by our expressions, given that mere correlations between words and extra-linguistic items will not do.)

7 Conclusion

I have argued that Wittgenstein and Davidson motivate the sceptical problem embodied in the paradox in similar ways. They also address it in a similar way in so far as they share the diagnosis of the sceptical problem: it rests on an incoherent assumption, the assumption that meaning rests on mere associations between words and extra-linguistic entities, conceived of independently of how we apprehend them. As a result, they both further argue that, in so far as the nature of meaning can be illuminated, it is only from within semantic contexts, by reflecting on words in meaningful use. For Wittgenstein, training into existing linguistic practices and ostensive training seem to play crucial roles in determining what people mean by their words and in making it possible for them to mean anything to begin with. But it remains unclear just what Wittgenstein thinks this commits him to. Are his scattered positive remarks to be taken to imply that only descriptions of our linguistic practices can illuminate meaning, thereby implying philosophical quietism? Or are they to be taken to imply a philosophically more robust view, according to which only by meaning by and large what members of one’s community mean can one mean anything by one’s words? What is clear is that Wittgenstein’s remarks are perfectly compatible with Davidson’s views, according to which a constructive, non-reductionist though it may be, treatment of the sceptical problem is available. Indeed, I take Davidson’s answer to be an improvement, not only on quietism, but also on the communitarian view that leaves it
mysterious in the end what makes linguistic practices possible in the first place.32

References

Kotatko, Petr, Peter Pagin, and Gabriel Segal, eds. 2001. Interpreting Davidson. Stanford, Calif.: CSLI.

32 I am grateful to Robert Myers and Olivia Sultanescu for their comments on an early draft of this chapter and for much helpful discussion of the issues. I would also like to take the opportunity to thank Barry Stroud whose teaching and writings on Wittgenstein have deeply influenced me.


I think there is a deep and distinctive accord between Donald Davidson’s conception of meaning and understanding and the treatment of those same notions in the later philosophy of Wittgenstein. Both philosophers presume that there can be nothing more to an expression’s meaning what it does than its being used in certain ways by speakers, and nothing more to speakers’ understanding an expression than their responding or being ready to respond to it in certain distinctive ways in interacting with others and with the shared world around them. This can seem beyond question. It is only because sounds or marks are used in some way or other that they mean what they do; no sounds or marks meant anything before human beings began to speak. But that very general connection between use and meaning is not the distinctive accord I see between Davidson and Wittgenstein.

Whether an appeal to the use of an expression can account for its meaning what it does depends on what the use of an expression is understood to be, or how that use is to be described. It is on this question that I see a promising accord between Davidson and Wittgenstein. Not every conception of meaning as use is equally promising.

Quine, for instance, also agreed that there is no more to the meaning of an expression than the use speakers and hearers make of it. For him use was to be described in terms of the conditions under which speakers would utter an expression or would assent to or dissent from its utterance by others. And for Quine those conditions could be specified by any description that is true if and only if speakers would in fact respond in those ways. This is vulnerable to the so-called “indeterminacy of translation”, according to which any expression uttered when and only when a rabbit is present means the same as the English words ‘rabbit’, ‘undetached rabbit part’, ‘rabbit-stage’, ‘rabbithood’, or any other expressions that differ in meaning from one another but are true of every circumstance in which a rabbit is present. This leaves what the expression in question means at best indeterminate. It does not account for differences of

1 Quine 1960, Chapter 2.
meaning we all recognize. This is not a consequence simply of the idea that meaning is determined by use; it is a result of a certain way of understanding what the “use” of an expression can legitimately be said to be.

It is perhaps not surprising that merely extensionally equivalent descriptions of the conditions for the appropriate use of a term do not suffice to fix what the term means. The same difficulty arises for conceptions of use understood as speakers’ dispositions to utter or to respond to a term under certain conditions, if those conditions could be specified in any extensionally equivalent terms. There is no such threat in Davidson’s or Wittgenstein’s conceptions of speakers’ use or understanding of expressions.

Both Davidson and Wittgenstein see language as a public social practice. Speakers’ meaning what they do by their words can therefore be understood and explained by starting from the undeniable fact that each of us is born into a linguistic culture in which we learn to speak by coming to participate in that culture and so acquire competences we did not have before. To describe in detail the linguistic competence we all acquire would be to describe the use and meaning of expressions in languages we understand.

In the early sections of *Philosophical Investigations* Wittgenstein explores the role of what he calls “demonstrative” or “ostensive” learning in the acquisition of language. That involves “ostensive teaching” in which a teacher draws a pupil’s attention to objects of a certain kind while uttering an expression whose use involves objects of that kind in some so-far-unspecified way. This requires more than the pupil’s simply noticing some similarity or other among the “ostended” objects. For one thing, there are too many similarities available; every object is similar to every other in many respects. If the objects the teacher points to when uttering the sound are both red and round, for instance, the question is which feature is relevant to what the expression means. Is the sound meant as a colour-word or a shape-word? Explicit instructions would be to no avail with a pupil who does not yet understand any words. By pointing, uttering, and gesturing, the teacher can eventually get the pupil to catch on to the similarity he is drawing his attention to. No one could learn to speak if he could not catch on in this way and so come to share judgements of relevant similarity with competent speakers.

That competence starts with what Wittgenstein calls an “associative connection” in the mind of the pupil between the sound uttered and objects of a particular kind. That “association” is a step in the right direction, but it alone does not give the pupil a determinate understanding of what the expression means. There must also be what Wittgenstein calls “instruction” or “training”. The pupil must learn how to use the expression – how to apply it to relevant items and to respond to it appropriately in communication with others. The same “associative connection” between expression and object, if combined with one kind of “training”, would result in a quite different
understanding of the term from that produced by a different “training”. The use a speaker learns to make of an expression is what determines what he means and understands by it.

To catch on to the “training” and to know how to use the new word correctly is to recognize what Wittgenstein calls the “post at which we station the word”, the “place in language, in grammar, we assign to the word”.\(^2\) That involves mastery of the structure of the sentences in which words of that kind appear, how it differs from the structure of other sentences, and under what conditions what is said to be so by such sentences is to be accepted or rejected.

Competent speakers of a language could observe a pupil who is learning that language and see that by this kind of “training” she has come to understand a certain expression as a colour-word rather than a shape-word. The pupil might still lack the word ‘colour’, so she could not yet say such a thing about herself, but she would show in her use of the new word that she understands it as a colour-word. If we find that what she means by the colour-word she has learned is red, we make use of our own understanding of the word, and of what it is for an object to be red, in saying what she means by the word. When she is fully competent she could say the same thing about herself: “By the word ‘red’ I mean red”. She then uses her recently acquired colour-word to report what she knows the word she mentions means. She says something she knows, and is now able to say, about the meaning of the word by having learned that language. All mature speakers of a language are in a position to say such things about the words of their language. By making use of words we understand, we can describe correctly the use and meaning of words that speakers we understand use and understand.

It is from just such a position that Wittgenstein describes the use and meaning of the words he is concerned with in his own investigations of what words mean or what speakers understand by them. The expressions he is concerned with are words of the languages we speak, and their use and meaning are described by using some of those very words. There is no restriction to describing the conditions of appropriate application of those words only “externally”, so to speak, in terms that presuppose or ascribe to the speaker no understanding of the meanings of any words. So there is no suggestion of behaviourism. This is part of the distinctive accord I find between Davidson and Wittgenstein. Each of them, from his own philosophical lights, describes speakers’ meaning and understanding of expressions from within their engaged, “internal”, and so intensionalist, understanding of the use of terms that we all employ in making sense of one another.

Davidson, like Wittgenstein, also stresses the importance of ostension, not just in learning language but also for what the words one learns mean. It is

\(^2\) Wittgenstein 1953, §29.
because the meanings of words learned ostensively are “tied to the sorts of objects and events ostended”, that “ostensive learning is crucial to the existence of objective thought and language”. That kind of learning involves what Davidson calls “triangulation”; a process of interaction between perceivers and objects in the world they perceive. The process in itself does not require language, nor is language always the result. Triangulation occurs whenever one creature sees and focuses on a certain object while also noticing another creature attending both to that same object and to the first creature’s attending to it. Two lionesses engage in triangulation when zeroing in on a prey while keeping each other in sight, for instance, or two baboons circling the same female very much aware of the tactics of the other. The point is that this is an essentially social process. No one can do it alone; it takes at least two agents, and it takes interaction between them and objects in their shared world.

Although triangulation can occur without language, Davidson holds that triangulation is “a necessary condition of thought and language”. What he calls triangulation is clearly present in the social interactions between teacher and pupil that Wittgenstein has in mind in his “ostensive learning”. In that situation triangulation serves to pick out the relevant similarities among the ostended objects that the pupil must catch on to. The teacher rewards what he finds to be similarities among the pupil’s responses to objects that the teacher regards as similar to one another in the relevant respect.

Thus the child, learning the word ‘table’, has already in effect noted that the teacher’s responses are similar (rewarding) when its own responses (mouthing ‘table’) are similar. The teacher on his part is training the child to make similar responses to what he (the teacher) perceives as similar stimuli. For this to work, it is clear that the innate similarity responses of child and teacher – what they naturally group together – must be much alike; otherwise the child will respond to what the teacher takes to be similar in ways the teacher does not find similar. A condition for being a speaker is that there must be others enough like oneself.

In the normal situation of a child learning a language from her elders, the pupil eventually comes to share the language with her teachers. But Davidson stresses that the success of a triangular “ostensive learning” procedure does not require that the speakers share a public vocabulary or common patterns of speech. The social conditions of language and thought can be recognized and appreciated even in the absence of a shared language. What the conditions of successful “ostensive learning” do imply, Davidson thinks, is that “there would be no saying what a speaker was talking or thinking about, no basis for claiming he could locate objects in an objective space and time, without interaction with a second person.”

Davidson, like Wittgenstein, also insists that the triangulation involved in “ostensive learning” of words is not sufficient for understanding what those words mean. A speaker’s having simply learned to utter a certain expression under certain conditions, or to assent to or dissent from an utterance under certain conditions, is not sufficient to determine what the speaker means by that expression, or that he means anything at all by it. Meaning and understanding make their appearance only because of what Davidson calls the “wide variety of social interactions” that “cluster around ostension”. Among those social interactions are the kinds of activities Wittgenstein speaks of as “training” or “instruction”. Speakers come to mean and understand something determinate by the expressions they are learning only when they have more or less mastered the grammatical structure of the language in which those expressions are used in sentences. Speakers’ mastery of that rich linguistic and cognitive structure is something that arises only gradually.

It is a fundamental idea of Davidson’s that the competence speakers acquire in meaning and understanding expressions of a language can be represented by a description of the compositional structure of the language they have mastered. That would be a description of something all competent speakers of the language know, or know how to do, by having learned the language. Davidson’s idea is to describe the linguistic structure they have mastered in what he calls a “truth-conditional theory of meaning” for that language. That is not a theory of something called “meanings” which are somehow attached to words. It is a description of the structural interrelations among all the sentences in the language with respect to their expressing something that is either true or false.

The idea is that, starting from specifications of the conditions under which certain sentences of relatively simple structure are true, a recursive account of the interrelations among the structures of all other sentences of the language would yield, for each sentence, a statement of the conditions under which that sentence is true. So for each sentence of the language there would be a “T-sentence” that states the conditions under which the sentence in question is true. Since each of those T-sentences is a consequence of the compositional structure of the language, they each could be said to state something that competent speakers know about the sentences they understand. This would explain how speakers can understand sentences they have never heard before. Their competence in their language would be embodied in the source of their knowledge of the truth of all those T-sentences: their grasp of the compositional structure of the language.

Development of a theory of meaning for a language along these lines might seem alien to the spirit of Wittgenstein’s later philosophy – although not, perhaps,

---

9 Davidson 2003, 694.
equally alien to the conception of language in the Tractatus. Crispin Wright, for instance, finds “a very fundamental tension” between Wittgenstein’s later conception of meaning and understanding and “thinking of language as available even to description by the sort of theory envisaged” by Davidson. But I think that is to misunderstand the point or function of a theory of meaning for a language as Davidson understands it. I will come back to this question.

The goal of Davidson’s theory of meaning for a language is a systematic description of a competent speaker’s knowledge of the structure of a language he understands. That knowledge is what Wittgenstein also has in mind in his metaphor of a speaker’s grasp of the “post” at which different kinds of expression are “stationed” in his language – the different “place, in grammar”, assigned to different kinds of expression. That grasp is something speakers acquire by learning the language.

In the same way, on Davidson’s view, it is by learning the language that competent speakers come to know the truth-conditions of the sentences they understand. In knowing those truth-conditions, they could be said to know the truth of the T-sentences that express those conditions. But knowing those T-sentences to be true is not of course speakers’ way into the language in the first place. T-sentences cannot convey the meanings of sentences to those who do not already understand them. You cannot know that the sentence ‘Snow is white’ is true if and only if snow is white, for instance, if you do not already understand the words ‘snow’ and ‘white’, and know what it is for something to be white. After all, those T-sentences that speakers come to know are contingent statements about certain linguistic expressions. The linguistic expressions themselves could have had different truth-conditions from those they actually have, or none at all. So the T-sentences speakers know to be true could have been false. As things are, the T-sentences of an actual language are not false, but that is because of the way the expressions they mention are in fact actually used. And that is what speakers who learn the language come to know.

We imagined earlier a beginning speaker who comes to learn what the word ‘red’ means by successful ostensive learning of the kind envisaged by Wittgenstein. She can report that by the word ‘red’ she means red, or that in applying the word ‘red’ to an object she means that the object is red. She expresses what she understands that word to mean by using the very word she understands within the very language in which the word means what it does.

The T-sentences generated by Davidson’s truth-conditional theory of meaning for a language do not themselves say what any expression means, but they also state the truth-conditions of the sentences of the language in a similarly engaged, “internal” way: they use the very words of the sentences they mention.

10 Wright 1981, 114.
with the meanings they have in that language. Even a theory of meaning for a different language states the truth-conditions of sentences of that other language by using (not simply mentioning) expressions in its own language in stating the conditions under which the mentioned foreign sentences are true. There is no other way of doing it.

The fundamental idea about the specification of meaning and understanding that Davidson and Wittgenstein share is that you cannot say what an expression means, or what someone understands by it, without using that very expression or others with the same meaning to say what the expression or the person means.

Being capable only of such “engaged, internal” specifications of meaning and understanding can be found dissatisfying. Michael Dummett, for instance, long insisted that a theory that gives the truth-conditions of sentences by using words of the very sentences whose truth-conditions are being specified is too “modest” to “satisfy the purpose for which, philosophically, we require a theory of meaning”.11 Dummett’s idea is that since a speaker’s employment of his language “rests on” his knowledge of it,12 possession of that knowledge or understanding must somehow be what explains a speaker’s successfully saying what he says and understanding the things he hears other speakers say. An adequate theory of meaning, for Dummett, must therefore be “a theory of understanding”.13

A “modest” theory of meaning for a language does concern itself in a certain way with speakers’ understanding; it gives what Dummett calls “a theoretical representation of a practical ability” that competent speakers have mastered.14 In that sense, a “modest” theory of meaning describes what speakers who understand the language know. But Dummett complains that it does not explain “what it is” for speakers to have that knowledge, or what their understanding their language in that way “consists in”.15 Dummett thinks a philosophical theory of meaning should explain how a speaker’s “general understanding of the language”16 is put to work in the speaker’s linguistic performances, or, as he sometimes puts it, it should explain how the “piece of internal equipment” – the understanding of the language possessed by the speaker – is “delivered” to him and “applied when the occasion for its application arises.”17 How does a speaker’s understanding the meanings of his words operate “to guide, prompt, or control the speaker’s utterances”?18 A “modest” Davidsonian theory that simply describes speakers’ competence does not explain that. For Dummett, only a more “full-blooded” “theory of meaning and understanding” can account for what he thinks we want to understand.

---

It is difficult to see exactly what a “full-blooded theory of meaning”, as Dummett understands it, is meant to do, and how it is expected to do it. It looks as if the theory, being “full-blooded” and not merely “modest”, would have to give an account of a speaker’s understanding of his language without presupposing any grasp on the speaker’s part of any of the concepts needed in understanding the language. The theory would have to explain “what it is” for a speaker to understand the expressions of his language in the ways he does without itself making use of those expressions in a way that credits the speaker with some prior understanding of them, or with possession of any of the concepts expressed by those words. It is not easy to understand how any theory could succeed in doing that. I will come back to this difficulty when I return to Wittgenstein.

For now it is enough to observe that Davidson does not think a theory of meaning, as he understands it, performs or is meant to perform the function Dummett complains it cannot fulfill. Davidson does not regard a theory of meaning for a language as a theory or device or “piece of internal equipment” that speakers consult in deciding what to say, or that they apply directly to utterances they hear to determine what those who utter them mean. Even when two speakers understand and speak the same language, as we say, Davidson denies that in communicating with one another they “operate on the basis of shared conventions, rules, or regularities” determining the meanings of their words. An “ability to operate in accord with a precise and specifiable set of syntactic and semantic rules” shared by different speakers is “neither necessary nor sufficient for successful linguistic communication”, Davidson holds.

Understanding what another person’s words mean is of course important in communication, but understanding what a person says requires more than knowing the meanings of the words she utters. Davidson illustrates the point with eccentric speakers like Sheridan’s Mrs. Malaprop. Mrs. Malaprop describes an insulting letter as “a nice derangement of epitaphs”, and a certain person as “as headstrong as an allegory on the banks of the Nile”. Direct application of a theory of meaning for English to Mrs. Malaprop’s utterances would have her speaking of epitaphs and allegories, but we all know what she means and says. Even when words are used normally and in accordance with the theory of meaning for the language, Davidson holds that the theory of meaning alone does not suffice for interpreting what a speaker says.

That is because saying something is doing something, and no theory of meaning for a language can tell you what a person is doing or will do on a particular occasion. That is something a hearer has to recognize, or sometimes figure out, even if he knows the speaker’s language. Understanding what others

say or mean is a matter of what Davidson calls “interpretation”. It involves making the best sense of what a person is intentionally doing in uttering the sounds he makes in the circumstances in which he makes them. That involves ascribing intentional attitudes of thought, meaning, and intention to the speaker. And to ascribe intentional attitudes to others the interpreter must himself possess the concepts he needs for ascribing any attitudes to others, and he must ascribe to others the concepts that those others must possess and understand in order to hold the intentional attitudes the interpreter ascribes to them.

Davidson starts from the idea that the point of language or speech is communication: “getting across to someone else what you have in mind by means of words that they interpret (understand) as you want them to”.21 This he sees as common to all verbal behaviour. Meaning, in the special sense in which we are interested when we talk of what an utterance literally means, gets its life from those situations in which someone intends (or assumes or expects) that his words will be understood in a certain way, and they are. In such cases we can say without hesitation: how he intended to be understood, and was understood, is what he, and his words, literally meant on that occasion.22

This is clearly not meant as a definition of meaning in terms of speaker’s intentions.23 There is no elimination of meaning or a reduction of it to something else. Having the intention that one’s audience understand one in a certain way involves having the idea of an expression’s being understood in one way rather than another, and so “assumes the notion of meaning”.24 That is not surprising if the goal of interpretation is to recognize the intentional attitudes out of which a speaker says and does what he does. Davidson argues that interpreters must find, in general, that those speakers they come to understand share with them many concepts, and so share with them many judgments about what is so in the world. That is a condition of successful interpretation. It is the source of what Davidson calls his “social externalism”: what we all think about and speak about and know are the very objects and people and states of affairs that surround us all.

Wittgenstein too makes sense of what people mean and say and think only from “within” the understanding we all have of people as active, speaking agents with intentional attitudes. A speaker who has caught on to her ostensive teaching and training understands the expressions she has learned. In her future use of those expressions, she goes on in the same way – in accordance with the relevant similarities – that she learned in her training. We recognize that she understands those expressions because we too understand them in the same

21 Davidson 2005b, 120. 22 Davidson 2005b, 120–121. 23 As H. P. Grice, e.g., envisaged in Grice 1957. 24 Davidson 2005b, 120.
way, and we recognize the similarities in terms of which she proceeds as essential to those expressions’ meaning what they do.

Wittgenstein addresses repeatedly and from different directions the question of understanding or knowing the meaning of expressions. He shows over and over why understanding cannot be simply the presence before the mind of an image or picture or set of instructions that the speaker attends to or somehow draws on in applying the expression or responding to its use. Not because there are no images or pictures “present to the mind”, but because any object – anything that could be attended to or consulted for guidance in using or responding to a certain expression – would itself have to be understood in a certain way by the speaker, or would have to mean something to her, in order to be any help to her in understanding the expression in question. To take seriously the idea that something present to the mind either is, or is what produces, a person’s understanding an expression in a particular way, is to be led into a regress. A person’s knowing the meaning of an expression cannot be understood, in general, as his possessing something – some object or instruction – that directs or guides or instructs him to respond to a situation or a remark in a certain way.

The threat of regress generated by an appeal to objects in the mind to account for meaning and understanding is a constant theme throughout the first half of Philosophical Investigations. It seems to me conclusive as an objection to the appeal to objects in the mind in accounting for understanding. What is needed in the place of objects in the mind is the idea that in coming to know a language and understand its expressions we acquire a certain ability or capacity. Such a capacity is not merely the presence or availability of certain expressions or other items. It is a general competence in using and responding to expressions in appropriate ways. In trying to understand such a capacity, Wittgenstein says it “dawned on him only later”\(^\text{25}\) that:

in philosophy we often compare the use of words with games, calculi with fixed rules, but cannot say that someone who is using language must be playing such a game.\(^\text{26}\)

To conclude in the face of this that “our languages only approximate to such calculi” is to stand, as Wittgenstein says, “on the very brink of a misunderstanding”.\(^\text{27}\) The heart of the first part of Philosophical Investigations is devoted to identifying and exposing the source of that misunderstanding. That source will have been revealed, Wittgenstein says:

only when one has attained greater clarity about the concepts of understanding, meaning something, and thinking. For it will then also become clear what may mislead us (and

\(^{25}\) Later than what? After the Tractatus? Or perhaps after discussions with Ramsey as referred to in Wittgenstein 1953, §81?

\(^{26}\) Wittgenstein 1953, §81.

\(^{27}\) Wittgenstein 1953, §81.
did mislead me) into thinking that if anyone utters a sentence and \textit{means} or \textit{understands} it, he is thereby operating a calculus according to definite rules.\footnote{Wittgenstein 1953, §81.}

In his discussions of understanding and meaning, Wittgenstein was fond of examples involving a rule for the development of an arithmetical series. To have caught on to the rule of such a series is to know “how to go on in the same way”, just as in the ostensive learning of words speakers acquire a capacity to “go on in the same way” in applying a word correctly to similar objects or situations beyond those they have encountered earlier. To best expose the potential misunderstanding of how this works, Wittgenstein takes as an example a strict arithmetical rule whose correct application at each point is fully determinate. He illustrates how, even in that case, a speaker who understands the rule, and understands the words that express that rule, and exhibits his understanding of them in applying the rule to particular cases, is not to be understood as operating a calculus according to definite rules.

Someone who is given the number 2 and the instruction to “Add two each time”, and understands what she has been told, knows how to go on. In grasping the rule, she could be said to know what number appears at each place in that series, even if she herself continues the series only up to 10. Her understanding far outstrips any actual performance she makes, however long it might be. But her application of the rule at each particular point could be said to be drawn from, or to issue from, her understanding of the general rule. It is \textit{because} she understands the general rule in the way she does that she knows to put ‘10’ right after ‘8’. The correctness of that step, and of each of the other steps she makes, is felt to be somehow “contained” in her understanding of the general rule, and needs only to be “extracted” from it.

It is when we ask exactly how someone who understands the general rule “extracts” from that understanding a particular application of the rule that we stand on the brink of the misunderstanding Wittgenstein warns against. A person who understands the rule knows that 10 is the number right after 8 in that series, and we feel there must be some way the person knows that. We are tempted to think there must be some operation by which she knowingly makes the transition from the general rule she understands to the particular application of it: some other rule or principle she has mastered that explains how her understanding the general rule leads to her knowing in particular that 10 comes right after 8.

This is the source of the question raised by the puzzled speaker in \textit{Philosophical Investigations} §198: “How can a rule teach me what I have to do at \textit{this} point?” If the person understands the general instruction ‘Add two each time’, and she has just written down ‘8’, she wants to know how the rule
shows her or tells her what to do next. If, as Wittgenstein warns, we have been ‘misled’ into thinking that “if anyone utters a sentence and means or understands it, he is thereby operating a calculus according to definite rules”, we will try to find the “calculus” the person is operating, the “definite rules” she grasps that take her from her understanding of the general rule and the number ‘8’ she has just written down to her now writing ‘10’. What rules or procedure or instructions direct her from the one to the other?

This question could be what lies behind Michael Dummett’s demands on a “philosophical theory of meaning”. He thinks any such theory must be “full-blooded”, as opposed to Davidson’s merely “modest” description only of what speakers know. According to Dummett the theory should explain how a speaker’s knowledge of what his words mean is “delivered” to him for application to particular cases, or how his understanding operates “to guide, prompt, or control” his performances. That is what the puzzled speaker in Philosophical Investigations §198 appears to be asking for. But we have seen that a “full-blooded” theory must apparently explain that procedure or guidance or instruction without itself presupposing any prior understanding on the part of the speaker of any of the expressions involved in the “delivery” of that guidance or instruction. The puzzled speaker who asks how the rule can teach him what to do at this point elaborates his quandary: “After all, [he says] whatever I do can, on some interpretation, be made compatible with the rule… So is whatever I do compatible with the rule?”

Here we have reached the fatal misunderstanding.

This speaker appears to seek just the kind of instruction or guidance that Dummett’s requirement on a “full-blooded” theory of meaning for a language is meant to provide: an explanation of how a speaker’s general understanding of his language is put to work by him in the particular applications he makes of it. This is what Dummett found missing from Davidson’s merely “modest” theory of meaning for a language. Much earlier, in his 1959 article “Wittgenstein’s Philosophy of Mathematics”, he also had found no such explanation in Wittgenstein’s accounts of inference and proof. In that paper, Dummett took what he referred to as the “considerations about rules presented in the Investigations and elsewhere” to carry the implication that even if the general rules of inference involved in a proof:

had been explicitly formulated at the start, and we had given our assent to them, our doing so would not in itself constitute recognition of each transition as a correct application of the rules… Hence at each step we are free to accept or reject the proof; there is nothing in our formulation of the axioms and of the rules of inference, and nothing in our minds when we accepted these before the proof was given, which of itself shows whether we shall accept the proof or not; and hence there is nothing which forces
us to accept the proof. If we accept the proof… we are making a new decision, and not merely making explicit a decision we had already made explicitly.³⁰

On these grounds Dummett attributed to Wittgenstein what he called a “full-blooded conventionalism” according to which “the logical necessity of any statement is always the direct expression of a linguistic convention. That a given statement is necessary consists always in our having expressly decided to treat that very statement as unassailable”. As Dummett observes, “this account is very difficult to accept”.³¹

The puzzled speaker in Philosophical Investigations §198 apparently feels a need for what Dummett thinks a “full-blooded” theory of meaning should provide. He appears to be looking for some particular way of understanding the expression of a rule he already accepts – ‘Add two each time’ – that will teach him or show him or in some way determine what application he is to make of that rule “at this point”.

Wittgenstein observes in Philosophical Investigations §201 that if an interpretation of the expression is needed, and if each “interpretation” offered is itself another expression and must be understood in turn, then still another “interpretation” will be needed to give the meaning of that expression, and so on down a never-ending regress. But people do understand expressions of general rules, and we all can see that they do and that they know how to apply rules to particular cases. So this quandary must be a result of a misunderstanding of how speakers do in fact understand the expressions of their language and know how to apply them. And there must be misunderstanding of how speakers reach conclusions that they know follow of necessity from something they understand. Wittgenstein says the threat of this regress of “interpretations” shows “that there 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’”.³² To say it is “exhibited” in what we call “following a rule” and “going against it” is to say that when we understand a certain rule we can usually recognize when our fellow speakers are following it. We can see that a person has the general competence involved in understanding the rule and exhibits it in applying the rule correctly. We can recognize the correctness of what the person does because we too understand the rule and the words that express it. We can see that someone understands the words ‘Add two each time’ and that he follows that rule correctly and so does what we call “following the rule ‘Add two each time’”, just as we can see that someone understands the word ‘red’ and applies it correctly to a red apple right before us. That is what we call “understanding and saying truly that a thing is red”.

Saul Kripke’s “skeptical paradox”, which is developed from materials he finds in Wittgenstein, leads to the conclusion that there is no such thing as a person’s using an expression with one determinate meaning rather than any other meaning compatible with all the applications he has made of the expression so far.33 The challenge is to find some “fact about a speaker” that could “constitute” his meaning the expression ‘+’ in one way rather than another. No objects or instructions in the mind, no current or past mental states or experiences, appear to suffice, since their mere presence does not show or imply that the speaker actually means plus by that expression. And adding further such states or expressions would lead only to regress.

The fact that the person we are considering is a fellow speaker whom we understand and know to be proficient in arithmetic is disqualified as a candidate “constitutive” fact of his meaning plus by ‘+’ on grounds of “circularity”. Under that restriction, the paradox succeeds, and so draws attention to what still needs to be understood if we are to make sense of someone’s meaning a certain thing by an expression. It succeeds in this form because the range of potentially “constitutive” facts is not allowed to extend to any of those facts we all know about what fellow speakers whom we understand say and mean. We do make sense of others, and of what they mean, but we can say or specify what they mean only by using words of our own to specify it. And we can do that only because we and our fellow speakers understand the terms and concepts we all use in making sense of one another as speaking agents with intentional attitudes.

On this fundamental point about the understanding of use and meaning, Davidson and Wittgenstein are fully in accord. What lies behind their agreement is the truth of an important observation Wittgenstein wrote in a notebook in 1931, not long after his return to philosophy:

The limit of language is shown by its being impossible to describe the fact that corresponds to (is the translation of) a sentence, without simply repeating the sentence.34

When Wittgenstein speaks here of “describing the fact that corresponds to (is the translation of) a sentence” I take him to mean saying what the sentence means. He says you cannot say what a sentence means without simply repeating the sentence. Perhaps you do not always strictly have to use exactly the same words. Maybe you could say what the sentence says slightly differently without repeating it, or saying in another language what it means. In saying you have to repeat the sentence, Wittgenstein is stressing that you cannot just mention the sentence, or quote it.

33 Kripke 1982.
34 Wittgenstein 1980, 10. I have discussed some of the implications of this remark in Stroud 2011.
I take the main point to be that you must use some words whose meaning you understand in order to say what a sentence means. You give the meaning of the sentence in uttering the words you utter, and anyone who understands you thereby knows what that sentence means. If you said only that a certain sentence you refer to means the same as a certain other sentence that is simply named or quoted or otherwise displayed, you would not thereby say or specify what the sentence means. You would say only that the sentence means the same as some other expression whose meaning you have not stated or specified. It is possible to know, and so to say, that two sentences mean the same without knowing what either sentence means.

When you do say what a certain expression means, or what a particular person means by an expression he uses, you must use some words whose meaning you yourself understand, which you take to express what the expression, or the other person, means. And you could not recognize what the expression or the person means if you were not master of the grammar and vocabulary needed to say what the expression or the person means. To that extent speakers must share with other speakers they understand a capacity to use the terms or concepts needed to express their shared understanding of those meanings and thoughts.

This is the basic idea at the heart of Davidson’s conception of meaning and understanding and his account of interpretation: the intensional vocabulary we use in speaking of what expressions or speakers mean cannot be reduced without remainder to something understood purely extensionally. There is no describing in extensional terms alone our saying and meaning and thinking the things we do. If we tried to understand the things human beings say and do “from outside” all meaning and intention, or as Wittgenstein puts it in Philosophical Grammar, if we treated those utterances and actions only as “a phenomenon or fact”:

Whatever phenomenon we saw, it couldn’t ever be intention; for that has to contain the very thing that is intended, and any phenomenon would be something complete in itself and unconcerned with anything outside itself, something merely dead if considered by itself … Every phenomenon seems dead in comparison with the living thought.35

This same idea is to be found in Davidson’s stress on what “interpreters” must share with those they “interpret” and understand. It is the source of his resistance to the idea of a “conceptual scheme” that is totally different from anything we can make sense of, and so is impenetrable to us but possibly available and intelligible to others.36 In this, Davidson concurs with Wittgenstein that “shared human behaviour is the system of reference by means of which we interpret an unknown language”.37 But the “behaviour”

referred to here is not “behaviour” as the philosophical “behaviourist” understands it: something describable without attributing any intentional attitudes to agents. For Davidson, then, as well as for Wittgenstein, “It is not only agreement in definitions, but also (odd as it may sound) agreement in judgements that is required for communication by means of language”.38

References


38 Wittgenstein 1953, §242.
1. If we are looking for a narrative thread that ties together the kaleidoscopic ruminations of the first 137 sections of the “album” of “philosophical remarks” (PI §4) that is Philosophical Investigations, an attractive candidate is the philosopher’s perennial quest to grasp “Das Wesen der menschlichen Sprache” (§1). Throughout these sections, Wittgenstein represents his diverse observations, examples and arguments as addressed to attempts to articulate language’s “essence”, either by serving as rejoinders to particular “pictures” of a supposed “essence”, or as directly challenging the idea that there is such an “essence” to be grasped at all.

Let us briefly remind ourselves of the pervasiveness of this framing.

The Investigations begins with a passage from Augustine, about which Wittgenstein’s first comment is that it “gives us a particular picture of the essence of human language” (§1). This “picture” is said to be the “roots” of a “philosophical concept of meaning” according to which “every word” has a “meaning” that is “the object for which it stands” (§1–2). “Augustine’s conception of language”, Wittgenstein writes, “surrounds the working of language with a haze which makes clear vision impossible” (§§4–5). The examination that follows of “primitive forms of language”, such as the “language-game” of the builders and their assistants chronicled in §2, proceeds from the hope that this examination will “disperse the fog” produced by Augustine’s conception (§5).

In §46 we are offered a snippet of another classic text, the Theaetetus. The quoted passage posits “primary elements” that admit of no “determination” other than a “bare name”. It concludes with the pronouncement that “the essence of speech is the combination of names” (§46). Wittgenstein tells us that he will “apply the method of §2 to the account in the Theaetetus” (§48). His ensuing examination of various primitive “language-games” encourages us to recognize the relativity to interest and topic of our application of the concepts of “simple” and “complex”, and to reflect on the implications of this fact for the nature of the understanding we can achieve by the “analysis” (§§60, 64) of language into “simple” or “primary” elements.
In §65, Wittgenstein suggests that the preceding discussion is likely to prompt an objection: “You take the easy way out! You talk about all sorts of language-games, but have nowhere said what the essence of a language-game, and hence of language, is” (§65). Wittgenstein does not regard this protest as trivial or beside the point. On the contrary, it raises “the great question that lies behind all these considerations” (§65). Moreover, the objection is partly correct: Wittgenstein has not tried to provide the kind of comprehension of language the protesters are after. But that is not because he wants to take the easy way out; it is because language has no such “essence” to comprehend. Wittgenstein says that he will “try to explain this”. In the sections that follow, he enlists the idea of “family resemblances” (§67) in a bid to expose as misbegotten our hankering for the imagined form of understanding or insight into language with which we associate the rubric of “essence”.

“These considerations”, Wittgenstein now writes in §89, “bring us to the problem: in what sense is logic something sublime?” The perception that “logic has a peculiar depth” (§§89), and our desire to plumb that depth, “finds expression in questions as to the essence of language, of propositions, of thought” (§92). The appearance of profundity surrounding such questions is fuelled by the perception that “the essence”, whatever it may be, “is hidden from us” (§92). We are subject to the “illusion that what is peculiar, profound, essential, in our investigation, resides in its trying to grasp the incomparable essence of language” (§97). It helps in puncturing this illusion, Wittgenstein tells us, to appreciate the force of the observation that “if the words ‘language’, ‘experience’, ‘world’, have a use, it must be as humble a one as that of the words ‘table’, ‘lamp’, ‘door’” (§97). This in turn is an instance of a dialectical move that has application elsewhere in philosophy: “When philosophers use a word—‘knowledge’, ‘being’, ‘object’, ‘I’, ‘proposition’, ‘name’—and try to grasp the essence of the thing, one must always ask oneself: is the word ever actually used in this way in the language-game which is its original home?” (§116)

2. The presence of the “essence” leitmotif is undeniable. But what are we to make of it? The philosophical topics explored in these sections, while interconnected, are wide-ranging and nuanced. (They are just as wide-ranging and nuanced here as they were when first taken up in the *Tractatus*, a text with which the early portion of the *Investigations* is obviously and explicitly in conversation.) If Wittgenstein has a reason for collecting the diverse bits of philosophy or proto-philosophy he discusses under one label, it is certainly not because they are addressed to some single, discrete philosophical question about linguistic meaning. The common element or theme indicated by the label, if there is one, must be a good deal more abstract.

I do think there is such an abstract commonality. What talk of “essence” signals throughout these sections, I suggest, is an aspiration on the part of
Wittgenstein’s interlocutor to achieve an insight into linguistic meaning that would be, in a sense I will shortly explain, external to our understanding of language in use. Whatever the details of the particular philosophical inquiry or debate within which a given picture or claim is entered, it counts as a picture of or claim about language’s “essence” by embodying this aspiration—and so by depending for its point and sense upon the aspiration’s intelligibility. Wittgenstein’s primary critical message throughout this region of the text is that the prospect of such external insight into linguistic meaning is a fantasy.

The best way to get a grip on what an external understanding of linguistic meaning would be is to consider the kind of understanding of meaning with which it is meant to contrast. Our acquaintance with linguistic meaning, at least in the first instance, resides in our speaking a language. To speak a language is to have the capacity to use that language to say things. Exercises of such a capacity are intentional: if all goes well, you say what you intend to say. It follows that, in general, when you utter a sentence of a language you speak on some occasion, you know what you are saying. Insofar as others share your language, you will typically know what they are saying as well. The capacity to speak a language, then, entails a capacity to know what speakers of a language say when they utter sentences of that language. We might call what is here understood utterance-meaning; it is what is said or meant by a particular use of the language.

Now, while you are in general free to say whatever you like in a natural language you speak, you are not free to choose whatever expressions you like to say it. You cannot, speaking English, utter, “This vest is an odd shade of fuchsia,” and thereby say that you’re eligible to vote in Illinois. Particular English words, when conjoined together into particular sentences, can be used in particular circumstances to say particular things. A fluent speaker or writer of English will have a good sense of what is and is not thus possible in the language as it stands; a creative speaker or writer will point the way to new possibilities ramifying out from the old. That words in a natural language have concrete possibilities for, and concrete limitations upon, what they may be used to express; that these possibilities and limitations are in general fixed independently of any given uses of these words; and that it is part of the competence of a speaker of a language to have a firm grasp of these possibilities and limitations—all of this together justifies our distinguishing a second kind of meaning associated with language. This is expression-meaning, a property or character that expressions in a natural language possess independently of particular instances of their use.

The existence of two related forms of linguistic meaning, utterance-meaning and expression-meaning, raises the question of their relative priority. The answer is that there is dependence in both directions. On the one hand, it is in part in virtue of what expressions mean, independently of our particular uses of
them, that we may use them to say the things that we do. There is thus a clear sense in which meanings of expressions precede and help provide for the meaning of their uses, for what speakers say or mean in using them on given occasions. On the other hand, our reflections to this point suggest that a speaker’s understanding of what a word means is ultimately a matter of her understanding of what the word is, and can be, used to say. We were motivated to introduce the idea of the understanding of expression-meaning from consideration of what a person who can speak a language must know. And the capacity to speak a language is, of course, a capacity to use a language. More specifically, it is the capacity to use a language to say what one intends to say. So far as this capacity is concerned, it is use, and the understanding of use, that is fundamental.1

This thought is reinforced by attention to what is at stake in questions about our understanding of the meanings of particular words. When asked to explain the meaning of a given expression, what we say or do in response aims to make clear to our audience what a person who uses this expression would be speaking of, or talking about, or otherwise meaning, when she uses the expression. It is no accident that students tested on vocabulary are asked to “use the word in a sentence”. The whole point of the exercise is to expand their boundaries of usage and comprehension of usage.

To reiterate: our engagement with linguistic meaning, in all its forms and guises, lies in our active exercise of our capacity to speak a language—to use it to say things—and thus to understand such uses. It is in exercising that capacity that we find application for the concept of linguistic meaning, both with regard to uses of expressions and to the expressions themselves. The “home” (§116) of this concept lies in these exercises: thought and talk about “meaning” expresses what we know about words and utterances when we bring this capacity to bear upon them.

At any rate, this is the primary, or initial, home of the concept. But need it be its sole home? The capacity to speak a language is itself something about which we may talk and think. It is a potential object of reflection. In particular, given the centrality of language to human life, and given its obvious connections with thought, knowledge, action and other important categories, it would seem an eminently fertile subject for philosophical inquiry. It’s obvious that consideration of meaning will factor prominently into any such inquiries. And as we undertake such inquiries, the perception can easily arise that there is a need for, and a prospect of achieving, a second home for the concept of linguistic meaning. Or, to vary the figure, we may find ourselves thinking we require,

1 C.f., Cavell on Wittgenstein’s talk of meaning as use: “‘The meaning is the use’ calls attention to the fact that what an expression means is a function of what it is used to mean or to say on specific occasions by human beings” (Cavell, 1999, p. 206).
and can have, a new perspective on the phenomenon of linguistic meaning, a perspective additional to that constituted by our exercise of the capacity to understand, in languages we speak, what people say and what words mean. To achieve such a perspective would be to find a way of recognizing and thinking about the meaningfulness of expressions and their uses that does not consist in our viewing those expressions and uses in the way that we do when we bring to bear upon them our capacity to speak and understand a language. The insights about meaning thus obtained would be in this sense external to our understanding of language in use.  

Of course, it is possible to recognize an expression or use of language as meaningful without understanding what it means. I know that words and utterances of Polish are meaningful, and I know this without exercising a capacity to speak Polish, for I have no such capacity. But that does not mean that I have a perspective on the meanings of Polish words and utterances that is external in the desired sense. My recognition of the meaningfulness of Polish words and utterances lies in my knowledge that there are people who speak Polish, and so that for these people the utterances and expressions of Polish have the same life that utterances and expressions of English have for me. There is nothing in this recognition to take me beyond what I know of linguistic meaning in possessing and exercising a capacity to speak a language. An external perspective must be a source of positive insights about meaning—about what it is, or about what makes or constitutes or provides for it—that does take me beyond such knowledge.

In philosophical reflection upon linguistic meaning, we may arrive at conclusions or observations that we were not in a position to formulate and recognize as correct in advance of this reflection. For example, we may uncover linkages between the concept of meaning and other fundamental concepts of human life and thought, such as that of truth, belief, intention, and so on. If we can bring these linkages to light and find accurate and productive ways to characterize and discuss them, this will be an intellectual achievement. The fruits of this achievement may fairly be regarded as new insight into, or new knowledge about, meaning and language. But the newness of the insights will not suffice to show that we have achieved an external perspective on meaning. Whether we have done so will depend upon whether recognition of these linkages puts us in a position to identify or explain or account for meaning in a way our understanding of which was not already implicit in exercises of our capacity to speak and understand language. It is arguable that many philosophers, historically, have been interested in these linkages precisely in virtue of the hope, explicit or inchoate, that they will show the way toward such a position. But in itself, the sheer prospect of identifying the linkages does

---

2 C.f., Stroud 2011, on Wittgenstein on our desire to get “outside language”.
not show that such a position is attainable. Indeed, it might be taken to cut the
other way: it might be taken to suggest that not only the concept of linguistic
meaning, but those of truth, belief, intention, and the rest are accessible only in
the active exercise of a set of interlocked capacities for thinking about others
and ourselves within which the capacity to speak and understand language is an
essential and irreducible part.

What, then, would be involved in genuinely occupying an external perspec-
tive on linguistic meaning? There are many ways we might try to convince
ourselves we have found the materials to achieve such a perspective. But what
any such attempt must involve is an effort to impose upon linguistic meaning
conceptual templates that are drawn from somewhere other than the forms of
thought about meaning in which we engage when we exercise our capacity to
speak and understand a language.

Consider, by analogy, the scientific treatment of a “natural kind” concept
such as that of water. It is no part of our ordinary concept of water that water has
a certain chemical composition. One can perfectly well talk and think about
water, and understand others’ ordinary thought and talk about water, without
knowing anything about its treatment in chemistry. The now-familiar identity
claim, “water is H₂O”—not to mention the more accurate (I am told), “Pure
water is a mixture of ions composed of hydrogen and oxygen atoms that
average out two to one”—embodies an alternative perspective upon water,
a perspective that is external to the recognitional-cum-practical capacities
constituting our possession of the ordinary concept. That our ordinary concept
of water is a “natural-kind” concept arguably provides for the intelligibility of
these macrophysical/microphysical identity claims. But the material that fills
out these claims—that furnishes our accounts of water’s “real essence”—is not
built into our ordinary concept or into any knowledge or capacity that concept
might plausibly be thought to implicate.

Some naturalistically minded contemporary philosophers appear to think of
linguistic meaning as a “natural kind” possessing a scientifically discoverable
“real essence”. But that is certainly not so of all, perhaps even of any, of the
interlocutors whom Wittgenstein engages. The point of the analogy is more
abstract. What Wittgenstein is concerned to root out, in our philosophical
reflections on linguistic meaning, is our inclination to seek insight into the
ground or nature of meaning by recasting it in terms of models, pictures or
structures that are not native to the forms of thought about meaning in which we
engage when we exercise our capacity to speak and understand a language.
A range of perceived philosophical puzzlements may fuel this inclination. But
whatever their motivation, our efforts to satisfy the inclination, Wittgenstein
will argue, are bound to yield only confusion.

3 See, e.g., many of the essays in Millikan 1993.
3. It is a measure of how fundamental—we might say, of how primitive—is the level at which this mode of criticism operates that Wittgenstein sees a fateful step in the wrong direction simply in our attaching explanatory or analytic significance to the thought that there are meanings. As soon as we think to posit meanings—or ‘propositions’, or ‘thoughts’—as distinctive elements, items or objects whose presence in a scene of language use somehow helps explain or account for the meaningful character of that use, we have embarked on the search for an external perspective on meaning. For the vision of meanings-as-special-objects is not a deliverance of the exercise of our capacity to speak and understand a language. On the contrary: from the perspective of that capacity, it is an excrescence.

This crucial claim is made in various places, but it is given special emphasis in the initial sections of the Big Typescript (sections which, in that text, precede and set the stage for the discussion of Augustine that Wittgenstein will later choose to begin the Investigations). Consider, for example, the following remark: “The question, ‘What do you mean?’ is answered by ‘I mean p’, and not ‘I mean what I mean by ‘p’” (BT 3). That is to say, we do not clear up a question about what we mean by drawing attention to a “meaning” associated with our use of a sentence, conceived as some kind of item that constitutes “what we mean” by that sentence. Rather, we just use a sentence to say what we mean.

The text goes on:

If one always expresses oneself in a system of language and so uses only propositions of this system to explain what a proposition means, then in the end meaning drops out of language completely, and thus out of consideration; what remains is language, the only thing we can consider. (BT 3)

The point here is not that the languages we speak are in fact meaningless systems of signs. The point is that we do not need to “consider” any items additional to the sentences (here translated “propositions”) we use in order to register the meaningfulness of these sentences and uses. Our appreciation of their meaningfulness stems not from our having in view any further such items, but simply from the fact that our relationship to these sentences is one of our using those sentences to express ourselves, to speak.

And again:

If I give someone an order then it is quite enough for me to give him signs. And I would never say: ‘These are mere words, and I have to get behind them.’

... “But I meant something when I said that!”—Fine, but how can we recover what it was? Surely only by his telling us .... You mean what you say. (BT 4)

That there is an item “behind” the words we use determining what we mean by them is not a thought we have when we are exercising our capacity to give or
understand orders, or to tell someone something or to grasp what someone has told us. It can seem fruitful, or necessary, only if we abstract from what we know of utterances in exercising our capacity to make and understand them, and then cast about for some alternative way of conceiving the now absent “meaning”.

4. What philosophical puzzlement might motivate our search for an external account of meaning, and how, beyond the bare idea of object-hood, of the reification of meaning, might we then seek to put flesh on our account? As I have said, there are many different possible answers to be gleaned from the *Investigations*, and elsewhere in the later writings. Perhaps the simplest are to be found in the path of reflection mapped in the early pages of *The Blue Book* (Wittgenstein 1965).

There we are said to be struck by the perception that linguistic signs, considered on their own, “seem dead” (BB 4). They are just “dashes on a bit of paper”—or if spoken, just noises. How could such inert bits of physical matter or physical events have the “life” that belongs to the meaningful? This worry can be pressed in terms of what is nowadays called the “normativity” of meaning. An item’s meaning determines conditions of correct use and response. A command’s meaning, for example, fixes what would count as correctly executing it. But how could a mere mark on a paper, or a mere noise, determine any particular condition on how it is to be understood or responded to? “If I give someone the order ‘fetch me a red flower from that meadow’, how is he to know what sort of flower to bring, as I have only given him a word?” (BB 3)

The answer to which we seem forced is that there must be something present on the scene additional to the mere word, something which accompanies the word and does the correctness-determining work that the word itself cannot. And so “we are tempted to suppose that the action of language consists of two parts” (BB 3). On the one hand, there is the “inorganic part”: “the handling of signs”. This part was obvious all along. But what we now take ourselves to have discovered is that there must be an “organic part” as well, a parallel track of events or processes that runs alongside the “inorganic” dealings with words. These further events and processes cannot just be more manipulations of dead signs. Precisely not: they must be such as to constitute the “life” we have found to be absent from such manipulations. It will now seem irresistible to locate these processes in the “mind”. For it seems open to conceive the “mind” as a “queer kind of medium”, capable of sustaining “mental processes” that “can bring about effects that no material mechanism could” (BB 3).

Let us retrace these steps in light of the previous discussion. Our puzzlement was said to arise from our sense that the linguistic expressions with which we are confronted, being mere marks or sounds, are “dead”. Of course, if these expressions belong to a language we speak, and we allow ourselves to bring to
bear upon them our capacity to understand that language, then neither these expressions, nor their uses on particular occasions, will seem “dead”. Exercising this capacity on Wittgenstein’s use of the English sentence, “Fetch me a red flower from the meadow”, his utterance will have for us the “life”, precisely, of an order to fetch a red flower from the meadow. Our perception that the utterance is after all dead, then, must reflect a conviction that we are not allowed to just help ourselves to what we know about it when we exercise upon it our capacity to understand English speech. The capacity, we must be supposing, is not self-vindicating. What it purports to reveal about the words and utterances of our language stands in need of validation, or explanation, from elsewhere.

Our way of attempting an external validation is to postulate special processes occurring in “the mechanism of the mind” (BB 3) that provide the “life” absent from our handling of “dead” signs. The concepts of process and of mechanism pick out ontological categories more determinate than the bare idea of an object or item. So we already seem to have more substance to our externalized conception of meaning than was to be gleaned from the passages from the Big Typescript quoted above. Moreover, the idea of a mechanism might seem especially suited to our explanatory purposes. The very idea of a mechanism is of a system whose characteristic processes can be described in terms of multiple families of concepts. As Wittgenstein puts the point in §149 in discussing the notion of a mental “apparatus”, processes implemented in a mechanism or apparatus are susceptible to two forms of characterization: one in terms of what the mechanism does or effects, and one in terms of how the mechanism is “constructed”. Described in terms of its effects, our posited processes are presumably to be characterized as processes that say, yield or constitute an understanding of Wittgenstein’s utterance as ordering the procurement of a red flower. To grasp the aptness of such a description would be to understand Wittgenstein’s order, and thus to rely upon the very capacity for which we are currently endeavouring to account. But in envisioning understanding as effected by a process in a “queer kind of medium”, we are holding out hope for a different way of characterizing that supposed process, a characterization, as Wittgenstein puts it, in terms of the mechanism’s construction. Such a characterization would not depend upon conceiving these processes as producers of meaning or understanding, but would rather exploit independently available characterizations of structures or configurations in the envisioned “queer medium”.

In fact, we have no inkling, and can have no inkling, of how such structures or configurations are to be described. We know that they cannot be physical, for “no physical mechanism could behave in this way” (BB 5). And so we think of them instead as having an “occult character” (BB 5). But we can get no farther than this. Our attempt to give an external account of meaning and
understanding in terms of the operations of the “mind” has ended in obscurity. Unfortunately, the felt pressure to find an external account of meaning may be strong enough that we do not recognize our failure for what it is. Instead, we may mistake its obscurity for profundity.4

5. Let us return at last to the *Investigations*. I have suggested that the interlocutor’s claims and pictures count as attempts at articulating language’s “essence”, in the manner of which Wittgenstein is critical, in virtue of being offered as attempts to fill out an external perspective upon linguistic meaning. This proposal, as we will see, illuminates various features of Wittgenstein’s discussion.

On its face, however, the proposal may seem a non-starter. The “picture” of the “essence of human language” Wittgenstein discerns in the passage from Augustine casts naming, or reference, as that “essence”. The concept of reference, unlike that of a process or a mechanism, is semantic. It is a concept of word meaning. And so the thought that words refer is not foreign, but native, to the kind of understanding or knowledge yielded by exercises of our capacity to speak a language. This raises the question of how reference could possibly be supposed by Wittgenstein’s interlocutors to provide a plausible basis or starting point for an external treatment of linguistic meaning.

The key to an answer to this question lies in the observation that the concept of the reference of a word is semantic, not pragmatic. It is a concept of expression-meaning, not utterance-meaning. Earlier, I suggested that utterance-meaning is conceptually prior to expression-meaning: that what a word means is ultimately to be understood as a matter of what it can, in conjunction with other words, be used to say. But this thought might be denied. Indeed, the distinction between the meaning of a word and what is meant by speakers in using a word might be taken to open the possibility that a word’s reference, belonging as it does to the first sort of meaning, could be accessible to us independently of our access to meaning of the second sort. We may still grant that what a word means is necessarily linked with what it can be used to say. But we might hope to exploit this linkage in the other direction, building from an external treatment of reference to an external understanding of the whole of meaning in use.

Viewed in this light, the identification of reference as the “essence” of language is one application of a more general thought: that expression-meaning—be it understood referentially or otherwise—is the key to the development of an external treatment of linguistic meaning. I take it that this more general idea is indeed in Wittgenstein’s sights. In §5 he speaks of the confusion that surrounds “the general concept of the meaning [der Bedeutung] of a word”,

---

4 For more on Wittgenstein’s treatment of his interlocutors’ many appeals to the idea of a mental mechanism, see Bridges 2017.
and there is no indication that he thinks the confusion arises only when we conceive meanings of words in terms of their naming of objects. This comment harkens back to remarks from *Philosophical Grammar* that are still more explicit on this point. There Wittgenstein writes, “The way Augustine describes the learning of language can show us the way of looking at language from which the concept of the meaning of words derives” (PG §20). And he expresses our enthrallment to that “concept” this way: “We say: the essential thing in a word is its *meaning*” (PG §22). That is, the very idea that words have meanings may seem to us to point toward, to begin to delineate, some picture of the “essence” of language.

The strategy we are now envisioning depends upon the prospect of finding a second home for the concept of a word’s possession of meaning, and of its possession of reference in particular. Such a home must be constructed from concepts, pictures, or claims whose application to words is not given within the understanding of expressions and utterances internal to our capacity to speak a language. What materials might be thought to be suitable?

There are several answers to this question suggested in *Investigations* §§1–137. One possibility is to try to conceive reference or signification as constituted by operations in the mind. We might suppose, for example, that a word’s referring to what it does is a matter of its being associated with a mental image of its referent (c.f., §6). If the hopelessness of that thought becomes clear to us, then we may seek recourse in the idea that the mind is able to forge referential connections between words and objects in virtue of its mysterious and unique powers. We adopt a “conception of naming as a process that is, so to speak, occult. Naming seems to be a strange connection of a word with an object … [W]e may indeed imagine naming to be some remarkable mental act, as it were the baptism of an object” (§38). Taking this route, we are squarely back in the territory examined in the *Blue Book*. In the sections after §137, similar ideas about understanding are considered in great detail.

Second, we may endeavour to cast referential relations as implemented by a set of linguistic practices that can be separated out from, and understood independently of, the rest of our use of language. The supposed discrete character of these special practices may even encourage the thought that they are not, strictly speaking, uses of language at all, but are instead preparatory to such uses. This is the suggestion of the interlocutor §27: “We name things and then we can talk about them: can refer to them in talk.” If naming happens in advance of talking, then, it might seem, we should be able to recognize a word as naming something without relying on our understanding of its role in talk. We might offer as an analogy: we sharpen a stone and only then can use it to cut things. Its sharpness must not simply be a matter of its being used to cut, and must therefore be cognizable in abstraction from that use.
Wittgenstein’s response to the interlocutor’s suggestions shows that he would regard any such analogy as profoundly misconceived. A stone might be sharpened and never be used to cut anything. But a word names nothing except insofar as it is used: “Naming is so far not a move in the language-game—any more than putting a piece in its place on the board is a move in chess. We may say: *nothing* has so far been done, when a thing has been named. It has not even *got* a name except in the language-game” (§49). And that is because to say what a word names is to say something about what is meant by those who use it: “What is the relation between name and thing named?—Well, what *is* it? Look at language-game (2) or at another one: there you can see the sort of thing this relation consists in” (§37).

Third, we may try to grasp referential relationships from outside our capacity to use and understand language by supposing that these relationships are to be identified by a special activity of “analysis” of language (§92). What is distinctive about this activity is that it is not just a matter of being explicit, or observant, or self-conscious, about patterns and structures discernible within what we know, or can come to recognize, in exercising our capacity to speak and understand a language. The envisioned activity is not in this sense a matter of examining “something that already lies open to view and that becomes surveyable by a rearrangement” (§92). Rather, it is a matter of getting at what “lies beneath the surface”. “The essence” that we are after “is hidden from us”. Digging for such an “essence” will require our going beyond, or beneath, what is already available to us as speakers. It will be to seek an external perspective.

Such a project is motivated by, and derives whatever sense it may have, from a perception that we in some way lack an adequate philosophical understanding of the status of language until we grasp its “hidden essence”. We might attempt to indicate the place of the missing understanding by speaking of, for example, the need to grasp the “super-order” between the “super-concepts” of “language” and “world” (§97). Reference, understood as a “bare”, unmediated confrontation between the words of a language and “the simple constituent parts of which reality is composed” (§47) might seem an essential tool in the articulation of such an order.

This third imagined strategy for achieving an external understanding of reference, and through reference an external understanding of meaning as such, is the most complex and difficult to understand. I will say no more about it except to register that this is where the later Wittgenstein’s treatment of the search for a hidden essence of language most explicitly confronts his earlier work:

My notion in the *Tractatus Logico-Philosophicus* was wrong: 1) because I wasn’t clear about the sense of the words “a logical product is hidden in a sentence” (and suchlike), 2) because I too thought that logical analysis had to bring to light what was hidden (as chemical and physical analysis does). (PG 210)
6. Let us now consider how the current reading illuminates Wittgenstein’s initial moves in response to “Augustine’s conception of language”, according to which “every word has a meaning” and “the meaning is the object for which the word stands”. I am suggesting that the inclination to think that these words express the “essence” of language amounts to the intuition that the idea of a word’s signifying an object can somehow be deployed to achieve external philosophical insight into linguistic meaning. At this very early stage of the *Investigations*, there is as yet little to no indication of what kind of philosophical puzzle or question might seem to motivate the search for such insight. This means that what is available at the outset for examination and criticism will be nothing more specific than the thought that the phenomenon of reference or signification is suited, in some way or another, to achieving some such insight of this kind. Thus whatever challenges are to be raised at this stage about the “philosophical concept of meaning” according to which signification or reference is language’s essence, these challenges must engage this highly general commitment. And indeed, it is precisely this general thought that Wittgenstein seeks to cast doubt upon in the sections immediately following §1. The main drift of these sections can be put succinctly as follows: we do not have and cannot have any grasp of what words refer to except in virtue of understanding what they can be used by speakers to say or otherwise mean.

Wittgenstein does not attempt to prove this denial—not surprisingly, given that it is difficult to see how anything approaching a proof could be in the offing when we are jousting with such abstract, indeterminate explanatory hopes. But he does still try to give the denial teeth. How does he do so?

To begin with, he emphasizes the great diversity of uses that words possess. He says that someone who characterizes the learning of language as Augustine does must be “thinking primarily” of proper names and common nouns, and forgetting about other kinds of words (§1). He describes a linguistic interchange involving a shopkeeper that provides an initial illustration of that diversity (§1). Augustine, he then tells us, is like someone who offers a circumscribed account of games, to which we can reply, “You seem to be thinking of board games, but there are others” (§3). He deploys various images—of tools in a tool box (§11), of the controls in a locomotive (§12)—as metaphors for the variety of “absolutely unlike” (§10) uses to which we put words that may seem similar enough on their visible or audible faces.

Some readers have taken the point of these remarks to be that it is straightforwardly wrong to claim that every word signifies or names an object.5 This would seem to provide a direct refutation of the picture of “essence” given in §1: presumably, the naming or signifying of objects by words cannot be the

5 Most famously, Baker and Baker 2005, Chapter 1.
“essence” of language if there are lots of words that don’t name or signify objects.

But this reading is off-target. Wittgenstein never outright claims that there are words that cannot be understood to signify anything. And to take his point to be such a flat-footed denial goes against conspicuous features of the text. One of the words whose use in the example in §1 is meant to cast doubt on “Augustine’s conception” is the word “five”. But in §10, Wittgenstein grants that it can perfectly acceptable to say that numerals “signify numbers”. If a word like “five” poses a problem for Augustine’s conception, it is apparently not by serving as a clear-cut counterexample to a well-defined universal generalization.

Furthermore, the attempt to portray naming or reference as the “essence” of language is, as I have suggested, just one possible way to try to make good on the thought that expression-meaning is the “essence” of language. If this more general thought is to be exorcised through recognition of the failure of the picture of reference as language’s “essence”, then the morals of the latter’s failure must be generalizable. It will not do, then, to rest much weight on the insistence that some expressions lack reference. For these expressions will nonetheless have meanings, even if not referential. What needs to be made plausible is that we have no way of understanding a word to possess a meaning except by conceiving of it as apt for use in saying this or that.

It is just this point that Wittgenstein’s highlighting of the variety of uses of words aims to help us to see. Failing to register this variety, it might seem possible to think that our appreciation of the concept of the meaning of a word is a matter of our bringing to bear a common template upon all words, a template whose application in each case is understood and justified in more or less the same way. The abstract relational property of standing for an object might, indeed, seem perfectly suited to this role. This thought might suggest in turn that we can grasp how this template is to be applied to a word in advance of knowing very much about the particular uses to which the word is put. And this may, finally, pave the way for the thought that we can fairly readily substitute for our ordinary concept of expression meaning some surrogate constructed out of external materials.6 But once we acknowledge just how radically diverse and variegated these uses can be, the whole train of thought will fall apart. For the initiating idea of a common template will lose its plausibility. We will come to recognize how utterly dependent is our grasp of a word’s “meaning” upon our appreciation of the particular ways in which we use the word. We will come to

---

6 A cousin to this line of thought can be found in contemporary work on “naturalized semantics”, in which simple, streamlined structures of meaning are represented as the core of meaningfulness as such. The simplicity of the structures is precisely what seems to open the prospect of their reconstruction out of “naturalistic” relations such as those of nomological connection or evolutionary fitness. The work of Fred Dretske (1981, 1988) exemplifies this approach.
see the truth in the slogan that “the meaning of a word is its use in the language” (§43; c.f., PG §23).

The intended import of Wittgenstein’s appeal to the diversity of our use of words becomes clearer when we juxtapose that appeal with his treatment of a case in which no such variety is at issue. In §2, Wittgenstein suggests that we “imagine a language for which the description given by Augustine is right” (§2). This is the “primitive” language of the builders and their assistants. In this language there is only one kind of use for a word: each word in the language is a call for the assistant to bring a building stone of a particular shape. Does this mean that the language is a good fit for “Augustine’s conception of language”? Evidently it does not:

Now what do the words of this language signify?—What is supposed to show what they signify, if not the kind of use they have? And we have already described that. So we are asking for the expression “This word signifies this” to be made a part of the description. In other words the description ought to take the form: “The word … signifies …”.

Of course, one can reduce the description of the use of the word “slab” to the statement that this word signifies this object. This will be done when, for example, it is merely a matter of removing the mistaken idea that the word “slab” refers to the shape of building-stone that we in fact call a “block”—but the kind of ‘referring’ this is, that is to say the rest of the use of the word, is already known.7

Wittgenstein does not deny that the words in language-game §2 signify objects. On the contrary, he grants that they do. His point is rather that what “shows what they signify” is “the kind of use” they have. Certainly these words refer to things, but “the kind of ‘referring’” in play is a function of “the rest of the use of the word”. That is to say, we grasp what these words refer to only by understanding Wittgenstein’s description of their use. Asking ourselves why we find it natural to say or think that “slab” in language §2 refers to a certain shape of building stone, we can do no better than repeat, or summarize, the information provided in Wittgenstein’s description.

Some philosophers who take themselves to be influenced by Wittgenstein promote a “use theory of meaning”, according to which the meaning of a word or sentence derives from facts about its use that may be described without mention of what speakers mean in using it.8 If Wittgenstein himself endorsed this view, we would expect him to provide descriptions of the use of language of the sort the view requires. But there is no sign anywhere that he is interested in doing so, and he certainly does not do so here. It is stipulated as part of the description of use in §2 that the builders’ words constitute a “language meant to serve for communication”. In particular, we are told “B has to pass the stones

---

7 For the final sentence of this passage I use Hacker’s and Schulte’s superior translation, in their 2009 Wiley-Blackwell edition of PI.
8 See especially Horwich 2005, Chapter 2.
and that in the order in which A needs them”. “For this purpose”, Wittgenstein goes on, they “use a language” in which A calls out a word of the language, and B brings the stone “he has learnt to bring at such-and-such a call”. It is simply given in Wittgenstein’s description, then, that the uses at issue are purposive speech acts. In particular, the description makes clear, as Wittgenstein says in §18, that these speech acts are commands.

The upshot of the reflection on language-game §2 undertaken in §10 is not that a word’s reference is fixed by facts about “use” that may be described in neutral, non-meaning-involving terms. Rather, it is that recognition of what a word refers to is available to us only via the exercise of a capacity to grasp characterizations of the use of language in which it is given that the speaker is saying and meaning particular things in using those words as she does. This implies that an appeal to reference cannot give us any insight into the “essence” of language of the sort we think we need.

It is this kind of point that is apt for generalization to expression of meaning as such. Take any putative semantic property you might think helpful to ascribe to words and other expressions. The claim will be that such ascriptions, insofar as they are intelligible to us, encapsulate patterns in how these expressions may be used to say things.

7. The primordial level at which Wittgenstein’s battles with “the search for the essence of human language” unfold gives his thought its distinctive power and reach. But it also presents an obstacle in bringing this thought to bear upon the work of actual philosophers.

Wittgenstein does not, in this region of his later work anymore than elsewhere, engage the characteristic results of philosophical theorizing: sophisticated theories, systems and arguments. His concern is not in where philosophical reflection ends up, but in where it begins, in the assumptions, ideas, pictures, impulses and confusions that lie at its roots. By the same token, he is not generally concerned to show that this or that philosophical thesis or theory is false, or that the arguments for them are unsound. It is far from clear that falseness and unsoundness are even in the cards. The worries cut deeper. They suggest that our attempts to articulate theses or construct arguments are controlled by intellectual aspirations that we do not really understand, indeed, that are not ultimately intelligible. But if that is so, then no claim or picture we put forward will have the significance we want it to have. Nor will they have any other significance, for what would give it to them if not our interests and intentions? So we are just spinning our wheels.

No philosopher is likely to accept this discouraging verdict on her activity. And given the great distance between the primitive pictures and notions upon which Wittgenstein focuses and the intricate subtleties of actual philosophical work, there will always seem room to deny that the work partakes of the impulses and tendencies Wittgenstein seeks to expose and critique.
Moreover, the denial might be correct: not all philosophical reflection on language, after all, is automatically guilty of hankering after “essence”. So how, in any given case, are we to tell?

I will close by brief consideration of an instructive test case: Donald Davidson. The case is instructive because while Davidson both shares many of Wittgenstein’s convictions about language and explicitly disavows any reductionist or naturalistic aims in his account of meaning and thought, I think we can nonetheless detect in some of his core theses and arguments the signs of an interest in uncovering language’s “essence”. That even Davidson does not fully avoid the lure of the quest for “essence” indicates the strength of its hold upon our philosophical imagination.

Davidson’s points of overlap with Wittgenstein are significant. First, his thinking on language is organized around an insistence on the conceptual priority of utterance-meaning over expression-meaning. It is true that he does not conceive that primacy in quite the way that Wittgenstein does, largely because he does not share Wittgenstein’s conception of use. Most notably, Davidson finds what is essential to the social dimension of language use in pair-wise communicative interaction, disparaging the theoretical importance of shared natural languages, and doing so for reasons which, if valid, would equally discredit Wittgenstein’s appeals to language games and practices as necessary contexts for meaning.9 But at the same time, Davidson also enriches our appreciation of the context required for meaningful language use. He reminds us that our capacity to ascribe meanings to a speaker’s utterances is a component of our capacity to grasp the whole mise-en-scène of the speaker’s rational engagement with the world. If we are to know what a speaker is saying, we must also grasp what she thinks and intends, and we cannot do that unless we can find rational intelligibility, and a firm grasp of reality as we understand it, in those thoughts and intentions. Among other valuable results, Davidson’s emphasis on the interdependence of meaning and “the propositional attitudes” provides further grounds for rejecting a mechanistic treatment of the mental.

Second, Davidson shares Wittgenstein’s distrust of hypostasizing talk of propositions, meanings, and other such items. He sees in it little explanatory power, and much opportunity for confusion. A major part of the motivation for his proposal to use a Tarski-style theory of truth as a theory of meaning is to relegate explicit talk of meaning to the background.

How about the other side of the ledger? Where I think we can see traces in Davidson of an un-Wittgensteinian desire to find an external perspective on linguistic meaning is in the enormous importance he assigns to the conceptual possibility of “radical interpretation”—i.e., of the derivation of a

9 See, especially, Davidson 1986.
comprehensive assignment of meaning, beliefs, and desires to a speaker solely on the basis of knowledge of what sentences of her idiolect she “holds true”. In principle, we can know that a person holds a sentence she utters to be true without knowing what she means by that sentence, and so correlatively without knowing what beliefs and desires she thereby expresses. Part of the burden of the famous essay, Davidson 1973, is to explain how we could exploit presumptions of charity and rationality to bootstrap knowledge of sentences held true into full-fledged knowledge of meaning, belief and desire. Seventeen years later, in Davidson 1990, the prospect of such a derivation is still front and centre, with the required principles, assumptions, and technical machinery spelled out with much greater care. In these essays and elsewhere, Davidson draws substantial and surprising philosophical conclusions from his account of the procedures of radical interpretation. But so far as I can see, he never provides a satisfactory explanation of why we should draw these conclusions, or indeed any philosophical conclusions, from the account.

One of Davidson’s conclusions is that it is a necessary condition on being a speaker and thinker that one’s utterances can in principle be subject to radical interpretation. But why should this be so? Suppose we were to discover some flaw in Davidson’s radical-interpretive method, which exploits ideas from Ramsey’s theory of subjective probability. Why should we be confident that the flaw could in principle be corrected? Davidson is careful to emphasize that he is not offering a reconstruction of how we actually do understand the speech of others. Why then, should we believe in, or care about, radical interpretation’s conceptual or logical possibility?10

A second conclusion is that meaning is indeterminate. Davidson represents this indeterminacy as no less benign and unexceptional than the relativity of an object’s numerical length to the units of measure: vary the unit, and you change the number.11 It is the interdependence of assignments of meanings and assignments of propositional attitudes that is here being analogized: vary the belief ascribed, say, and you must adjust your assignment of meaning accordingly. But what, then, is analogous to the side of the object that is measured? The answer must be the speaker’s linguistic behaviour. And if the analogy is to make sense, the behaviour, in itself, must not be capable of deciding between different paired assignments of meanings and attitudes, anymore than an object’s side can be such as to decide between descriptions of it as 12 inches long vs 1 foot long. If we suppose that a person’s behaviour, in the relevant sense, is constituted by facts about what sentences she asserts or holds true,

10 Davidson’s later work on ‘triangulation’ introduces material relevant to this question. For a criticism of that work, see Bridges 2006.
then this result is achieved. It seems evident that, primed by his reflections on radical interpretation, Davidson does conceive the speaker’s behaviour in this way. But why? Why not suppose that the proper conception of the behaviour to which our ascriptions of attitudes and meaning are answerable is one according to which it is not merely given that a speaker asserts and holds true the sentence, say, “It’s raining,” but it is given as well that in uttering this sentence, she says, precisely, that it is raining?

These gaps in Davidson’s arguments might be filled if we could find a reason for demanding that philosophy show how meaning arises from, or is grounded in, facts about which sentences a speaker asserts or holds true. One possible justification for this demand is that its satisfaction would be a step toward achieving externality.12 Is there anything else to be said for it?

References


---

12 This is in effect David Lewis’s justification for taking an interest in the problem of radical interpretation. See Lewis 1974, p.110.
Do at least some non-human animals (henceforth simply “animals”) have minds comparable to those of humans? On this question, there are two basic stances. Differentialists maintain that there are categorial differences separating us from animals; assimilationists hold that the differences are merely quantitative; they diagnose continuity between our mental capacities and those of higher animals (see Brandom 2000, 2–3). The main differentialist argument has always been that mental powers, at least of a more advanced kind, require language. This position can be labelled “lingualism.” The reasons could be factual ones established by empirical science. Alternatively, mental capacities could require linguistic competence for loosely speaking a priori, conceptual reasons.

Both Wittgenstein and Davidson are generally interpreted as a priori lingualists. This chapter tries to provide a more nuanced account of their positions, to compare them, and to assess them for their philosophical merits. It starts out by highlighting the respective backgrounds of their discussions of animal mentality, which turn out to overlap to a considerable extent. Both adopt a third-person perspective on mental phenomena without lapsing into behaviorism. Both are also united in granting animals “lower” mental faculties like sensation, by contrast to some who profess to follow their lead. A fundamental contrast opens up with respect to so-called intentional states like believing, desiring, and intending (following Davidson 1984, 156; 2001, 98–100, I shall also refer to these as “thought” or “thinking”). Whereas Davidson denies that non-linguistic animals can possess any thoughts, Wittgenstein favors an intermediate position. Animals are capable of beliefs, desires, and intentions of a simple kind. As regards this issue, I end up on Wittgenstein’s side. Davidson’s arguments against animal thought revolving around holism and the connection between belief and the concept of belief are uncompelling. At the same time, by raising the specter of indeterminacy regarding the content of the intentional states of animals he has posed a formidable challenge to assimilationism.
1 Wittgenstein and Davidson: General Similarities

Several areas invite comparison (see Glock 2014). Leaving aside important disagreements, notably concerning the value of formal logic, there are interesting parallels at the strategic level. In different ways, both are part of the linguistic turn of twentieth century analytic philosophy. They ascribe a central philosophical role to language, albeit for somewhat different reasons. Furthermore, they propound conceptions of language and the mind that shun both Platonism and Cartesianism. Linguistic expressions acquire meaning not by being associated with either abstract entities or private mental processes, but by having a certain role or function in our practices. More generally, in both Davidson and the later Wittgenstein there are important themes that are loosely speaking pragmatist. This includes not just a stress on the philosophical importance of human action, but also more specific ideas: a holistic perspective that views individual expressions, sentences and thoughts as part of a larger context, a context which is ultimately constituted by human activities, and a communitarian tendency to view those activities as inherently social. Finally, both share an anthropological approach to philosophy. Even their discussions of topics apparently unrelated to human affairs – e.g., truth, modality, mathematics, causation – revolve ultimately around a philosophical anthropology, a conception of human behavior and human capacities in general, and of language and linguistic capacities in particular. Turning to our current topic, this anthropological perspective naturally incorporates philosophical comparisons between humans and animals.

2 Animal Minds

Wittgenstein and Davidson developed views on animal minds that have been influential and remain of great significance. In both, moreover, these views are shaped by the fact that they approach mental phenomena from a third-person perspective. They do not appeal to phenomena – whether mental or neurophysiological – that cannot be manifested in behavior or of which we cannot become aware even in principle. In particular, while both acknowledge that mental capacities and their exercises depend causally on neurophysiological events of which we are by-and-large ignorant, they both shun appeal to hidden sub-personal agents or factors of a mental and epistemic kind, such as homunculi or modules. Finally, both hold that one cannot attribute beliefs to creatures that are totally incapable of manifesting these beliefs. Davidson’s rationale is that one cannot make sense of the notion of a belief as a private attitude completely detached from behavior and its explanation (Davidson 1984, 170; 1985a, 475–6). Ascribing beliefs, desires and meanings is essentially part of an interpretative and communicative
enterprise, namely, that of human beings’ understanding one another. The master-problem of Davidson’s work is: What is it to understand other human beings? His solution is provided by a unified theory of meaning and thought that intertwines the philosophy of language, the philosophy of mind, and the theory of action (Davidson 2004). Understanding what a person does, what she believes and desires, and what she means by her utterances are part of a single enterprise.

Wittgenstein’s rationale for insisting that thoughts must be manifestable in behavior is compatible albeit distinct. We can ascribe a thought that \( p \) to a creature \( a \) only if something \( \text{counts as a thinking that} \ p \) rather than that \( q \). For one thing, we identify thoughts/beliefs by identifying their \text{linguistic expressions} (see Wittgenstein 1958, 4–5, 161; 1953, §502; 1997, 108, 237). Although thoughts are not identical with their linguistic expression, they are not entities beyond language either. The answer to the question “What do you think?” is not a description of an inner process, but an expression of my thoughts in words, e.g., “I think that it will rain.” If I am challenged by a mentalist to express the thought “behind” that utterance, I do not re-examine some private process for the sake of describing it more accurately. Nor, if challenged by a Platonist, do I undertake another attempt to latch on to the proper denizen of an abstract realm. Instead, I simply paraphrase my utterance into other symbols. I am not prevented from doing more by an epistemological obstacle. Rather, my opponent’s insistence on something more ultimate evinces a confusion about the very concept of thought.

This third-person perspective contrasts with a less austere approach taken by many philosophers of mind and cognitive scientists, one that posits mental faculties, processes, and sub-personal agencies that are unreflected in behavior and are such that subjects cannot become aware of them even in principle. I shall not attempt to defend the third-person approach against this alternative here. Two points bear stating, however. First, a third-person perspective is adopted not just by common folks, but also by cognitive ethology and comparative psychology in their scientific investigations of animal mentality. Second, such a third-person approach is not tantamount to logical behaviorism (see Glock 2006, 152–9). Thoughts can be ascribed on the basis of behavior, without therefore being reducible to behavioral dispositions, as Davidson (1985a, 476) points out. Furthermore, both Wittgenstein and Davidson reject the idea (occasionally mooted by Ryle) that we ascribe psychological predicates to ourselves on the basis of observing our own behavior. Both respect the first-person authority that attaches to first-person present-tense psychological statements, yet without explaining it as resting on introspection, an infallible perception of one’s own mind (see Glock 1996, 304–9; 2003, 163–7). For Wittgenstein, the ascription of psychological predicates to other people remains \text{conceptually} (or “criterially”) connected to
behavior. That connection is not one of logical equivalence between propositions (namely psychological and behavioral ones), however. Rather, our mental terms would not mean what they do if they were not bound up with some behavioral criteria or other, however diverse, context-dependent, and defeasible. It is part of the concepts of particular mental phenomena that they have a characteristic manifestation in behavior. And it is part of mental concepts in general that they have some such manifestation. We would have no use for these expressions if they did not. Unless certain patterns of behavior – including facial expressions, posture and linguistic utterances – counted as manifestations of pain in a specific context, the term “pain” would not mean what it actually does (Glock 1996, 93–7).

In spite of such overlaps in the philosophy of mind, as regards animal minds, the differences between Wittgenstein and Davidson are more significant than commonly assumed. The main reason is that Wittgenstein is more accommodating to the possibility of animal thought than many of his interpreters allow. Davidson insists that even for simple beliefs, notably beliefs concerning perceptible features of one’s immediate environment, the prerequisite manifestations must include linguistic behavior. Consequently he denies that non-linguistic animals have any thoughts. He has thereby become the leading representative of a position on animal minds that is differentialist, and, more specifically, lingualist. Wittgenstein (by contrast to some of his disciples) adopts a position closer to assimilationism, and to common sense – for once, many will be tempted to say. Animals are capable of having beliefs, desires, and intentions of a simple kind, namely those that can be expressed in non-linguistic behavior. The objective of the next section is to substantiate that exegetical claim.

3 Wittgenstein: A Brute to the Brutes?

Rumor has it that Wittgenstein was a brute to the brutes (e.g., Rollin 1989, 137–41; Singer 1990, 14; Frey 1980, 101–10). He supposedly denied that non-linguistic animals can have thoughts on account of his private language argument. Some animal lovers then consider a caricature of the private language argument – often derived from A. J. Ayer (1954) – which they demolish to their own unbridled satisfaction.

Whatever its merits, however, the private language argument does not imply that animals do not think (see DeGrazia 1994). Its conclusion is that there is no such thing as a language that cannot be communicated or taught to others even in principle. This in turn implies that we cannot credit animals lacking a public language with thinking on the grounds that they possess a private one. However, to reach the further conclusion that such animals cannot think would require the lingualist assumption that thinking requires possession of
a language to begin with. Far from endorsing that assumption Wittgenstein actually casts doubt on it (Glock 2006, 140–7). He criticized – albeit obliquely – his own earlier idea of an inner “language of thought” (Wittgenstein 1958, 34–5; 1974, 144–5). Thinking is a “widely ramified concept.” But it never consists in either words or images crossing one’s mind. Inner goings on – whether pictorial or symbolic – are neither necessary nor sufficient for thought (see Wittgenstein 1953, §§318–21, §343; 1958, 41; 1988, 247–8).

To be sure, Wittgenstein does raise doubts about animal thinking that do not rest immediately on the private language argument.

One does not say of table and chair: ‘it is thinking now’, nor ‘it isn’t thinking now’, nor ‘it never thinks’; nor do we say it of plants or of fish, hardly of dogs; but of humans. And not of all humans. (Wittgenstein 1967, §128)

But this verdict applies to only one of the ramifications of the notion of thinking, namely that of ratiocination or reasoning. For we shall see that Wittgenstein positively grants that animals can think in the sense of believing something to be the case.

‘So these concepts are valid only for the whole human being?’—No: for some of them have application to animals as well. (Wittgenstein 1980, §328)

This passage is pregnant with connotations. On the one hand, it is connected to Wittgenstein’s insistence that mental concepts apply to whole organisms rather than their parts, most famously in Wittgenstein 1953, §281, discussed below. On the other hand, it distances Wittgenstein from blanket differenti- alist denials of animal minds. Unfortunately, the context leaves it entirely unclear what concepts Wittgenstein has in mind. Other passages are more helpful in this regard. Turning first to “lower” mental faculties, Wittgenstein explicitly countenanced the ascription of sensation, perception and concio- usness to animals.

Only of a living human being and what resembles (behaves like) a human being can one say: it has sensations; it sees; is blind; hears; is deaf; is conscious or unconscious. (Wittgenstein 1953, §281, see also §§282–7, §§359–61).

It is clear, moreover, that according to Wittgenstein some animals do behave like a human being in the relevant respects. Again, while he denies that dogs can simulate pain (infra), his treatment of the issue implies that they can be in pain. Indeed, he even countenances the ascription of pain to insects:

Look at a stone and imagine it having sensations.—One says to oneself: How could one so much as get the idea of ascribing a sensation to a thing? One might as well ascribe it to a number!—And now look at a wriggling fly and at once these difficulties vanish and pain seems able to get a foothold here, where before everything was, so to speak, too smooth for it. (Wittgenstein 1953, §284)
Next, Wittgenstein grants that animals have intentions because, like sensations, they can be displayed in non-linguistic behavior.

What is the natural expression of an intention?—Look at a cat when it stalks a bird; or a beast when it wants to escape (Connection with propositions about sensations.) (Wittgenstein 1953, §647)

Commenting on this passage, Anscombe (1957, §2) grants that animals can have intentions, yet insists that their expression is conventional and hence linguistic. She nevertheless protests that one “might as well call a car’s stalling the expression of its being about to stop.” The analogy founders, however, given that by Anscombe’s own lights animals yet presumably not cars can have intentions.

As concerns emotions and beliefs, Wittgenstein’s verdict is both ambivalent and tentative:

One can imagine an animal angry, frightened, unhappy, happy, startled. But hopeful? And why not?

The dog believes his master is at the door. But can he also believe that his master will come the day after tomorrow? And what can he not do here?—How do I do it?—How am I supposed to answer this?

Can only those hope who can talk? Only those who have mastered the use of a language. That is to say, the phenomena of hope are modes of this complicated form of life. (If a concept refers to a character of human handwriting, it has no application to beings that do not write.) (Wittgenstein 1953, 174)

1 Rodriguez (2013) criticizes the interpretation of Wittgenstein as a moderate differentialist put forward by me, among others. His rationale is that there is no expressive behavior common to humans and non-linguistic subjects (children and animals) that could sustain the ascription of the same mental properties even of a simple kind. He aptly summarizes his position as follows: “one, animals are expressive, and, therefore, minded beings; two, animals lack certain mental states and abilities, due to their exclusion from the linguistic, human form of life where those states and abilities have their home; three, there is no set of shared mental states and abilities in animals, children and adult humans, whose expressive nature is independent of linguistic-cum-conceptual abilities” (2013, 114–15). The first two claims are correct, since passages such as these tie some mental capacities to a linguistic form of life. Yet Rodriguez does not adduce a single passage that supports the third claim. His interpretation is instead driven by a “form of life” holism. Since in adult humans all expressive behavior is part of a form of life that includes language, those forms of behavior cannot be equivalent to any expressive behavior in a non-linguistic form of life. But that is simply a holistic fallacy. While it has been committed by members of the Pittsburgh school, there is no evidence that Wittgenstein succumbed to it. One might point to his famous holistic claim that “light gradually dawns over the whole” (Wittgenstein 1969, §141). However, that holism needs to be distinguished from Davidson’s web of belief holism discussed below. And there is no evidence that Wittgenstein used this holism to exclude animal belief, etc. On the contrary, as regards animals he seems to have set store more on the “gradually” than on the “whole.” Concerning simple mental phenomena, there is no categorial divide between us and animals. As Wittgenstein 1953, §650 demonstrates, he held that some expressive behavior is shared between humans and animals. Most importantly, without any qualifications or qualms about univocality, in numerous passages he ascribes mental terms to animals that he also ascribes.
And in a similar vein:

We say the dog is afraid his master will beat him; but not, he is afraid his master will beat him tomorrow. Why not? (Wittgenstein 1953, §650)

A first point to note is that neither passage features standard noun phrases of the form *that* … Should we conclude from this that Wittgenstein rejects animal intentions and beliefs in a demanding sense, one which requires a “propositional content”? In that case his position might not be that remote from Davidson’s (see next section). But the textual evidence does not bear out this interpretation. Not all bona fide ascriptions of intentional states feature the prefix “that.” In English, alternatives include infinitival constructions (“The dog takes his master to be at the door”) and the implicit quotation employed in the English translation of the passages from *Philosophical Investigations*. In German, the alternative is to employ the conjunctive mood in the content sentence, just as in Wittgenstein’s original. If he had intended the conjunctive construction commonly regarded as equivalent to a that-ish construction to signify a lesser, non-propositional kind of believing, he would have needed to make this explicit.

Having addressed this complication, we must distinguish two parameters here: the kind of “propositional attitude” on the one hand, the kind of “propositional content” on the other. Expressing the point in a less presumptuous formal mode: it is one question what intentional verbs can be applied to animals; it is another question what noun-clauses, and in particular what that-clauses or alternative constructions, can follow these intentional verbs.

Wittgenstein reasons that a dog can satisfy some intentional verbs but not others, and that these can be complemented by some noun-clauses yet not others. Let’s attend to the noun-clauses first. Some thoughts cannot be manifested in non-linguistic behavior because they are about things remote in space and time. For example, the dog can manifest its fear of being beaten by cowering with its tail between its legs. But nothing in its behavioral repertoire can be a manifestation of its fear now that it will be beaten tomorrow. To be sure, Eike von Savigny has ingeniously envisaged canines capable of nuanced and complex behavior that might give license to such an ascription (von Savigny 1995, 43–4). However, such ascriptions are at best highly speculative and at worst hopelessly underdetermined. Furthermore, a modified conceptual limitation remains: the sort of behavior that dogs are actually capable of cannot license the ascription of intentional states with complex contents.

to humans. Admittedly, there is an important kernel of truth in the holistic picture, which I try to bring out in the last section. The addition of language makes for an important difference concerning the overall import of even simple mental phenomena. But that is a difference in what it means, e.g., to be in pain or to believe things, i.e., what the consequences are; it is not a difference concerning what “pain” or “believe” mean linguistically in the two cases.
With respect to the intentional verb, there are two possible rationales for Wittgenstein’s restrictive claims. One is what one might call his “form of life contextualism” (the interpretation pursued by von Savigny). A dog can look forward to being taken for a walk by fetching the lead, wagging its tail excitedly, etc. But this expressive behavior is not embedded in the appropriate way in a social environment. The other is a more strictly behavioral argument, which forms part of his “ability approach” to the mind. Thus, nothing distinguishes a dog hoping to be taken for a walk from its being joyously confident of being taken for a walk. In my view, the ability interpretation is superior. For better or worse, Wittgenstein does not allude to the kind of social life an animal is capable of. For instance, the difference between solitary bears and pack hunters like wolves never registers in his reflections. What does register is the kind of expressive behavior, including the kind of facial expressions.

The same two options also apply to another famous passage:

Why can’t a dog simulate pain? Is he too honest? Could one teach a dog to simulate pain? Perhaps it is possible to teach him to howl on particular occasions as if he were in pain, even when he is not. But the surroundings which are necessary for his behavior to be real simulation are missing. (Wittgenstein 1953, §250).

*Prima facie* this supports the contextualist interpretation. But another passage points in the opposite direction.

A child has much to learn before it can pretend (A dog cannot be a hypocrite, but neither can he be sincere). (Wittgenstein 1953, 229)

Here the emphasis is on what the dog is, or rather is not, capable of doing. Similarly in the following passage:

A clever dog might perhaps be taught to give a kind of whine of pain but it would never get as far as conscious imitation. (Wittgenstein 1953, §631)

Once again it is not the social context, but the absence of an ability, namely that of intending to deceive which bars dogs from simulation and hypocrisy. The argument is very similar to the debate in current cognitive ethology about so-called Machiavellian intelligence (see Byrne 1995). Although several primates are capable of deceiving other creatures, the crucial question is whether the deceptions are performed with the *intention* to deceive. Such an intention can be displayed in non-linguistic behavior, e.g., in the hide and peek behavior of chimpanzees. But this requires more complex behavior than dogs are actually capable of.

4 Davidson’s Lingualism

Davidson is the most important contemporary proponent of lingualism. His claim that “a creature cannot have a thought unless it has a language”
(Davidson 1985a, 477; see his 1984, 163) is famous among philosophers, infamous among zoologists and pet owners. He stands firmly within an Aristotelian tradition that conceives of humans as the “rational animal” (Davidson 1980, 223, see also the title of his 1985a). At the same time, his case for lingualism combines several arguments that are not just sophisticated but also original.

Davidson is rightly unimpressed by the charge that to draw any qualitative distinction between humans and animals is deplorably anthropocentric or insufficiently naturalistic. There is no gainsaying the fact that there is both biochemical similarity and evolutionary continuity between us and certain non-linguistic animals. But it does not follow that they must approximate to our mental life. Although it is probable that our closest evolutionary ancestors without language shared many of our other mental capacities, these ancestors are extinct; and there is no guarantee that the biologically closest extant species is mentally close to us. It so happens that the chimpanzees share 98% of our DNA. However, it does not follow that they share 98% of our mental life, simply because small biochemical differences may lead to significant differences in terms of our mental vocabulary. That vocabulary captures neither genetic nor neurophysiological differences but differences in the kinds of capacities we humans are interested in. To that extent, our mental concepts themselves may be anthropocentric; yet it does not follow that it is anthropocentric to insist that these concepts preclude application to non-linguistic creatures (Davidson 1985a, 473). Or, as Wittgenstein points out: “After all, there is nothing astonishing about some concepts only being applicable to a being that, e.g., possesses a language” (Wittgenstein 1967, §520).

The so-called Pittsburgh school derives from Sellars and lays claim to Wittgenstein’s mantle. But it occasionally veers towards an extreme differentialism that is skeptical even about the possibility of animal sensations. Davidson’s differentialism is more moderate, and in this respect once again Aristotelian, in being confined to thought, also known as intentional states.2 His doubts arise out of the connection between intentional states, their contents, concepts and language. Davidson illustrates the difficulty through a by-now classic example of Malcolm’s.

---

2 The evidence for this interpretation is negative or indirect. With respect to other central planks of his philosophy of mind – e.g., anomalous monism and his account of first-person authority – Davidson states explicitly that they are not meant to apply to sensations (Davidson 1970, 221; 1984, 3). Not so for his qualms about animal minds. But his differentialist arguments are confined to intentional states. And he writes: “On the moral issue of how we should treat dumb creatures, I see no reason to be less kind to creatures without thoughts or language than to those with; on the contrary” (Davidson 1985a, 474n). Leaving aside the question of why for Davidson thought makes for a negative rather than positive moral difference, the idea of kindness implies that animals are capable of suffering, and perhaps of certain types of emotions.
Suppose our dog is chasing the neighbor’s cat. The latter runs full tilt toward the oak tree, but suddenly swerves at the last moment and disappears up a nearby maple. The dog doesn’t see this maneuver and on arriving at the oak tree he rears up on his hind feet, paws at the trunk as if trying to scale it, and barks excitedly into the branches above. We who observe this whole episode from a window say, “He thinks that the cat went up that oak tree.” (Malcolm 1973, 13)

Malcolm claims that we would be right to say this, and Davidson acknowledges that it is prima facie plausible. Moreover, he grants that ascribing intentional states to animals is a useful and perhaps even inevitable façon de parler. Nevertheless, he insists, strictly speaking Malcolm’s dog cannot believe anything. In attributing thoughts to animals, we merely treat them as if they were capable of acting for reasons (beliefs and desires), just as one might explain the movements of a heat-seeking missile by ascribing to it the desire to destroy an airplane (Davidson 1985a, 474–8; 1984, 155).

Davidson’s stance is based on a line of reasoning which shapes the general contours of contemporary debates on animal thought, even though it is rarely stated in so many words. This “linguist master-argument” (Glock 2010a) runs as follows:

<table>
<thead>
<tr>
<th>Concept thesis</th>
<th>Thinking (having thoughts) requires concept-possession.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Language thesis</td>
<td>Concept-possession requires language.</td>
</tr>
<tr>
<td>Linguist conclusion</td>
<td>Thinking requires language.</td>
</tr>
<tr>
<td>Dumbness thesis</td>
<td>Animals lack language.</td>
</tr>
<tr>
<td>Differentialist conclusion</td>
<td>Animals cannot think.</td>
</tr>
</tbody>
</table>

The argument is valid. But is it sound? That depends on three questions. First, if animals cannot have concepts, does that preclude them from having thoughts (beliefs, desires intentions, etc.)? Second, can non-linguistic creatures possess concepts? Third, are there animals with (limited) linguistic powers? Neither of our two protagonists has much to say about that third question; I shall leave it aside as well and assume in the sequel that animals lack language.

5 Thoughts and Concepts

One will have to answer the first question in the affirmative if one accepts the idea that thoughts are mental occurrences or abstract entities that have

3 To be sure, there is Wittgenstein’s puzzling remark “If a lion could talk, we could not understand him” (Wittgenstein 1953, 223; see also §25). On one reading this means that we could not understand a lion who utters English sentences like “I’m not interested in you, I just had an antelope,” which is obviously false (although one might, following Austin, question whether such a talkative creature could count as a lion). On a charitable reading, it means that if lions had a feline language of complex growls, roars, etc., we could never come to learn it. Why? Because their form of life and their behavioral repertoire are so alien to us. We could not make head or tail of their facial expressions, gestures, and demeanour. Moreover, our ability to interact even with a tame lion is strictly limited.
concepts as their components, because that implies that one cannot have or grasp the thought without having or grasping its constituent concepts. However, Davidson rejects the idea that propositional attitudes are relations between a subject and an abstract or mental phenomenon, and regards them simply as modifications of a person (Davidson 1994, 232). He is right to do so. Although that-clauses can grammatically function as accusatives, they no more refer to an object than noun-phrases like “everything” or “the past” (Glock 1997).

If this is correct, for a to believe that p, a need not stand in a relation to an object (a proposition) which would involve standing in a relation to components of that object (concepts). This conclusion is compatible with Davidson’s insistence that a creature cannot be credited with a belief unless it can be credited with the constituent concepts. But it blocks the obvious argument for the concept thesis, namely that concepts are the building blocks of thoughts. Davidson’s approach to human belief is “holophrastic”: we ascribe beliefs to linguistic creatures on the basis of their asssenting to sentences as a whole (Davidson 1984, 4, 22, 220–5; 1997, 25). The possibility he ignores is that rejecting the building-block picture in the linguistic case invites an analogous move in the case of beliefs, a holodoxastic approach that starts out from the whole belief and imputes a structure – whether conceptual or not – on that basis.

According to such an approach, what matters is precisely a kind of “modification” – if a creature can be correct or mistaken as to how things are, it can have beliefs. Although the sentences we use in ascribing thoughts have components, our ascriptions are not based on a prior ascription of these components. Instead, they are based on the subject manifesting certain perceptual capacities, attitudes, and emotions. In animals and pre-linguistic children, these manifestations will obviously not include assent to sentences. But they will include forms of behavior, postures, and facial expressions that higher animals share with us. When we say that Malcolm’s dog believes that the cat went up that oak tree, we do not do so on the grounds that it picks out objects and classifies them in a way that corresponds to the singular and general terms we use in the attribution. Rather, we simply note the dog’s reaction to its environment. We regard these reactions as directed towards particular objects, creatures and events, because we assume that dogs have certain perceptual capacities and wants, assumptions which require rudimentary knowledge of the way dogs live – what they can discriminate and recognize, what they tend to dislike, etc. What holodoxastic belief requires is not possession of concepts in any weighty sense (see below), but only the ability to identify distinct objects and to discriminate them as being of different kinds. Because of its reliance on behavioral reactions, the holodoxastic move is confined to simple beliefs, notably
about perceptible features of the subject’s environment. But it suffices to blunt the force of the line “No thoughts without concepts!”

A difficulty remains nonetheless. Granting thoughts to animals while denying them concepts suggests that there is an incongruity between ascribing thoughts to animals and ascribing thoughts to linguistic creatures. In the second case, our ascriptions impute to the believer a grasp of the concepts involved, whereas in the first they do not. This creates a pressure for holding that intentional verbs such as “believes” are ambiguous, referring either to a holodoxastic, behavioral phenomenon or to a conceptual, linguistic one.

However, we can grant that there are important differences between the beliefs of conceptual and non-conceptual creatures, yet resist the pressure towards postulating distinct objects and hence distinct attitudes. A certain disparity between the terms used in a belief report and those that could be used by the subject is present even in the linguistic case, without constituting a fundamental incongruity. The terms that occur in the content-clause are in general dictated not so much by the creature whose belief we report, but by the concerns of speaker and audience. Thus, “Sarah thinks that the charlatan you introduced me to is about to give her a biscuit” can be in order, whether Sarah is an adult, a child that lacks the concept of a charlatan, or a dog (Rundle 1997, 83). Consequently, it is far from obvious that attribution of beliefs requires attribution of concepts, especially if one follows Davidson in denying that to have a belief is to stand in a relation to an object.

6 Belief, Truth, and Triangulation

This anti-representationalist rebuttal of the concept thesis implies that animals can indeed believe or know things, if only holodoxastically. Precisely that is contested by other considerations in favor of the concept thesis, considerations that rely neither on the building-block model of contents nor on the relational conception of intentional states. Davidson has developed an argument that does not invoke concepts as constituents of individual propositional contents; instead, it makes the general possibility of believing that \( p \) dependent on grasp of two concepts that need not occur in \( p \), namely truth and falsehood. According to him, \( a \) can only have beliefs if \( a \) has mastered, if only inarticulately, the concept of objective truth.

Davidson starts out from the observation that it is essential to beliefs that they can be true or false. From this he infers that \( a \) “cannot have a belief unless he understands the possibility of being mistaken, and this requires grasping the contrast between truth and error – true belief and false belief” (Davidson 1984, 170).

Davidson goes on to make this grasp dependent on linguistic communication – in line with the language thesis. \( a \) can only believe that \( p \) if \( a \) is capable of
recognizing that a’s own belief that p may be at odds with the objective facts. And this in turn is supposed to be possible only through a process of “triangulation” that is analogous to the measuring of spatial distances. a must be capable of triangulating in the sense of linguistically communicating with another subject b about the world, more specifically, about whether or not p (see Davidson 1985a, 479–80; 1997, 26–7).

Davidson’s idea of triangulation is both ingenious and fertile. There are interesting parallels to cutting-edge ethology and developmental psychology, notably research into joint attention by the Leipzig group (e.g., Tomasello 2014). Nevertheless, it does not furnish an a priori conceptual argument in support of the linguist conclusion. To begin, one can turn the measuring analogy against Davidson. In the case of establishing spatial distances, there are several alternatives (more or less technologically sophisticated) to triangulation. By a similar token, why shouldn’t a be able to correct her beliefs by adopting a new perspective herself, rather than by conversing with another subject? Unless this prima facie unproblematic option can be ruled out, the idea of triangulation fails to support the language thesis for the concepts of truth and falsehood.

There is an avenue for meeting this challenge, which is reminiscent of Wittgenstein’s private language argument.4 If a changes her perspective and then reaches a different conclusion about the situation, what distinguishes her correcting an objective error from the situation having changed? One prima facie plausible option is the communication with b, who has held the object fast in sight without altering her perspective. But is that the only possibility? Can’t a keep the object in sight to a degree sufficient for ruling out pertinent changes? If so, triangular communication is not necessary for a grasp of the idea of objective truth. Moreover, unless a can trust her own individual ability to gage the objective situation, how could she rely on the more presumptuous and precarious process of communication with b to undergird her judgment? If so, triangular communication is not sufficient for the idea of objective truth.5

Even if the triangulation argument fails, it remains plausible on independent grounds to tie grasp of logico-semantic concepts like truth to linguistic abilities. However, does one need to master the conceptual pair true and false in order to

4 As reconstructed in Glock 1996, 309–15. This rejoinder to an objection raised in Glock 2000, 57–60 occurred to me on reading Tomasello’s remark: “If I see something in one way, and then round the corner to see it in another, this does not give me two perspectives on the same thing, as I do not have multiple perspectives available to me simultaneously for comparison. But when two people are attending to the same thing simultaneously – and it is in their common ground that they are both doing so – then ‘space is created’ (to use Davidson’s, 2001, metaphor) for an understanding of different perspectives to arise” (Tomasello 2014, 44–5, my emphasis). But the diagnosis of a connection to the private language-argument and the elaboration of the comeback are mine.

5 For a more positive assessment of the triangulation argument, see Verheggen 2013.
have beliefs to begin with? A first cavil is that a might possess a non-conceptual understanding of the difference between true and false beliefs. One could further insist that some primates demonstrate an understanding of that difference – whether conceptual or non-conceptual. Thus monkeys are credited with recognizing mistaken beliefs of their own in certain wager games (Kornell et al. 2007; Esken 2012). Similarly, the tactical deceptions reported, e.g., for baboons and chimpanzees (Byrne 1995, Chapter. 8–9) provide grounds for holding that they can recognize mistaken beliefs in others. These findings and anecdotes may be exaggerated or allow of alternative explanations; yet the described patterns of interaction are perfectly conceivable among non-linguistic creatures.

In any case, Davidson’s case for lingualism is unpersuasive, since his argument for the concept thesis concerning truth/falsehood founders. The fact that a cannot grasp or recognize the contrast between two properties F and G does not entail that a cannot possess either F or G. a is not immune to developing a malignant tumor simply because a is ignorant of the difference between malignant and benign tumors. To be sure, Davidson tries to demonstrate that in the case of the contrasts knowing/(merely) believing and true/false, there are specific reasons why a can instantiate them only if a has mastered these concepts. He reasons as follows:

(i) A belief is something that “can be true or false” (Davidson 2001, 104);
(ii) aBp ⇒ a can be mistaken in believing that p;
(iii) a can be mistaken in believing that p ⇒ a can recognize that a is mistaken in believing that p;
(iv) a can recognize that a is mistaken in believing that p ⇒ a has the concept of a mistake.

If one thinks of self-regarding meta-cognition in animals as both real and conceptual, one can accept these steps without committing to the lingualist conclusion. Alternatively, one can repudiate (iv) by appealing to the option of non-conceptual meta-cognition. However, (ii) is already contentious, since it excludes belief in necessary propositions and bars an omniscient being from believing things. Admittedly, it would be acceptable if in these two cases, at least, knowledge excluded rather than implied belief. But one need not subscribe to the moribund tripartite conception of knowledge as true justified belief to reject that defense. On the other hand, in considering animal mentality we are entitled to disregard the case of necessary propositions and the dubious notion of omniscience. Even then, however, (iii) goes astray, in so far as it makes the possibility of believing erroneously dependent on second-order beliefs to the effect that a previously held belief is mistaken. A “first-order” change can lead a from falsely believing that p to truly believing that not-p. That a is genuinely erring and recognizing its mistake is manifest in non-linguistic reactions such as surprise.
Davidson recognizes the importance of surprise. Indeed, he regards the possibility of surprise as essential for having beliefs. But as Sorabji (1993, 43) remarks, the top flight of British diplomats are trained to avoid being surprised, surprise being an inefficient reaction surplus to requirements. And there could be fully linguistic humans who lack the terminology and/or facial expressions and demeanors of surprise altogether.

Davidson insists that surprise requires second-order correction of belief (Davidson 1985a, 479). In ordinary parlance, however, “surprise” is the name for a particular kind of reaction to things being otherwise than one expected or believed (as well as for facts or events which elicit that reaction). Such reactions are displayed not just linguistically, but in forms of behavior and facial expressions which are not the prerogative of language-users. We would not hesitate to speak of a chimpanzee as being surprised by discovering that what looked like a banana is merely a decoy, provided its behavior or facial expressions show signs of disorientation or disappointment (DeGrazia 1996, 148–9). What is more, by recording looking-time and attention-span the “habituation-paradigm” may allow cognitive ethologists and developmental psychologists to operationalize the idea of attention and surprise for non-linguistic subjects (for Wittgensteinian reservations on this score see Tissaw 2013).

Davidson, by contrast, distinguishes between being startled and being surprised. Perhaps he has in mind the difference between being surprised by an object or event and being surprised that things are thus-and-so. In any event, he conceives of surprise independently of behavioral manifestations and as involving beliefs about beliefs by definition. Given this stipulation, the soundness of (iii) hinges straightforwardly on whether a creature can only have beliefs if it is capable of also believing that a prior belief was false.

One should grant that a can believe that p only if it is also capable of believing something that is incompatible with that belief, notably that not-p. But Davidson has no argument to rule out the possibility of a simply switching from a belief that p to a belief that q, without that switch involving a believing that its original belief was false. a can manifest, recognize, and correct error not just through referring to its own prior beliefs, but also through correcting its behavior – notably on account of the deliverances of its senses – by pursuing a persistent goal in a more apposite manner. For example, Malcolm’s dog first believes that the cat is in the oak tree, as witnessed by its barking up the oak, and then that it went up the maple tree, as witnessed by its barking up the maple (see Glock 2000, 54–6; Hacker 2007, 237 n).

7 Animal Concepts

Regarding the language thesis, one finds the same spectrum of opinion as regarding animal minds in general. Davidson, together with Frege and
Dummett, occupies one end. According to them, while non-linguistic subjects may be able to perceive, they lack concepts of any kind. At the other end are cognitivists, who have no qualms about ascribing complex concepts to animals. An intermediate position is occupied, for example, by Kenny, who maintains that animals can possess some concepts, namely those that can be manifested in non-linguistic behavior (Kenny 1989, 36–7; also DeGrazia 1996, 155–6).

Whether this position is defensible naturally depends on what one makes of concepts and concept-possession (see Glock 2010b). One construal is that concepts are principles of discrimination and that to possess a concept is to have the ability to recognize or discriminate different types of things (e.g., Price 1953, 355). On that account, animals certainly possess concepts: it is evident from their behavior – in the wild and in the laboratory – that they can distinguish between a host of different colors, tastes, sounds, shapes, stuffs, types of creatures, etc. Moreover, many of these capacities are learned rather than innate. Davidson admits this, but nevertheless resists the conclusion that animals can have concepts (Davidson 1985a, 480). He adduces several arguments to this effect.

The first is a reductio ad absurdum. “Unless we want to attribute concepts to butterflies and olive trees, we should not count mere ability to discriminate between red and green or moist and dry as having a concept, not even if such selective behavior is learned” (Davidson 1997, 25). Leaving aside butterflies for the moment, I agree that it would be absurd to credit olive trees with concepts. But this absurdity does not follow from treating concepts as powers of discrimination. Olive trees do not discriminate between moist and dry soil, since discrimination is a prerogative of sentient creatures, that is, animals.

We must distinguish between mere differential reaction to causal inputs, which is a universal feature of physical phenomena, and discrimination, which is tied to creatures with perceptual capacities.

Davidson’s second argument is that there is a fundamental difference between classification and discrimination: whereas the former is required for concept-possession, only the latter is available to non-linguistic creatures. “To have a concept is to classify objects or properties or events or situations,” or, more accurately, to be able to do so. Powers of discrimination, by contrast, are mere “dispositions,” and therefore, “as Wittgenstein emphasized, have no normative force.” Such dispositions do not involve the ability to recognize.

---

6 Once again Davidson is in good, or at any rate Wittgensteinian, company, namely Geach (1957, 16–17).
7 Biologists speak, for example, of the immune system as discriminating between different antigens. But discrimination in this sense is a functional notion distinct both from perceptual discrimination by sentient creatures and from mere differential reaction (as in the case of Davidson’s olive tree). It is the biological role of the immune system to react differentially, whereas it is a biological liability of the olive tree to require a certain kind of soil.
a mistake, and hence involve no knowledge of the difference between correct and incorrect behavior (Davidson 1997, 24–5; see his 1985a, 480).

Davidson is right to hold that there is a type of classification that differs from mere discrimination in its normative dimension. He is also right to suggest that it is the absence of such classification that makes us reluctant to credit butterflies with concepts. Again, he is right to maintain that the normativity required for classification presupposes that the classifier can make a mistake that she is capable of recognizing. Finally, he is right to detect a proximity to Wittgenstein here. Going beyond Davidson’s hint, it concerns the latter’s distinction between the mere disposition to behave in accordance with a rule – as butterflies do when they land only on red petals – and of genuinely following a rule.8

Unfortunately, the precise nature of this normativity is far from straightforward. It is tempting, and perhaps in the spirit of both Wittgenstein and Davidson, to tie it to intentional behavior. Concept-possession may appear as a “two-way power” in Kenny’s terminology. Unlike “one-way” or “natural powers,” i.e., mere dispositions, these two-way powers are not automatically exercised given certain antecedent conditions. Rather, their possessors can exercise them or refrain from exercising them at will. Greenhouse gases have the disposition to trap light reflected from the surface, and will inevitably do so given certain conditions. By contrast, I can choose whether or not to exercise my ability to cycle to work.

But now, concept-possession – like linguistic understanding – is not a two-way power. Its exercise is not (uniformly) subject to the will. I might be able to decide not to employ certain concepts actively in a complex train of thought. As Kenny remarks, however: “Looking up at the flashing lights of the advertisements in Piccadilly Circus, one cannot prevent oneself from understanding their message. (How much more beautiful they would be, G. K. Chesterton once remarked, if only one could not read!”) (Kenny 1989, 22).

Similarly, while we can decide whether or not to take a look, once we do look, we cannot decide whether or not to see something within our field of vision. And as conceptually gifted creatures we cannot help conceptualizing what we see, in the sense of recognizing it as being of a certain kind. At most, we can try to reconceptualize what we see by thinking of less obvious categories that also apply to it.

Nevertheless, the Wittgenstein-Davidson line retains a kernel of truth: attributions of concepts possess a normative force lacking in the attribution of mere dispositions to discriminate. A subject that fails to discriminate in accordance with a disposition is not subject to correction. By contrast, concept-possession, like some other intellectual abilities, occupies a halfway

8 For the distinction see Baker and Hacker (1985, 154–8). It is used to explain the difference between discrimination and classification in Glock (2000, 44–9).
position. Unlike two-way powers it is not subject to the will; yet unlike one-way powers its exercise allows of normative assessment. Crucially, that assessment must at least include reference to the subject’s own perspective. The subject must be capable of committing a mistake that it can recognize and correct, at least in principle. Otherwise, a is merely diverging from our expectations or from a statistical norm. As Davidson points out, a slippery road may be a danger or a nuisance, but it does not commit a mistake. *Mutatis mutandis*, a butterfly that fails to discriminate between red and green, can commit mistakes of a type not requiring a capacity for correction, e.g., reducing its biological fitness. Yet without it, such failures are not misapplications of *its own standards*.9

One moral of these reflections: conceptual classifications are susceptible to justification, criticism, and correction. Obviously, non-linguistic creatures cannot justify or criticize discriminations either to themselves or to others. But that is no bar to their *being justified* in their discriminations and to their *correcting* themselves and others. In fact, animals and pre-linguistic infants capable of learning and/or teaching are an obvious case in point. This is where Davidson’s third argument comes in. He suggests that animals and pre-linguistic children cannot be genuinely taught, but only conditioned. Because they are sentient, their behavior can be altered by means of inflicting pleasure or pain.

But the point remains: we improve the road, from our point of view, by spreading sand or salt; we improve the child, from our point of view, by causing pleasure or pain. In neither case does this process, by itself, teach road or child the distinction between correct and incorrect behavior. To correct behavior is not, in itself, to teach *that* the behavior is incorrect. Toilet training a child is like fixing a bathtub so it will not overflow; neither apparatus nor organism masters a concept in the process. (Davidson 1997, 25)

This verdict is in line with Wittgenstein’s well-known distinction between teaching and training (Wittgenstein 1953, §§5–6), and with pronouncements by McDowell (e.g., in his 1994, 115). But neither Davidson nor McDowell explains why it should follow a priori from a subject’s lack of language that its behavior is purely mechanical.

Furthermore, Davidson’s characterization is unduly behaviorist. The acts and omissions of non-linguistic creatures are not always explicable solely by reference to immediate biological stimuli and imperatives. Both pre-linguistic children and the great apes can refrain from a particular action, either by pursuing their goals in a different way or by forsaking them, at least temporarily.10 By the same token, in a particular situation they are capable of

---

9 The point needs to be labored against teleological theories of content, but not here.

10 This will be obvious to the parents of any two-year old. As regards animals, see Goodall 1986; Menzel 1974; Byrne 1995, 150, 187–9, 225, Chapter 7; Dupré 1996, Tomasello and Call 1997, 7–12; White 2007, 85–91.
either heeding or disregarding a difference. Unlike bathtubs, some animals are capable of distinguishing objects of types \( F \) and \( G \) in one situation, and of ignoring the difference in another. It would be wrong to ascribe such classification to butterflies. But it seems equally wrong to deny classification to chimpanzees capable of selecting or making tools in advance of attempting a task. For, they deliberately distinguish between different kinds of objects (e.g., leaves and blades of grass) in some situations, but may disregard the difference in other situations, or if they are not in the mood.

On this view, non-linguistic concepts may be the prerogative of infants, great apes, and dolphins. Nonetheless, concept possession depends not on language, but on discriminatory behavior that is sufficiently complex, flexible and deliberate (see Glock 2010a).

8 Holism

There is an even more demanding conception, according to which classification as just described does not suffice for concept-possession. \( a \) only masters a concept if \( a \) can draw inferences from the fact that \( x \) is \( F \). At this juncture, the language thesis can be backed up by a further consideration. It concerns “the intrinsically holistic character of the propositional attitudes,” the alleged fact that “to have one is to have a full complement” (Davidson 1985a, 473; similarly Stich 1979). Since at least some members of that complement are definitely beyond the ken of animals, they lack even the simple beliefs commonly ascribed to them. According to Davidson, Malcolm’s dog cannot believe of an object that it is a tree, “unless we suppose the dog has many general beliefs about trees: that they are growing things, that they need soil and water, that they have leaves or needles, that they burn. There is no fixed list of things someone with the concept of a tree must believe, but without many general beliefs there would be no reason to identify a belief as a belief about a tree, much less an oak tree” (Davidson 1985a, 475).

For one thing, Davidson contends that specific concepts that occur in our attributions even of simple thoughts presuppose general beliefs, with which animals cannot be credited. But his examples are far from compelling. He insists that \( a \) can only believe that the cat went up the oak or that the sun is behind clouds if \( a \) also knows that trees burn and that clouds consist of water vapor. But this would restrict the possession of many beliefs to moderately educated contemporaries. Furthermore, it threatens to entail that any alteration in general beliefs amounts to a conceptual change, with the consequence that two scientific theories featuring apparently incompatible empirical claims cannot be talking about the same phenomena (Fodor and LePore 1992, ch. 1). As Davidson candidly admits, his holism implies that the Ptolemeans could not believe that the earth is flat, since this would
amount to rejecting a belief that he treats as constitutive of our concept of the earth.

For another, like other proponents of the language thesis, Davidson implicitly relies on general holistic principles. But these are threatened by a dilemma: they are either too strong, since they would also exclude plausible cases of human thought, or too weak, since they cannot rule out all types of animal thought. The strongest holistic principle lingualists could invoke runs:

\[(A)\quad (aBp \land (p \Rightarrow q)) \Rightarrow aBq.\]

(A) is excessively restrictive. Human beings can believe, for example, the axioms of Euclidean geometry without believing all the theorems entailed by them. According to a modally mitigated version, a only needs to be capable of believing (learning, understanding) the consequences of its beliefs:

\[(B)\quad (aBp \land (p \Rightarrow q)) \Rightarrow \Diamond aBp.\]

Even (B) is too demanding, however. For it is possible to believe the Euclidean axioms without even being capable of understanding all of the theorems (some of which are highly complex). It is more plausible to maintain that a need be capable of appreciating only some rather than all of the things entailed by p. Schematically, this would amount to:

\[(C)\quad aBp \Rightarrow \exists q(q \neq p \land (p \Rightarrow q)) \land \Diamond aBq.\]

According to this principle, if a human being is incapable of even grasping any of the theorems entailed by Euclid’s axioms, his beliefs in the axioms do not have the same content as the belief of someone capable of understanding these theorems; and to that extent, the two have different beliefs.

The problem for lingualism is that (C) only offers moderate support to the language thesis. For prima facie there are animals capable of appreciating some consequences of simple perceptual beliefs. This holds especially of consequences that satisfy two conditions. First, we can formalize them in the propositional calculus, without imputing subject/predicate or quantificational structure. Second, they are of an immediate kind. For example, by consistently barking up the oak and completely ignoring the pine even when prompted by us to do otherwise, Malcolm’s dog could display the belief that the cat is not on the pine tree. Even non-linguistic creatures can in principle be guided not just by what they perceive, but also by what follows from what they perceive.

\[11\text{ I leave aside the question of whether there might be quantitative measures of more or less immediate consequences through a proximity ordering.}\]
On the other hand, the failure of this lingualist reasoning does not imply that a creature could entertain just a single belief. Indeed, the complex and flexible demeanor required for conceptual belief appears incompatible with a behavioral repertoire capable of exhibiting just a single belief. These reflections do not establish, however, that the web of which any belief must be part need extend as far as the web of sophisticated human thought. There may be larger and smaller webs. What kind of network is required may depend on the belief and the creature concerned. From the fact that an animal lacks our web of beliefs and our concepts, it does not follow that it has no beliefs and no concepts (see also Bekoff and Jamieson 1991, 19–20; Dupré 1996, 332; DeGrazia 1996, 154–8; Allen 1999, 39).

As a last resort, a holistic defender of the language thesis can raise the bar for concept-possession even higher. It does not suffice for a to be capable of entertaining some consequences of a purported belief, a must also be capable of believing that the latter entails the former. That is to say:

\[ (D) \, aBp \Rightarrow \exists q (q \neq p \& (p \Rightarrow q) \& \Box aB(p \Rightarrow q)). \]

According to (D), a can only possess concepts and hence beliefs if a can draw inferences. But leaving aside the question of whether highly intelligent animals can draw disjunctive and transitive inferences (see Andrews 2015, 96–105), the question is whether this capacity is a prerequisite for having concepts understood as principles of classification rather than as rules of inference.

9 The Specter of Indeterminacy

Even if holistic objections to attributing thoughts to animals fail, one related difficulty remains: without verbal manifestations we cannot draw the fine distinctions between different thoughts (beliefs, desires) expressed in the same non-verbal behavior. Thought-attributions to humans create “intensional” contexts: if we substitute co-referential terms within the content-clause, this may lead from a true attribution (e.g., “Sarah believes that Cicero was Roman”) to a false one (e.g., “Sarah believes that Tully was Roman”). In the case of animals, by contrast, substitution of co-referential expressions often leads from attributions that we commonly regard as true to attributions which are absurd or unintelligible. The oak tree that the cat went up happens to be the oldest tree in sight and the same tree the cat went up last time the dog chased it. But does Malcolm’s dog believe that the cat went up the oldest tree in sight, or the one it went up last time? Equally, a dog can know that its master is at the door. But does it also know that the president of the bank is at the door? We have no clue how to settle or even understand such questions (Davidson 1984, 163; Chater...
and Heyes 1994). The reason is that the dog can think neither that its master is the president, nor that he is not.

One response to this failure of intensionality is to hold that in the sentence (1) The dog thinks that the cat went up that oak tree
the expression “that oak tree” occurs transparently (in Quine’s terminology). Accordingly, (1) is paraphrased so as to avoid problems of intensionality, through a ‘de re’ description:
(2) The dog thinks, with respect to that oak tree, that the cat went up it.

But this response presupposes that “the de re description picks out an object the believer could somehow pick out” (Davidson 1985a, 474–5). If Malcolm’s dog were incapable of distinguishing the oak tree from among other objects (e.g., the pine tree or the garden fence), we might nevertheless causally explain his behavior by reference to the oak tree, just as we explain the convulsions of an oyster by reference to its being pricked with a needle. But (2) would no longer be appropriate. For de re constructions like “with respect to” or “of” require an anaphoric referent in the subsequent content-clause, an “it” which a disabled dog could not distinguish from other things.

However, this requirement might be met by non-linguistic creatures on account of their powers of discrimination. Non-linguistic subjects cannot know objects by different descriptions, but they might recognize them by different features, and they can fail to recognize that these are features of one and the same object. Thus Malcolm’s dog might react in one way to a man with heavy footsteps being at the door, in another to his master being at the door, because he has failed to realize that his master has adopted a heavy foot-step. In that case he can believe that the heavy stepper is at the door, but not that his master is at the door, or vice versa.

For all that, attributions of intentional states to animals remain under-determined, as Stich (1979) has argued by elaborating on Davidson. In their case we cannot diagnose the fine-grained differences that only manifest themselves through the subject’s linguistic utterances. Notwithstanding this noteworthy restriction, however, at this juncture we should heed Wittgenstein’s insight that determinacy is less pronounced in and less essential to human thought and speech than Fregean ideals and formal semantics allow. It makes sense to ascribe to animals intentional states of a simple kind, namely those that can be identified on the basis of non-linguistic behavior. This verdict applies to both parameters of intentional states, the kind of intentional state (believing, desiring, etc.) and the kind of “content.” Thus, one may grant that a dog can know, believe or see that p, but deny that it can hope that p. Similarly, a dog can believe that its master is at the door, but not that its master will return the day after tomorrow (Wittgenstein 1967, 174).
10 Conclusion

Where does all this leave the lingualist thesis that the capacity for thought requires the capacity for language? According to Davidson, intensionality, concepts and holism point in the direction of language, but they do not amount to a demonstration that language is necessary to thought. Indeed, what these considerations suggest is only that there probably can’t be much thought without language. (Davidson 1985a, 477)

This verdict is puzzling. Davidson’s radical holism according to which to have one thought is “to have a full complement” (Davidson 1985a, 473) is incompatible with the conclusion of merely restricting the scope of animal thoughts in the direction of which it is supposed to point. By contrast, a modest holism such as the one encapsulated in (C) favors such a restriction, since it allows that thoughts come in larger or smaller packages. And packages that include beliefs manifestable only in linguistic behavior are – following Wittgenstein and Davidson – the preserve of language-users.

If this is on the right track, the problem with the arguments concerning intensionality, concepts and holism is not that they are uncompelling, but that all they compel us to is confining animal thought to simple cases. On one occasion, Davidson himself accepts this conclusion. “The fewer acceptable transformations [of thought attributions] the less thought” (Davidson 1985b, 252). To that extent, our discussion reinforces the intermediate position. In fact, it lends support to a heretical idea. Pet lovers and zoologists attribute a greater variety of thoughts to animals not because they suffer from anthropomorphism, as Davidson insinuates (Davidson 1985a, 474n1; 1984, 164), but because they are better acquainted with their behavior and perceptual capacities.

At the same time, Davidson’s reflections point to ways in which attributions of even simple thoughts to animals are not mere extrapolations from the human case. Animals do not just have thoughts of a simpler kind; their having these thoughts amounts to something simpler, because it is part of a smaller logical-cum-conceptual space. In the case of animals, there is at most an analogue of intensionality. In so far as thought-ascriptions to animals are holodoxastic, they are also restricted to thoughts about perceptible features of the environment. Furthermore, they lack conceptual connections which apply in the human case: we cannot infer from the fact that the dog thinks that $x$ is $F$ that the dog grasps the concept $F$. Furthermore, even if animals can have concepts, these are not just confined to concepts of a (roughly speaking) perceptual kind. Animals are also incapable of satisfying one of the two criteria we standardly use in attributing concepts to humans. They may be able to apply principles of classification, but not to explain them. In fact, the two restrictions are linked. A chimpanzee may discriminate between its keeper and other humans just as deliberately as it does
between red and black ants. But we are more inclined to ascribe to it the concept of redness than the concept of a keeper, because there is so much more to explain with respect to the latter. Finally, animals cannot avow thoughts, specifying what precisely they are thinking. This makes for an important parallel between Davidson’s rejection of animal thought and Wittgenstein’s more tentative and qualified verdict:

So one might distinguish two chimpanzees with respect to their mode of operation, and say of the one, he is thinking, and of the other, he is not thinking. But here to be sure we would not have the complete employment of ‘think’. The word would refer to a mode of behavior. It only acquires the meaning of a mental activity through its special employment in the first person (Wittgenstein 1980, §§229–30).

Attributions of simple thoughts to complex animals are intelligible and often warranted. But they are neither intensional nor conceptual nor holistic in the way attributions of the same thoughts to linguistic subjects are. And they do not specify what went through the subject’s head, not even in the Wittgensteinian sense of specifying thoughts the subject itself would be capable of avowing. In short, they are part of a more “primitive” language-game played with mental terms (Wittgenstein 1992, 39).

The best analogy for this relative impoverishment is not Davidson’s anthropomorphic explanation of missiles, but one he has suggested in discussion. Attributing thoughts to animals is like using numerals for the purpose of labeling members of a football team. Although natural numbers stand in complex relations of order and numerical difference, these relations are ignored in this context. What matters here is not the numerical difference between two numbers, nor even which one is greater, but only that no two numbers should be used for the same player.

The analogy is illuminating. Thought attributions to animals employ a rich conceptual apparatus to an area in which many of the logical connections that constitute that apparatus do not apply. But it breaks down in one important respect. Attributing thoughts to animals is not simply an impoverished application of a rich technique. For that richer technique evolves around a central core of cases in which creatures believe, know, or desire things on account of their wants and perceptual capacities. These biological basics of belief are shared by humans and animals. At the same time, when we move from this core area in the direction of conceptual thought, we also move in the direction of linguistic thought. The features non-linguistic creatures must possess to be capable of conceptual thought – intentionality, complexity, flexibility – correspond to those features by which theorists from Descartes through Wittgenstein to Chomsky have distinguished language from more basic systems of communication. In this respect, at least, the reflections of Wittgenstein and Davidson confirm rather than negate the connection between thought and language.
References


‘The question can be raised: Is a state that I recognize on the basis of someone’s utterances really the same as the state he does not recognize this way?’ (Wittgenstein 1992, 8–9)

‘If the mental states of others are known only through their behavioural and other outward manifestations, while this is not true of our own mental states, why should we think our own mental states are anything like those of others?’ (Davidson 1991, 207)

1 Introduction

There are striking parallels between Davidson’s conception of the mental and Wittgenstein’s. At the most general level, Davidson’s pronouncement that ‘the mental is not an ontological but a conceptual category’ (Davidson 1987, 114) echoes Wittgenstein’s statement that ‘“mental”… is not a metaphysical, but a logical, epithet’ (Wittgenstein 1992, 63). More specifically, Davidson’s idea that mental concepts belong to a sui generis system of description and explanation, which is governed by norms distinct from those that govern the physical scheme of description and explanation (see Davidson 1970), echoes Wittgenstein’s idea that the language-game of applying mental terms to ourselves and others is ‘autonomous’ (Wittgenstein 1992, 40), and that ‘the inner differs from the outer in its logic’ (Wittgenstein 1992, 62). And Davidson’s insistence that ‘there are no rules’ for correctly interpreting a person’s words and thoughts, ‘no rules in any strict sense, as opposed to rough maxims and methodological generalities’ (Davidson 1986, 107), echoes Wittgenstein’s idea that there are no ‘exact rules of evidence’ for ascribing inner states (Wittgenstein 1992, 94): that such rules as there are ‘do not form a system… Unlike calculating rules’ (Wittgenstein 1953, part II xi §355 [p. 227]).

There are many other points of contact between Davidson’s views and Wittgenstein’s.

1 References to passages in Philosophical Investigations Part II are given in two forms: by the § numbers introduced in the 4th edition (in which Part II appears as ‘Philosophy of Psychology':
Despite the deep similarities, however, there is little direct discussion of Wittgenstein in Davidson’s writings. But one point at which Davidson does explicitly engage with Wittgenstein comes in his writings on first-person authority. And in this case, his tone is critical; he presses an objection to a view about our knowledge of our own and others’ minds that he associates with Wittgenstein and Strawson. The Wittgenstein/Strawson view, as Davidson presents it,

tries to relieve worries about ‘our knowledge of other minds’ by remarking that it is an essential aspect of our use of certain mental predicates that we apply them to others on the basis of behavioural evidence but to ourselves without benefit of such aid. (Davidson 1987, 16)

Davidson agrees that this is an essential feature of mental predicates. But, he complains, if that is all we say about the relation between first-person and third-person uses of mental terms, we invite scepticism about whether mental terms have the same meaning in their first-person and third-person uses; and we invite scepticism about whether the mental properties we ascribe to others on the basis of their behaviour are the same as the mental properties we ascribe to ourselves without evidence. For ‘why should we think that a predicate that is sometimes applied on the basis of observation, and sometimes not, is unambiguous’? (Davidson 1984, 8). An adequate account of mental terms and concepts, Davidson insists, must answer that question. But the Wittgensteinian approach does not even attempt to address it. It ‘correctly describe[s] the asymmetry between first and other person ascriptions of mental predicates [but it does] nothing to explain it’ (Davidson 1984, 8).

In Section 2, I explore Davidson’s own answer to the explanatory challenge he poses. I argue that Davidson’s account explains less than he seems to suggest; he must, in the end, accept that it is simply a basic, irreducible fact that a person can, without observation, reliably self-ascribe the mental

A Fragment”), and by the page numbers of the 1st and 2nd editions, which are cited in square brackets.

2 A notable exception is Davidson’s explicit endorsement of what he takes to be Wittgenstein’s reason for holding that thought and language are essentially social: that ‘unless a language is shared, there is no way to distinguish between using the language correctly and using it incorrectly; only communication with another can supply an objective check’ (Davidson 1991, 209–10). For other references to Wittgenstein in connection with that point, see, e.g., Davidson 1992, 116; 1994, 124; 1997, 129; and 2001b passim.

3 Davidson’s discussion moves fairly freely between talk of mental terms and concepts, on the one hand, and talk of mental states or properties, on the other. My discussion follows his in that respect.

4 See Davidson 1987, 16 and Davidson 1991, 207 for other statements of the same objection. Davidson is characteristically cautious about explicitly attributing ‘the Wittgensteinian style of answer’ (Davidson 1987, 17) to Wittgenstein himself; he writes of ‘Wittgenstein’s insight (if it is Wittgenstein’s)’ (Davidson 1987, 16). Note also that Davidson 1999 acknowledges that there is more to Strawson’s view than Davidson’s original critique allowed.
properties that others ascribe to her on the basis of her behaviour. Then I turn to
Wittgenstein’s position. In Section 3 I consider and reject the suggestion that
Wittgensteinian considerations show Davidson’s challenge to be in some way
illegitimate or misconceived. In Section 4, I argue that Wittgenstein’s discus-
sion of mental concepts offers a plausible answer to Davidson’s challenge.

2 Davidson on Self-Ascription and Other-Ascription

Davidson asks what reason there is to think that mental terms have the same
meanings, and pick out the same properties, in their first-person and third-
person uses. That question seems a legitimate one – though, as we shall see
below, some might question its legitimacy on supposedly Wittgensteinian
grounds. Consider a different kind of case. We make claims about birds on
the basis of different sorts of perceptual evidence: I can know that there is
a cuckoo nearby by hearing its distinctive call, but also, by seeing it fly by. How
would we respond to the question, what reason there is to think that the term
‘cuckoo’ has the same meaning, and picks out the same thing, when applied on
the basis of auditory experience and on the basis of visual experience? We would certainly not say that it is just a primitive part of the meaning of
the word ‘cuckoo’ that it can be applied on the basis both of auditory and
of visual experience. Rather, we would appeal to our conception of what
a cuckoo is – a physical object of a certain kind – to show how it is that one
and the same thing can be known about in these different ways. Davidson’s
general strategy in answering his own question about the mental case is broadly
the same. His idea is to set out a conception of thought – more specifically,
a conception of the content of thought – that will show how it is that one and the
same property (the property of having the content, that \( p \)) can be known about
in two quite different ways.

We can distinguish two elements in Davidson’s explanatory challenge: two
questions he thinks an account of mental concepts or properties must address.
1. Are the mental properties that I ascribe to others on the basis of their
behaviour the same as the properties that I self-ascribe without reference
to my behaviour?
2. Given that they are the same properties:
   a) How is it possible for me to know on the basis of behavioural evidence
      that others have those properties?; and
   b) How is it possible for me to know without evidence that I have those
      same properties?

Davidson’s answer to his own challenge turns on the idea that, in the simplest
and most basic cases, the meanings of my words – and so, according to
Davidson, the contents of my thoughts – are determined by the way I apply
those words to things and kinds in my environment. As Davidson puts it,
‘whatever she regularly… apply[ies] them to gives her words the meaning they have and her thoughts the contents they have’ (Davidson 1987, 37). On this view, using the word ‘cat’ to mean cat, for instance, is a matter of being generally disposed to apply the word to cats (and to withhold it from things that are not cats). How does that conception of meaning and content help in addressing Davidson’s explanatory challenge?

At first sight, Davidson’s broadly dispositional conception of meaning might seem to give an immediate and satisfying answer to question 1. Using a word with a given meaning is a matter of being disposed to apply it in a certain way: I ascribe that dispositional property to others when I say that they use the word ‘cat’ to mean cat; I ascribe the same property to myself when I say that I use the word ‘cat’ to mean cat. There seems no temptation (as there may be a temptation in the case of properties like pain) to think that there are two different properties in play: a dispositional property that we ascribe to others, and a non-dispositional property that we ascribe to ourselves.

But on reflection, we can see that the mere appeal to dispositional properties does not by itself answer question 1. As Davidson himself points out, a creature can interact with its environment, and with other creatures, in complex and regular ways without having thoughts about an objective world and without acting for reasons. Such a creature could produce sounds in regular ways in response to objects and events in its environment without thereby using a language: its sounds could carry information (in a familiar, causal sense) without having meanings; in producing those sounds, it would not be saying that things were some way. (See for instance Davidson 1997, 128; Davidson 2001b, 11–12). So we can distinguish two different kinds of disposition that a creature might have if it is disposed to produce the sound ‘cat’ in response to cats in its environment. It may, like me, have a genuinely linguistic disposition: the disposition to use the word ‘cat’ to say that something is a cat. Alternatively, it may have something more basic: the mere disposition to produce the sound ‘cat’ in the presence of cats. So even if we agree that the property I ascribe in the first-person case and the property I ascribe in the third-person case are both dispositional properties, there is room to doubt whether the same dispositional property is involved in each case. We are left with the same question as before:

---

5 Davidson recognizes, of course, that this is an oversimplification. It is possible for someone to use the word ‘cat’ to mean cat without being able to distinguish cats from non-cats. Such a person will have no general disposition to apply the word to the cats in her environment and withhold it from the non-cats (see Davidson 1988, 48–9). But language and thought, Davidson insists, depend on there being a core of cases for which something close to the simple dispositional view is true. These are the cases ‘that anchor language to the world’ (Davidson 1988, 45: for other statements of the same idea, see Davidson 1987, 29; 1990, 197). And it is these cases that are the focus of his explanation of the univocality of, and asymmetry between, first-person and third-person ascriptions of content.

6 For a full discussion of Davidson’s account of first-person authority, see Child 2013.
what reason is there to think that the dispositional properties I ascribe to others on the basis of their behaviour are the same as the dispositional properties I self-ascribe without observation?

Davidson thinks that we do know that our fellow human beings are thinking, reasoning, language-using subjects, and not mere information-processors that react in regular ways to the environment and each other. So we do know that the dispositional properties we ascribe to others are the genuinely mental and linguistic properties that we ascribe to ourselves. And he says something about how we know it, with his story about triangulation: about the complex pattern of interactions between two or more people, communicating about a shared environment, which he takes to be a necessary condition for thought. Davidson emphasizes, however, that the difference between a full-blown subject of thought and a mere information-processor is basic and unanalysable; it cannot be explained in terms of anything else. A genuine subject grasps the distinction between how she takes the world to be and how it is: she has the idea of error; the concept of objective truth. That is the difference between a subject of thought and a mere information-processor. But we cannot explain that difference in other terms. In Davidson’s picture, then, this most fundamental feature of the mental – what makes something a genuinely mental state at all – must be accepted as basic and unanalysable. We might wonder why this kind of unanalysability or inexplicability is more acceptable than the inexplicability that Davidson objects to in the Strawson/Wittgenstein account. I will return to that point shortly.

What about question 2? Davidson’s view of meaning makes it easy to see how I can know what someone else means by observing their behaviour. Using a word with a given meaning is a matter of being disposed to apply it in a particular way; and I can tell how someone is disposed to apply a word by observing how she does in fact apply it. That answers the first half of the question: 2(a). But how does Davidson answer the second half of the question: 2(b)? How do I know what my own words mean, without reference to my linguistic behaviour?

There is a debate in the literature about how exactly we should understand Davidson’s account of our knowledge of the meanings of our own words. It is clear that, if the meanings of a speaker’s words are determined by the way that he himself applies those words, he cannot generally misapply his own words. As Davidson puts it: ‘a person cannot generally misuse his own words, because it is that use which gives his words their meaning’ (Davidson, 1993, 250). (Analogously: a person cannot generally act out of character, because it is the

---

8 I bracket any difficulties relating to the point just discussed in the text, about how we know that others are genuinely using language, rather than simply producing sounds in regular and consistent ways.
way she generally acts that determines what her character is.) But how does that relate to the idea that a speaker generally knows what his words mean? How do we get from the claim that a speaker cannot generally misapply his own words to the claim that he generally knows what those words mean? (In terms of our analogy: How do we get from the claim that a person cannot generally act out of character to the claim that a person generally knows what her character is?)

Some commentators take Davidson to hold the deflationary view that knowing the meaning of a word just consists in the ability to use it correctly. On this deflationary view, there is no gap between the claim that a speaker cannot generally misapply her own words and the claim that she generally knows what her words mean; for all there is to knowing what one’s words mean is being able to apply them correctly. If we read Davidson in this way, it will follow directly from his account of what determines meaning that speakers generally know what their words mean – and what their thoughts are about.

Others interpret Davidson as taking a less minimalist view of knowledge of the meanings of one’s words. On this less minimalist interpretation, knowing what my word ‘cat’ means is not merely a matter of having a practical ability to use it correctly. It is a piece of higher-level, propositional knowledge: the knowledge that my word ‘cat’ refers to cats. But, on Davidson’s view of meaning, this is a kind of higher-level knowledge that speakers cannot fail to have – at least in the simple, basic cases on which we are focusing. The reason is this. For Davidson, using a word with a given meaning is, in the simplest, most basic cases, a matter of being disposed to apply it to things of a particular kind in one’s immediate environment. So a speaker who means something by a word must in general be able to give a correct ostensive statement of its meaning: ‘That’s a cat’, for instance. Explaining the meaning of a word in that way draws directly on the very disposition that using the word ‘cat’ to mean cat consists in. So it is no mystery that I can know the meaning of my word, and the content of the associated thought, without observing my linguistic behaviour.

However, even supposing that Davidson offers a completely satisfying explanation of our knowledge of the contents of our own and other people’s thoughts, he does not give anything like a general explanation of our knowledge of mental states. For his account only deals with our knowledge of the contents of thoughts. He says nothing at all about how I know what attitudes I and others have towards those contents: about how I know that I or someone else believes that p, or intends to Φ, and so on. But could we develop the basic

11 I have argued elsewhere that this account of our knowledge of the meanings of our own words does not in fact adequately explain the character of that knowledge (see Child 2013, 541–2). I will not repeat that criticism here.
intuitions behind Davidson’s account of our knowledge of the contents of attitudes to give a more general account of knowledge of those attitudes themselves?

The starting point would be the idea that mental properties are essentially dispositional properties. Davidson certainly accepts that idea. He says explicitly, for instance, that a pro-attitude is ‘a disposition to act under specified conditions in specific ways’ (Davidson 1987, 108); and he takes a parallel view of belief, intention, and the rest. Now as before, if mental properties are dispositional properties of this sort, it is easy to see how we can in principle know other people’s mental properties on the basis of their behaviour. But how easy is it, in the general case, to understand how a dispositional property can be self-ascribed without evidence? We have many dispositional properties that we are not in a position to know that we have without observation. Being strong, being allergic to pollen, and being friendly, for example, are dispositional properties. But I cannot know that I have any of those properties without observing how I behave or react in the relevant circumstances. What explains why it is different with mental properties? How is it possible to know without observing my behaviour that I have ‘a disposition to act under specified conditions in specific ways’; how do I know without observation that I believe that $p$, or intend to $\Phi$?

However we interpret Davidson’s explanation of our knowledge of the meanings of our words, it is hard to see how that explanation could be adapted to answer the more general question. On the minimalist interpretation of Davidson’s view, knowing what one’s words mean just consists in being disposed to apply them correctly. But we cannot explain a subject’s knowledge of her own attitudes in a parallel way. Believing that $p$ may be a disposition to act in certain ways. But knowing that I believe that $p$ is not merely a matter of having that very disposition; it involves more than simply believing that $p$. On the less minimalist interpretation, Davidson’s explanation of our knowledge of the meanings of our words turns on the fact that judging that one’s word ‘cat’ means cat, for instance, is a direct manifestation of the disposition that makes it true that one’s word has that meaning: the disposition to apply the word ‘cat’ to cats. But again, that is not a plausible model for our knowledge of what we believe: for judging that I believe that $p$ is not, in the same way, a direct manifestation of the disposition that makes it true that I believe that $p$.

Where does that leave Davidson’s response to his own challenge? As far as I can see, Davidson’s account ultimately takes it as a basic, unanalysable fact about such properties as belief, desire, and intention – which he conceives as dispositions to act in certain ways – that subjects can reliably self-ascribe those dispositional properties without observing their own behaviour. That propositional attitudes have that feature is part of the difference between genuine propositional attitudes and the more basic dispositional properties possessed
by simpler creatures. But, by Davidson’s own insistence, there is no way of explaining that difference in other terms. Davidson complains that the Wittgenstein-Strawson view correctly describes the asymmetry between self-ascriptions and other-ascriptions, but does not explain it. But, if I am right, the same is true of his own account.

3 Davidson’s Challenge and Wittgenstein’s Conception of Philosophy

How would Wittgenstein respond to Davidson’s demand for an explanation of how mental terms can have the same meanings, and ascribe the same properties, in their first-person and third-person uses? It might be thought that Wittgenstein would either reject Davidson’s challenge altogether or else claim that the issue it raises is trivial and entirely linguistic. In the current section, I explore and reject those two reactions; from Wittgenstein’s point of view, I argue, the question Davidson poses is legitimate and substantive.

3.1 Is the Demand for an Explanation of Sameness of Meaning Legitimate?

On one view, Wittgenstein would simply reject Davidson’s question; the demand for an explanation of the univocity of mental terms in first-person and third-person ascriptions, he would say, is illegitimate. I shall consider two possible arguments for ascribing that position to Wittgenstein.

A first argument appeals to Wittgenstein’s general insistence that it is not the job of philosophy to explain anything: ‘Philosophy just puts everything before us, and neither explains nor deduces anything’ (Wittgenstein 1953, §126); ‘There must not be anything hypothetical in our considerations. All explanation must disappear, and description alone must take its place’ (Wittgenstein 1953, §109). And Wittgenstein explicitly applies that general point to the case of mental concepts in particular: for example, ‘I look at this language-game [of applying mental terms to ourselves and others] as autonomous. I merely want to describe it, or look at it, not justify it’ (Wittgenstein 1992, 40). So from Wittgenstein’s point of view, it will be said, it is simply a mistake for philosophy to address explanatory questions like Davidson’s.

How should we respond to that argument? It is quite true that Wittgenstein rejects the demand for explanation in philosophy – in the sense of ‘explanation’ expressed by the German word Erklärung: that is to say, hypothetical or causal.

12 For other well-known statements of opposition to explanation in philosophy, see Wittgenstein 1953, §496 and §654.
But Wittgenstein does not reject the idea of explanation in philosophy altogether. After all, he is clear that philosophy can yield understanding: Verstehen. A key aim for philosophy, as he sees it, is to produce ‘surveyable representations’ of our practices. And ‘a surveyable representation’, he says, ‘produces precisely that kind of understanding which consists in “seeing connections”’ (Wittgenstein 1953, §122). But producing understanding is, precisely, a kind of explaining. There is, then, no general Wittgensteinian prohibition on explanation in philosophy. There is only a prohibition on a particular kind of explanation: causal or hypothetical explanation. And when Davidson asks for an explanation of the sameness of meaning of mental terms in their first-person and third-person uses, he is obviously not asking for a causal explanation. He is asking for an account of our use of mental terms – a surveyable representation, we could say – that will allow us to understand how those terms can be univocal, given the very different bases on which they are applied.14 There is nothing essentially unWittgensteinian in Davidson’s question; nor is there any presumption about the form of an adequate answer that offends against any Wittgensteinian principle.

A second possible argument is that Wittgenstein would reject Davidson’s demand for an explanation of the univocality of mental terms on the grounds that it misconstrues the dialectical position. We could put the point like this. ‘There is in Wittgenstein’s view no genuine question about the univocality of mental terms. On the contrary, it is perfectly clear in ordinary life that the term “pain”, say, has the same meaning in its first-person and third-person uses. There is nothing intrinsically puzzling about that, and no need for a philosophical account to explain it. If we are sometimes tempted to think that the word “pain” must mean something different in its first-person and third-person applications, or that the claim that it has the same meaning stands in need of a substantive philosophical justification, those are not ideas that we ordinarily take seriously; nor are they the kind of data from which philosophical reflection must begin. On the contrary, they are artefacts of a distinctively philosophical picture of the mental: the picture of the mental as an inner realm that lies concealed behind a person’s behaviour; a realm with which only the subject herself is acquainted and which others can only know indirectly, if at all. We are easily gripped by that picture when we stand back from the practice of applying mental terms to ourselves and others and try to achieve a reflective understanding of the practice. And when we conceive of the mental in terms of

13 Erklärung is the word used in each of the passages quoted or cited above: Wittgenstein 1953, §§109, 126, 496 and 654.

14 In his response to a paper by Anita Avramides about his views on this topic, Davidson comments on the kind of explanation at issue: ‘it seems to me… that, in practice, at least in the present context, there is no significant difference between description and explanation’ (Davidson 1999, 155).
this “inner-outer picture”, questions about the univocality of mental terms can seem genuinely pressing. But in Wittgenstein’s view, we should not accept the inner-outer picture and then look for a substantive answer to the question, how it is possible for our mental terms to be univocal. The task for philosophy is, rather, to diagnose the sources of the inner-outer picture and to expose the picture as an illusion. Once that is done, the suggestion that the word “pain” is ambiguous will be seen to have no plausibility, and there will be no need for any positive explanation to justify the common-sense conviction that “pain” and other mental terms are univocal.

That line of thought plainly reflects an important strand in Wittgenstein’s philosophy. And it is clearly true that Wittgenstein does not think that the univocality of mental terms is genuinely in doubt. But that does not show that, from Wittgenstein’s point of view, Davidson’s question is misplaced or illegitimate. One strand in Wittgenstein’s philosophy is the aim of achieving a reflective understanding of a practice that we have mastered, but of which we do not command a clear view. In the current case, it is one thing to say that we have a practice of ascribing mental states to others on the basis of their behaviour and to ourselves without evidence. It is another thing to have a reflective understanding of the practice: an understanding that allows us to see why we are entitled to treat the self-ascription and other-ascription of mental terms as elements of a single practice, rather than thinking of our terms as ambiguous. That is a recognizably Wittgensteinian project. And it is reasonable to describe the project as that of giving an account of our practice that makes it intelligible that the property we ascribe to ourselves is the same as the property we ascribe to others. It is wrong, then, to think that Wittgenstein would simply reject the request for an account that shows why mental terms are not ambiguous in their first-person and third-person uses. That, at any rate, is what I shall aim to show.

There is a further and related point. Wittgenstein regards the inner-outer picture as a bad picture of the mental and of our use of mental terms. If the bad picture were right, we can suppose, then ‘pain’ and other mental terms really would be ambiguous in their first-person and third-person uses. But if there is a bad picture of the mental, on which the word ‘pain’ really would be ambiguous, there must be a good view – an accurate account of our practice – that will allow us to see why the term ‘pain’ is not in fact ambiguous. It cannot be that there is literally nothing to say about how we actually use the term and why that use does not make it ambiguous. All that Davidson is asking for when he presses his explanatory question is an account of our mental terms that does that. And, I shall argue, when we look at the detail of Wittgenstein’s treatment of mental terms we can see that he does offer a positive account of just that kind.
3.2 Is the Answer to Davidson’s Challenge Trivial?

A second response to Davidson’s explanatory challenge holds that, from Wittgenstein’s perspective, the answer to Davidson’s question is trivial. On this view, whether or not the property of pain that we ascribe to ourselves is the same as the property we ascribe to others is determined by how we choose to count properties. And that is a merely verbal matter; nothing substantial turns on the choice. As things are, we regard the property we self-ascribe when we use the word ‘pain’ of ourselves, and the property we ascribe to others in third-person uses of the word ‘pain’, as the same property. But we could just as well have chosen to regard it as a different property. And if we had done so, we would not be getting things wrong – failing to recognize an objective fact about the identity of the properties ascribed in first-person and third-person uses of the word. For there is no absolute or objective standard of sameness of property: whether or not the same property is involved in the two kinds of case is entirely a matter of what we choose to count as sameness of property.

Someone might cite passages like the following in favour of ascribing that view to Wittgenstein:

If I were to reserve the word ‘pain’ solely for what I had previously called ‘my pain’, and others ‘LW’s pain’, I’d do other people no injustice, so long as a notation were provided in which the loss of the word ‘pain’ in other contexts were somehow made good. Other people would still be pitied, treated by doctors, and so on. It would, of course, be no objection to this way of talking to say ‘But look here, other people have just the same as you!’ (Wittgenstein 1953, §403)

On the present interpretation, the point Wittgenstein is making is this. If we use the same word, ‘pain’, in each case, then we will say that the property that other people have is the same as the property that I have. If we reserve the word ‘pain’ for the first-person case, and use a different word in the third-person case, then we will say that other people have a different property from me. But nothing important turns on which we say. Provided everything else stays the same, the difference between the two positions is purely verbal.

And consider this passage, where Wittgenstein explicitly raises Davidson’s question:

The question can be raised: Is a state that I recognize on the basis of someone’s utterances really the same as the state he does not recognize this way? And the answer is a decision. (Wittgenstein 1992, 8–9; see Wittgenstein 1982, §428 for a similar remark.)

Wittgenstein says it is ‘a decision’ whether to regard the states we self-ascribe and the states we ascribe to others as the same. His point, on the present view, is that there is nothing essential to the practice of using mental terms of ourselves and others that makes it true either that the states are the same or that they are different. We can decide to call them ‘the same’; or we can decide to call them
‘different’. But that is just a matter of how we choose to represent things; it is not a substantial question in its own right.

There is something right and something wrong in that account of Wittgenstein’s approach to questions about sameness of meanings and properties. It is true that, in Wittgenstein’s view, standards of sameness of property are not given by nature but are determined by human practices of description and classification. And it is true that, in his view, a system of classification that counted the property of pain that we self-ascribe without evidence as different from the property we ascribe to others on the basis of their behaviour would not be wrong; it would simply be a different system from ours. But it is wrong to think that Wittgenstein regards these questions about sameness of meaning, or sameness of property, as merely verbal and entirely trivial. And it is wrong to think that he takes such questions to have no substantive answers. That, at least, is what I shall argue in Section 4.

4 A Wittgensteinian Response to Davidson’s Challenge

There is a general question about when we should say that a word has a single meaning in its various uses, and when we should say that it has two (or more) different meanings. Wittgenstein writes:

What does it mean to say that the ‘is’ in ‘The rose is red’ has a different meaning from the ‘is’ in ‘Two times two is four’? If it is answered that it means that different rules are valid for these two words, the retort is that we have only one word here. – And if I attend only to the grammatical rules, these do allow the use of the word ‘is’ in both kinds of context. – But the rule which shows that the word ‘is’ has different meanings in these sentences is the one allowing us to replace the word ‘is’ in the second sentence by the sign of equality, and forbidding this substitution in the first sentence. (Wittgenstein 1953, §558)

He returns to the same question a few sections later:

Now isn’t it remarkable that I say that the word ‘is’ is used with two different meanings (as copula and as sign of equality), and wouldn’t want to say that its meaning is its use; its use, namely, as copula and as sign of equality?

One would like to say that these two kinds of use don’t yield a single meaning; the union under one head, effected by the same word, is an inessential coincidence. (Wittgenstein 1953, §561)15

He discusses an analogy:

Let’s say that the meaning of a piece is its role in the game. – Now let it be decided by lot, before a game of chess begins, which of the players gets white. For this, one player holds a king in each closed hand, while the other chooses one of the two hands, trusting to

---

15 See Wittgenstein 1982, §§272–291 for related considerations about sameness and difference of meaning.
luck. Will it be counted as part of the role of the king in chess that it is used to draw lots in this way? …

But, after all, the game is supposed to be determined by the rules! So, if a rule of the game prescribes that the kings are to be used for drawing lots before a game of chess, then that is an essential part of the game. What objection might one make to this? That one does not see the point of this prescription. Perhaps as one likewise wouldn’t see the point of a rule by which each piece had to be turned round three times before one moved it. If we found this rule in a board-game, we’d be surprised and would speculate about the purpose of the rule. (‘Was this prescription meant to prevent one from moving without due consideration?’)

If I understand the character of the game aright, I might say, then this isn’t an essential part of it. (Wittgenstein 1953, §§563, 567–8)

We can spell out some of the lessons of that discussion.

In the case of chess, Wittgenstein’s reflections suggest the following view. The use of the kings to draw lots is not essential to the game, even if it is included in the rules. And the reason it is not essential is that it does not matter to the point of the game; no use is made in the game of the fact that the kings are used to draw lots; it would make no significant difference to the game if we used something else to draw lots. So the role of the king in the game of chess, and its role in drawing lots are distinct roles. Of course we can yoke the two roles together, by stipulating in the rules of the game that the king is to be used in drawing lots. But for all that, they remain distinct roles; they do not form a unity.

Wittgenstein takes a parallel view about questions of sameness and difference of meaning. Why should we say that the word ‘is’ has two different uses, and two different meanings, rather than saying that it has a single use, ‘as copula and as sign of equality’? Because the two meanings are explained in distinct and unconnected ways; there is no point in using the same word in both cases; no use is made of the fact that it is the same word.16 For those reasons, the two uses do not form a unity. Now Wittgenstein suggests that it is (or can be) a ‘decision’ whether a given case is a case of ambiguity or, rather, a case of a word’s having one meaning and two kinds of use.17 But the lesson of the quoted discussion from Wittgenstein 1953, §§558 ff (and of the related discussion in Wittgenstein 1982, §§272 ff) is that in answering the question, ‘One meaning or two?’, we are not making an arbitrary decision. On the contrary, what we say is answerable to independent facts about the practice of using the word. Those facts do not absolutely require us to answer the question in one way or the other. But if it is to keep contact with the ordinary

---

16 For the first of these points, see also Wittgenstein 1982, §280: ‘If you say that [a word] has two meanings then you have to use an explanation to distinguish between them. (There can be various reasons for doing that.)’

17 For the notion of decision in this context, see, e.g., Wittgenstein 1982, §275: ‘You visit a tribe; they have a language; in this language you hear a word (a sound) – does it have one meaning, or several? How will you find out, how will you decide?’ (second emphasis added).
notion of meaning, the way we count meanings must answer to the kinds of fact about our practice of using the word that we have just mentioned.

In the light of all of that, what should we say about the use of the word ‘pain’ in first-person and third-person ascriptions? Wittgenstein says that it is ‘a decision’ whether to say that the properties we ascribe to ourselves without reference to our behaviour are the same as the properties we ascribe to others on the basis of their behaviour. His point is that it would not be wrong to use different words in first-person and third-person ascriptions of what we currently regard as the same mental properties; nor would it be wrong to regard the properties ascribed as different properties. Nonetheless, it is clear that, in Wittgenstein’s view, the overall use of the word ‘pain’, for instance, encompassing both first-person and third-person applications, has an internal unity. Its use in first-person ascriptions and third-person ascriptions are not distinct, self-standing practices, arbitrarily yoked together under the single label ‘pain’. Rather, they are parts of an integrated whole. Wittgenstein puts the point like this:

One has to look at the concepts ‘to be in pain’ and ‘to simulate pain’ in the *third and first* person. Or: the infinitive covers all persons and tenses. Only the whole is the instrument, the concept. (Wittgenstein 1992, 37)

Similarly:

One may have the thought: ‘How remarkable that the *single* meaning of the word “to feel” (and of the other psychological verbs) is compounded of heterogeneous components, the meanings of the *first* and of the *third* person.’

But what can be more different than the profile and the front view of a face; and yet the concepts of our language are so formed, that the one appears merely as a variation of the other. And of course it is easy to give a ground in facts of nature for this structure of concepts. (Heterogeneous things; arrow-head and arrow-shaft.) (Wittgenstein 1980a, §45)

Similarly, Wittgenstein suggests, though the first-person and third-person uses of mental terms are in themselves very different, they belong to a single overall use with a genuine, natural unity.

But it is one thing to claim that the first-person and third-person uses of ‘pain’ form a single meaning, with an internal unity. It is another thing to justify that claim – to explain in detail why we should say, with Wittgenstein, that only the whole use of the word ‘pain’, in the first and third person, is an ‘instrument’. The basic reason, Wittgenstein thinks, is that the two uses are mutually interdependent; they cannot be given distinct explanations. We can bring out the interdependence in both directions.

In the first place, the first-person use of mental terms is not independent of their third-person, behaviour-based application: their application to other people on the basis of their circumstances and behaviour. In teaching the child to replace her natural, pre-linguistic expressions of pain with the linguistic
expression ‘I’m in pain’, we teach her to make first-person uses of the word that are not grounded in observations of her behaviour. But, at the outset, we teach her to make such first-person uses only in circumstances where she already expresses her pain non-linguistically and, therefore, where we can apply the word ‘pain’ to her on the basis of her behaviour. And she only qualifies as having grasped the meaning of the word if her first-person applications are consistent with the way we apply the word to her on the basis of her circumstances and behaviour. Later on, we accept her self-ascriptions of pain as decisive in cases where we have no independent basis for making a third-person ascription of pain. But even then, her overall use of the term must not part company from the third-person criteria. If she regularly applied the word ‘pain’ to herself in a way that conflicted with the external evidence, we would start to doubt whether she really did understand the word after all. By contrast, if the word ‘pain’ were ambiguous, there would be no reason why a person’s applications of the word ‘pain’ to herself would need to harmonize with the way that others applied the word ‘pain’ to her on the basis of her behaviour; for whether or not the first-person term ‘pain’ was true of a person would have nothing to do with whether or not the third-person term was true of her.

The first-person use of mental terms is not, then, a distinct, self-standing practice, independent of their third-person use. Similarly, Wittgenstein thinks, the third-person use of mental terms is not independent of the first-person use of those terms by the subjects of our third-person ascriptions. A word that was applied to others on the basis of their behaviour, but which subjects could not learn to apply to themselves without reference to their behaviour, he suggests, would not be a word for a kind of mental phenomenon at all. He writes:

One might distinguish between two chimpanzees with respect to the way in which they work, and say of the one that he is thinking and of the other that he is not. But here of course we wouldn’t have the complete employment of ‘think’. The word would have reference to a mode of behaviour. Not until it finds its particular use in the first person does it acquire the meaning of mental activity. (Wittgenstein 1980b, §§229–30)

In other passages, too, Wittgenstein imagines counterparts of our ordinary mental terms that relate only to behaviour.

Look at it purely behaviouristically: Someone says, ‘Man thinks, wishes, is happy, angry, etc.’. Imagine that these words were only about certain forms of behaviour on certain occasions. One could suppose that whoever talks about human beings in this way had first observed these kinds of behaviour in other beings and was now saying that these phenomena could also be observed in human beings. That would be like our saying this of a species of animals. (Wittgenstein 1980b, §33)

One could imagine a concept of fear, for instance, that had application only to beasts, and therefore pertained only to behaviour. (Wittgenstein 1980b, §333)
The cases are not developed. But Wittgenstein seems to trace the fact that these terms pick out behaviour, or behavioural dispositions, rather than genuinely mental phenomena to the fact that all there is to their use is their application to others on the basis of their behaviour; they have no relation to any first-person use by the creatures to which the terms are applied. The lesson of that is that the third-person use of a term only counts as the use of a mental term if it is connected in appropriate ways with related first-person uses.\textsuperscript{18} As before, the first-person and third-person uses are not distinct and self-standing: they are parts of a unified whole. That is why Wittgenstein says that it is the ‘whole use’ of mental terms, encompassing both first-person and third-person uses, that is ‘the instrument’.

4.1 An Objection to the Wittgensteinian View; and a Reply

We can anticipate the following objection to the Wittgensteinian view that we have just described. ‘Wittgenstein suggests that the third-person use of mental terms is not independent of their first-person use: a term that we apply to a creature on the basis of its behaviour only picks out a mental phenomenon if the creature itself can make first-person uses of that term without reference to its behaviour. That may be true for concepts like belief and intention; it is quite plausible to think – with Wittgenstein, Davidson, McDowell, and others – that such concepts only properly apply to creatures that can think of themselves as accepting and rejecting things for reasons.\textsuperscript{19} But it seems completely implausible to hold that it is a condition for the truth of the third-person ascription “That creature is in pain” that the creature should be capable of self-ascribing pain. It may be true that it is not until the word “thinking” finds its particular use in the first person that it acquires the meaning of mental activity. But it is surely not true that it is not until the word “pain” finds its particular use in the first person that it acquires the meaning of sensation.’

How should a defender of Wittgenstein respond to this objection? One possibility would be to bite the bullet, insist that the dependence of the third-person use of mental terms on their first-person use is a perfectly general feature of such terms, and accept that creatures that cannot self-ascribe pain cannot experience pain: or at least, that they cannot experience pain in exactly the same sense that we do.\textsuperscript{20} It is possible that Wittgenstein would respond in

\textsuperscript{18} Notice the close similarity between, on the one hand, Wittgenstein’s distinction between the purely behavioural concepts of fear, belief, desire etc. and their fully mental counterparts and, on the other hand, Davidson’s distinction between the concepts we can apply to creatures that merely react to their environment in regular ways and the fully mental concepts that apply to genuine subjects of thought.

\textsuperscript{19} We should note, however, that many other philosophers take this to be too demanding a requirement.

\textsuperscript{20} For a position of that kind, see McDowell 1994, Chapter 6.
that way. Consider the question: ‘Is the sensation we ascribe to a fox when we say that the fox feels pain the same as the sensation we ascribe to a person when we say that she feels pain?’ Wittgenstein would certainly allow that we could legitimately count the fox’s sensation as different from ours; it would not be wrong to classify sensations like that. However, the current proposal is not simply that that way of classifying sensations is one possible form of classification: the idea is, rather, that it is implicit in the ordinary notion of sensation. And it is far from clear that Wittgenstein would endorse that stronger view.

A different response – which seems closer to Wittgenstein – would involve distinguishing between the conditions of applicability of different mental concepts. On this view, concepts like thought, belief, and intention only fully apply to creatures that are capable of self-ascribing those concepts, as Wittgenstein insists. But for concepts like pain and visual experience, there is no similar condition. There is a precedent in Wittgenstein’s discussion of experience for distinctions of this kind between the conditions of application of different mental concepts. He says there that there are some experiences that a person can only have if she has mastered a particular technique: for instance, a person looking at a triangle can only have the experience of seeing ‘now this as apex, that as base – now this as apex, that as base’ if she is ‘capable of making certain applications of the figure with facility’ (Wittgenstein 1953, part II xi §222 [p. 208]). He continues:

But how odd for this to be the logical condition of someone’s having such-and-such an experience! After all, you don’t say that one ‘has toothache’ only if one is capable of doing such-and-such. – From this it follows that we cannot be dealing with the same concept of experience here. It is a different concept, even though related. (Wittgenstein 1953, part II xi §223 [p. 208])

It seems plausible that Wittgenstein would draw a similar distinction in the present case between concepts such as thought, wishing, happiness, anger, and fear, which in his view apply only to creatures that are capable of self-ascribing them, and concepts such as pain, for which there is no such requirement.

4.2 Conclusion

What should we conclude about Wittgenstein’s attitude to the question, whether mental terms mean the same in their first-person and third-person uses, and whether the mental properties we ascribe to ourselves without evidence are the same as the properties we ascribe to others on the basis of their behaviour?

In the first place, from Wittgenstein’s point of view these questions about the univocality of mental terms and the sameness of mental properties are perfectly legitimate. And they are to be answered by achieving a reflective understanding of our practice of ascribing mental properties to ourselves and others – rather
than by reference to any supposedly more basic metaphysical facts about the sameness or difference of the properties our terms pick out.

The point of saying that we make a ‘decision’ when we answer these questions is that no particular answer is forced on us by the facts: it is possible to count properties in different ways, and no way of counting properties is simply wrong. Notwithstanding that point, however, the question ‘same property or different?’ does not in this case call for any active decision; it is so overwhelmingly natural to count one’s own pain and other people’s pain, say, as instances of the same property that it would be quixotic to say anything else.

Relatedly, the naturalness of counting mental properties in the way we do is not a merely psychological fact about us. And the ‘decision’ to count properties that way is not arbitrary. It has a deep rationale in the underlying character of our practice of applying mental terms to ourselves and others, specifically, in the interdependence of the first-person and third-person uses of those terms.

There remains a question in this area that Wittgenstein’s philosophy does not attempt to answer: How is it possible for a person reliably to self-ascribe without evidence properties that have the dispositional element that mental properties do? What explains that ability? Wittgenstein’s response to that question would be this. When we train people in the use of mental terms, they do in fact learn to make mental self-ascriptions, without reference to their behaviour, that cohere with the ascriptions that others can make on the basis of that behaviour. That we can learn to use mental terms in that way is a deep, contingent fact about us. But it is a fact that philosophy, at least, cannot explain. That, I have argued, is a further parallel between Wittgenstein’s views and Davidson’s.21

References


Davidson, Donald. 1984. ‘First Person Authority.’ In Davidson 2001a.


21 Earlier versions of this paper were presented at conferences and workshops at Birkbeck College, London; Keio University, Tokyo; the American University of Beirut; and York University, Toronto. I am grateful to the organizers of those events, in particular Claudine Verheggen; and to the participants, for much helpful discussion.
The Harmony of Thought and Reality
Wittgenstein and Davidson vs McDowell

Tim Thornton

Introduction

I find it helpful ... to reflect on a remark of Wittgenstein’s: “When we say, and mean, that such-and-such is the case, we—and our meaning—do not stop anywhere short of the fact; but we mean: this is so” [Wittgenstein 1953, §95]. Wittgenstein calls this a paradox. That is because, especially in conjunction with the fact that “thought can be of what is not the case”, it can prompt a reaction in which our minds boggle over what seems a miraculous power of thinking in the most general sense, in this case meaning what one says, to “catch reality in its net”. But Wittgenstein also says, rightly, that the remark “has the form of a truism”.

To a first approximation, this passage summarises both the central problem, and the outlines of its preferred solution to that problem, of John McDowell’s 1994 book Mind and World (McDowell 1994). Only “to a first approximation” because the fact that McDowell invokes Wittgenstein to introduce it suggests the need for some care in characterising his project. McDowell aspires to follow Wittgenstein in rejecting an explanatory view of philosophy in which genuine problems call for substantive philosophical explanations. His aim is not so much to explain how mind can make contact with the world, or the world can impact on the mind, as show how apparent problems with the very idea of such relations can be eased. Following Wittgenstein, McDowell aims to dissolve philosophical problems rather than address them head on. So in this case the threat of an apparent dualism of mind and world is to be disarmed by a better understanding of both.
Mind and World sketches a picture in which thought is open to the world through experience. Impediments to that view that assume a prior dualism of norm and nature can be disarmed through a particular account of the content of experience. Having an experience, understood as a conceptually structured representation, is both normative—it justifies beliefs—but also a passive and natural event. For the last twenty years, this picture has come under fire and McDowell has been forced to modify key aspects of it whilst maintaining the main idea that experience has a representational content.

Despite his Wittgensteinian meta-philosophical aims, I will argue that McDowell’s account fails in its therapeutic intent and has increasingly taken the shape of a philosophical theory of perception of decreasing intuitiveness. But the roots of the difficulties of more recent developments lie in the approach taken in Mind and World to the account of the lack of a gap—or, to put it another way, the harmony—between thought and reality. A better approach is to be found in the work of both Wittgenstein and, coming from a different tradition, Donald Davidson. Further, by comparing Davidson and Wittgenstein on this point, it is possible to point to a different way of scratching the itch that McDowell feels.

McDowell and the Lack of Gap Between Thought and World

The second half of my opening quotation starts to outline McDowell’s response in Mind and World to Wittgenstein’s truistic paradox although, as he concedes, in a style with which Wittgenstein would not have been comfortable:

[T]here is no ontological gap between the sort of thing one can mean, or generally the sort of thing one can think, and the sort of thing that can be the case. When one thinks truly, what one thinks is what is the case. So since the world is everything that is the case (as he himself once wrote), there is no gap between thought, as such, and the world. Of course thought can be distanced from the world by being false, but there is no distance from the world implicit in the very idea of thought. (McDowell 1994, 27)

The idea in the last sentence, that cases of truth and falsity call for different treatment, suggests the influence of one of four antecedent commitments I will describe in this section: disjunctivism. The others are what Brandom calls the “rational constraint constraint”, transcendental empiricism and Sellars’s Myth of Jones.

McDowell’s commitment to disjunctivism dates back to his paper ‘Criteria, defeasibility, and knowledge’ (McDowell 1982). That paper criticises a widespread assumption among Wittgensteinians that knowledge of other minds relies on criteria understood as a priori but defeasible behavioural types. In any particular case, an expression or sign apparently of pain may or may not mean that the person expressing it is actually in pain depending on
whether the criteria are satisfied and undefeated or are satisfied but also defeated by other features of the context.

This prompts the question: on the assumption that it is sometimes possible to know someone else’s mental state, how can it be based “on an experiential intake that falls short of the fact known… in the sense [of] … being compatible with there being no such fact”? (McDowell 1982, 459). If one knows something, then it cannot be the case that – “for all one knows” – things may be otherwise. But if criteria fall short of implying the fact that they are supposed to enable one to know, then they cannot themselves rule out the possibility that the fact does not obtain. So if our everyday concept of knowledge does rule this out, then it cannot be based on perception that the criteria, so understood, for a particular mental state are satisfied.

Rather than assuming that, in the case of pretence, the criteria for mental states are satisfied but are also defeated, McDowell argues that in such a case the criteria only appear to be satisfied. This is a rejection of the idea that criteria are defeasible types of situation in common between veridical and misleading cases. Instead, when criteria are satisfied, one’s experience itself rules out the possibility of acting or pretence. The distinction between cases where behavioural criteria are satisfied and cases where they merely seem to be so applies also to experience more generally and is the basis of McDowell’s broader epistemic disjunctivism. In the more general case, the target idea of neutral behaviour being a veil obscuring others’ inaccessible mental states is replaced by a target view of experience which is neutral between veridical cases of perception and indistinguishable cases of illusion.

Traditional epistemology accords a deep significance to the fact that perception is fallible. It is supposed to show something like this: however favourable a perceiver’s cognitive stance may be, we cannot make sense of subjective states of affairs constituted by a subject’s letting the objective world reveal itself to her … That is supposed to show that the genuinely subjective states of affairs involved in perception can never be more than what a perceiver has in a misleading case. [But] This strains our hold on the very idea of a glimpse of reality … We cannot have the fact itself impressing itself on a perceiver. (McDowell 1994, 112–3)

Instead of this traditional view, McDowell argues that perceptual states or experiences are divided into two sorts: those where objects really are present as they appear and thus the warrant provided by the state for the corresponding perceptual belief is truth-guaranteeing; and those in which the capacity misfires and things are not as they seem. When things do go well, the capacity gives the subject an indefeasible warrant for their perceptual belief.

In Mind and World, McDowell offers this idea in the form of a gloss on the previous comment about meaning from §95 in Wittgenstein [1953] 2009 (in my opening quotation):
To paraphrase Wittgenstein, when we see that such-and-such is the case, we, and our seeing, do not stop anywhere short of the fact. What we see is: that such-and-such is the case. (McDowell 1994, 29)

But this quotation leaves hanging the nature of the harmony of thought and reality in non-veridical cases.

As well as disjunctivism, three other assumptions underpin the positive account of the relation of thought and reality in Mind and World. Two concern the positive attractions of both sides of what McDowell calls the “interminable oscillation” between, on the one hand, coherence theories in which there is no friction on thought imposed by contact with the world and, on the other, the Myth of the Given which attempts to ensure such friction but, impossibly, from outside the space of reasons (McDowell 1994, 9).

The partial truth contained in coherentism is the importance of rational relations governing thought. McDowell quotes Davidson asserting that only a belief can serve as a reason for another belief (McDowell 1994, 14). Although he rejects this claim – arguing that experiences can also serve as reasons for beliefs – he does not dispute the central role of reasons but rather augments their range. Robert Brandom calls this commitment “the rational constraint constraint” adding that a successful account must “make intelligible how perceptual experience imposes not merely causal, but rational constraints on thinking” (Brandom 1998, 369). There is a further echo of Davidson’s slogan in the assumption that McDowell makes that rational relations hold only between conceptually structured relata. Whilst he does extend the reach of reasons to experiences, on his account experiences are belief-like.

The positive motivation for adopting the Myth of the Given is the assumption that thought must have some sort of external worldly constraint to prevent what McDowell calls “frictionless spinning in the void”. He assumes that experience needs to be invoked to play this role, hence a form of “transcendental empiricism”.

I do indeed take a transcendentally motivated empiricism to be innocent. Liberated from a certain discardable conception of the natural, it does not present us with a mystery about the very idea of objective purport. (McDowell 1998c, 405)

The fourth antecedent assumption is less obvious in the text or citations of Mind and World itself but is implicit in the way the account of experience is constructed (it is discussed in detail in McDowell 1998b, 221–238). It is a debt to the kind of theory building that Sellars describes in his Myth of Jones (Sellars 1997, 102–117). Just as the mythical Jones postulates mental states as theoretical entities to explain behaviour on the model of utterances, so McDowell argues that experiences can have conceptual contents on the same model. A clear expression of this is in a later response to a phenomenologically inspired criticism offered by Bill Brewer based on how experience can
represent the different apparent lengths of lines in the Müller-Lyer illusion. Brewer asks:

Is the line with inward hashes represented [in experience] as shorter than it actually is; or is the line with outward hashes represented as longer than it actually is; or both; and by how much in each case? (Brewer 2008, 19, italics added)

None of the options seems plausible. McDowell’s reply likens the content of experiences to the contents of utterances.

Suppose I say, of two lines that are in fact the same length, that one, say A, is longer than the other, say B. It does not follow that I am saying that A is longer than it is, or saying that B is shorter than it is, or saying that both of those things are the case. One of those things would have to be so if what I say were true, but I am not saying of any one of them that it is so. A ‘By how much?’ question does not arise. Just so with an experience that represents one line as longer than the other. (McDowell 2008b, 201, italics added)

The response disarms the objection by comparing the content of experiences to the content of utterances with the same level of determinacy. The content of experiences is modelled on the content of utterances akin to the Myth of Jones.

With these background assumptions in place, the main argument for the picture of world, experience and judgement in Mind and World runs, schematically, as follows:

1. The very idea of mental states having content requires, when all goes well, their rational friction, rather than merely causal contact, with the world.
2. Our only understanding of rational relations requires both relata have conceptual structure.
3. Since experience provides the friction on thought, it must have a conceptual structure.
4. And since, in order to provide genuine friction with the world, experience must be, as it seems pre-philosophically, a kind of direct openness to the world – rather than a self-standing configuration in the inner realm as Descartes proposed – then the world itself must also have conceptual structure.

In sum:

In a particular experience in which one is not misled, what one takes in is that things are thus and so. That things are thus and so is the content of the experience, and it can also be the content of a judgement: it becomes the content of a judgement if the subject decides to take the experience at face value. So it is conceptual content. But that things are thus and so is also, if one is not misled, an aspect of the layout of the world: it is how things are. Thus the idea of conceptually structured operations of receptivity puts us in a position to speak of experience as openness to the layout of reality. Experience enables the layout of reality itself to exert a rational influence on what a subject thinks. (McDowell 1994, 26)
This suggests an account of the harmony of thought and reality, when all goes well, through the common content of experience, judgement and what is the case: the totality of which comprises the world. There is a difference between experience and judgement in that, in the former, conceptual understanding is passively drawn on whereas in judgement it is actively exercised. But in the *Mind and World* account, experiences contain the same kind of claims as judgements: a mark of the influence of Sellars’s Myth of Jones. Hence they mediate between the judgements made on their basis and the world construed as everything that is the case, the set of true (Fregean) thoughts.

Two features of this picture are particularly worth noting here. First, in the spirit of disjunctivism, no account is given of what happens when one is misled aside from the remark about thought being *distanced* from the world by being false. Cases of mere appearance are derivative of the primary case of successfully glimpsing the world. Second, there is at least a prima facie tension between thinking of the world as itself having the conceptual structure of thought and as being made up of extra-conceptual objects and properties. McDowell anticipates the criticism that his account is idealist in the afterword to *Mind and World*.

In Lecture II, I exploit Wittgenstein’s “truism” to discourage the idea of a gap between thought as such and the world. An objector might say something like this: “You make it look as if your drift is not idealistic, so long as you consider the world only as something whose elements are *things that are the case*. In that context, you can exploit the claim that it is no more than a truism that when one’s thought is true, what one thinks is what is the case. But as soon as we try to accommodate the sense in which the world is populated by *things*, by objects (and there had better be such a sense), it will emerge that your image of erasing an outer boundary around the realm of thought must be idealistic in tenor… [I]t obliterates … a possibility of direct contact between minds and *objects* which must surely be external to the realm of thought … ”. (McDowell 1994, 179)

His response is to draw on Gareth Evans’ neo-Fregean account of singular thought. Anticipating that the objection would be raised only by a non-Fregean direct-reference theorist who assumes that Fregean theory can accommodate different senses only by using *descriptive* specifications (based on Russell’s Theory of Descriptions), McDowell argues that Evans’s work *The Varieties of Reference* shows in detail how object-dependent or *de re* senses can also be incorporated into a neo-Fregean theory (Evans 1980). But whilst this indicates the logical space for an account in which thought need not meet reality via a specification (akin to Russell’s Theory of Descriptions) nor depend on an extra-conceptual act of pointing (such as Russellian rather than neo-Fregean direct reference theories), it is not clear that it addresses the worry about idealism. Indeed McDowell comments: “On a proper understanding of the Fregean apparatus, my exploitation of Wittgenstein’s truism… can indeed be reformulated by saying thought and reality meet in the realm of sense”
(McDowell 1994, 180). Whilst this flags the proposed reconciliation of object-dependence and neo-Fregean approaches to thoughts it does not address the worry that the reality of the realm of sense is mind-dependent.

In the twenty years since the publication of *Mind and World*, McDowell has retracted two key claims:

I used to assume that to conceive experiences as actualisations of conceptual capacities, we would need to credit experiences with propositional content, the sort of content judgements have. And I used to assume that the content of an experience would need to include everything the experience enables its subject to know noninferentially. But both these assumptions now strike me as wrong. (McDowell 2008a, 3)

On the new picture, the conceptual content of an experience need not include everything it enables a subject to judge to be the case. The experience itself contains only some of the conceptual content available to the subject. Secondly, it has a form of unity distinct from explicit utterance or silent judgement: it is intuitional rather than propositional.

The resulting picture has continued to come under heavy criticism. It comes under fire both from those who argue that perception has content but it is non-conceptual (e.g., Heck 2000) and from the other side: that perception does not itself have any content and simply makes worldly objects available for judgement (Travis 2013). McDowell continues to argue that there is no necessary conflict between his disjunctivism and a representational model of content but he now suggests that the content is the same in veridical perception and illusion but the way that content is had is different (McDowell 2013).

My focus is none of the details of the picture, of whether it yields a coherent theory of perception. Rather the question is whether it sheds light on the paradoxical but truistic harmony of thought and reality.

In the recent account, the same content is no longer mirrored from world to experience to thought. Thought and world may share the same content, the totality of what is the case, but their relation is mediated by a content with both a different form of unity and structured by only a subset of the concepts available to the subject. McDowell suggests they may comprise the “proper sensibles of sight and common sensibles accessible to sight” (2008a, 4). But he goes on to suggest that this might be broader than colours and geometric shapes.

The concept of a bird, like the concept of a cardinal, need not be part of the content of the experience … But perhaps we can say it is given to me in such an experience, not something I know by bringing a conceptual capacity to bear on what I anyway see, that what I see is an animal … because ‘animal’ captures the intuition’s categorical form, the distinctive kind of unity it has … In an intuition unified by a form capturable by ‘animal’, we might recognize content, under the head of modes of space occupancy, that could not figure in intuitions of inanimate objects. We might think of common
sensibles accessible to sight as including, for instance, postures such as perching and modes of locomotion such as hopping or flying. (McDowell 2008a, 5)

In other words, whilst bird might not be part of the conceptual structure of an intuition, perching animal might be. But the selection of such concepts seems far from simple. It cannot be read off from either the worldly facts seen nor the judgements about them grounded in experience. Hence it seems to be part of an explanatory theory of perception rather than the dissolution of a sense of paradox about the harmony of thought and reality.

Second, the connection between perceptual experience and judgement is also no longer simple. McDowell contrasts the transparent relation (again with echoes of the Myth of Jones) between inner judgement and spoken discourse with the more complex connection between judgement and the intuitional unity of perceptual experience.

The unity of intuitional content is given, not a result of our putting significances together. Even if discursive exploitation of some content given in an intuition does not require one to acquire a new discursive capacity, one needs to carve out that content from the intuition’s unarticulated content before one can put it together with other bits of content in discursive activity. Intuiting does not do this carving out for one. (McDowell 2008a, 5)

Even when the content of an explicit judgement can be articulated in words with no demonstrative element (such as ‘that shade of red!’), still the linguistic articulation needs to be “carved out” of a unity which has a quite distinct form. (Hence the word ‘given’ does more than merely reflect being struck by a thought: something with propositional unity.) But whatever such “carving” is, it interrupts a truistic connection between worldly facts, experience and judgements.

This might suggest a rationale for returning to the apparently more innocent picture in Mind and World. But even if that picture could be stabilised against the need to make the changes McDowell has, in fact, proposed, it still implies that the world itself inherits the conceptual structure of thought. Harmony of thought and reality is underpinned by making reality thought-like. That suggests that the price of this philosophical account is subscribing to what McDowell describes as “Absolute Idealism” (McDowell 1994, 44). Elsewhere he offers the following sympathetic sketch of “German Idealism”:

Kant – to resort to a thumbnail caricature – established that the world (something that consists of things that are the case …) cannot be constitutively independent of the space of concepts, the space where subjectivity has its being. Kant himself preserved a residual role for the idea of something more brutally alien to the realm of thought than that, something that co-operates with mind in the transcendental constitution of the world. But Kant’s successors saw (in one way or another) that the fundamental thesis, that the world cannot be constitutively independent of the space of concepts, does not require
But although he insists it need not be idealistic in any obvious sense, it is hard to see why not. At the very least, it leaves as something of a mystery the relation between the world understood as everything that is the case, the set of true Fregean thoughts and the world as made up of objects and properties (cf. Dodd 1995). If as quoted above, thought and reality meet in the realm of sense, that idea alone does not seem to connect thought and the world of things or referents. That is sufficient reason to see whether there is an alternative account of the harmony of thought and reality that comes at a cheaper price.

**Wittgenstein on the Harmony of Thought and Reality**

In the passage set out at the start of this chapter, McDowell endorses Wittgenstein’s remark that “When we say, and mean, that such-and-such is the case, we – and our meaning – do not stop anywhere short of the fact; but we mean: this – is – so.” Wittgenstein adds that this can seem paradoxical because “thought can be of what is not the case” (Wittgenstein 1953, §95). As I have described, however, influenced by his disjunctivism, McDowell plays down this aspect of thought’s connection to reality. Veridical and illusory cases call for different treatment.

By contrast, as some later remarks in the *Philosophical Investigations* that pick up a theme from the passage McDowell quotes make clear, Wittgenstein’s discussion of the harmony of thought and reality is symmetric with respect to truth and falsity.

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. [128] And in this: that if I want to explain the word “*red*” to someone, in the sentence “That is not *red*”, I do so by pointing to something that *is* red. (Wittgenstein [1953] 2009, §429)

Light can be shed on this surprising claim by looking at the discussion of the connection between an expectation and what satisfies it a few remarks later.

I see someone aiming a gun and say “I expect a bang.” The shot is fired. – What! – was that what you expected? So did that bang somehow already exist in your expectation? Or is it just that your expectation agrees in some other respect with what occurred; that that noise was not contained in your expectation, and merely supervened as an accidental property when the expectation was being fulfilled? – But no, if the noise had not occurred, my expectation would not have been fulfilled; the noise fulfilled it; it was not an accompaniment of the fulfilment like a second guest accompanying the one I expected. Was the feature of the event that was not also in the expectation something accidental, an extra provided by fate? – But then, what was *not* an extra? Did something
of the shot already occur in my expectation? – Then what was extra? for wasn’t I expecting the whole shot.

“The bang was not as loud as I had expected.” – “Then was there a louder bang in your expectation?” (Wittgenstein [1953] 2009, §442)

The paragraph explores the conflicting temptations both to assert and to deny that the bang somehow already existed in the expectation. Unless it did, then it seems that it, that very bang, was not what was expected. But not everything about the actual event was anticipated which suggests that it was not somehow contained. Further, the case of a frustrated expectation, such as the expectation of a louder bang than actually happens, suggests that the expectation somehow contains a second, louder bang than the actual bang.

These conflicting temptations suggest a dilemma which generalises from the satisfaction of expectations to other propositional attitudes and to the truth of beliefs.

- If the contents of expectations or other propositional attitudes that are satisfied, or of true beliefs, are constituted by the very facts (possibly in the future) they are about, what are the contents of false thoughts or unsatisfied expectations or other propositional attitudes?
- If the contents of expectations that are not satisfied, or of false beliefs, are free-standing mental states constitutively independent of worldly states of affairs, how do such thoughts ever bear on, or connect to, worldly facts?

The difficulty with the second horn of the dilemma reflects the apparently conflicting demands on understanding the meaning of a word which is discussed earlier in the Philosophical Investigations (Wittgenstein [1953] 2009, §139 ff.). Such understanding can occur in a flash but also governs a potentially unlimited number of subsequent uses of the word. The problem highlighted there is that the sort of freestanding objects that look to be good candidates for what can come to mind do not seem to possess the right normative properties to govern subsequent correct use of words. Pictures, for example, stand in need of particular rules of projection. But such rules can only themselves be depicted in another mental picture on pain of vicious regress.

The dilemma also captures the awkward combination of truism and paradox Wittgenstein ([1953] 2009) describes in §95. It seems obvious and truistic that a successful expectation or a true belief concerns the fact that satisfies or makes it true. But that idea is made paradoxical by unsatisfied expectations or false beliefs: they cannot inherit their content by simply standing in for non-existent worldly events or non-obtaining facts.

Wittgenstein’s next paragraph suggests that whatever response he will offer to the dilemma will be, unlike McDowell’s, symmetric to truth and falsity. Changing from the case of expectations to imagining, he says:
“The red which you imagine is surely not the same (not the same thing) as the red which you see in front of you; so how can you say that it is what you imagined?” – But haven’t we an analogous case with the sentences “Here is a red patch” and “Here there isn’t a red patch”. The word “red” occurs in both; so this word can’t indicate the presence of something red. (Wittgenstein [1953] 2009, §443)

As Arrington comments, this passage seems to go too fast (Arrington 1991, 181). The interlocutor stresses the apparent distinction between an object of thought and a worldly feature; Wittgenstein replies with a baffling analogy involving two contrasting assertions. So why should one think that colours in the imagination must be different to physical coloured samples? In this kind of case there is a plausible motivation. Imagining the colour red might be assumed to involve having a red mental image before the mind’s eye. If so, the question of whether the red mental image is the same colour as a worldly sample of redness is pressing.

The response, however – the suggested “analogous case” – appears to concern something quite different, the meaning of the word ‘red’ in two contrasting sentences albeit with an echo of the claim of what the harmony of thought and reality comprises in §429 (Wittgenstein [1953] 2009). But as Arrington highlights, the second sentence – ‘Here there isn’t a red patch’ – functions no less than the previous sentence but without the presence of anything red (Arrington 1991, 181). And hence the suggested diagnosis: if imagining something red is modelled on a sentence about redness rather than a red mental image, the problem of comparing inner and outer reds can be side-stepped. This line of thought is clearer in the next pair of remarks.

One may have the feeling that in the sentence “I expect he is coming” one is using the words “he is coming” in a different sense from the one they have in the assertion “He is coming”. But if that were so, how could I say that my expectation had been fulfilled? If I wanted to explain the words “he” and “is coming”, say by means of ostensive explanations, the same explanations of these words would go for both sentences. 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 [131] 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?! (Wittgenstein [1953] 2009, §444)

It is in language that an expectation and its fulfilment make contact. (Wittgenstein [1953] 2009, §445)

Again the passage concerns the connection between a mental state and a worldly event that satisfies or fulfils it. This time, the initial worry is with the content of a first person linguistic expression of an expectation and a description of an event that satisfies it. But why might one think that the
phrase ‘he is coming’ has two different senses depending on whether it is part of the former or an assertion concerning the latter? One reason would be if one thought that in the two uses, it stood for two different things: a merely mental object and a worldly event. This would echo the previous difference between colour as imagined and actual sample.

Two considerations count against this assumption, however. First, there has to be a connection between the meaning of the two phrases if the assertion can be used to report whether the expectation was fulfilled or not. Second, the same explanations of meaning – for ‘he’ and ‘is coming’ – can be used for the phrase ‘he is coming’ in either context. And hence, jumping ahead, the connection between the expectation and what fulfils it is not an independent explanation of the re-use of the same fragment of language. Rather, the connection holds in virtue of the re-use of the same linguistic phrase. ‘He is coming’ can be used both to individuate the expectation and to report the event or fact that satisfies it.

This, however, prompts a further question of justification. In virtue of what is it right to use the same phrase both to describe an expectation and to report an event given that the one sequence of events has not the “slightest similarity” to the other? Wittgenstein’s suggestion is that a subject may spontaneously avow an expectation. Hence the justification of a description of an expectation using the phrase ‘he is coming’ starts with its being expressed using these same words. This however leaves a question over the interpretation of both this thought and the summary: it is in language that an expectation and its fulfilment make contact. Does the avowal constitute the expectation as the expectation it is or does it report it? And if it reports it, in virtue of what antecedent fact is such a report correct, or not? Likewise, is the linguistically mediated link between expectation and fulfilment a matter of description or construction?

One reason to be wary of answering ‘description’ is that that seems to push the line of inquiry back to the nature of the mental object that merits the description ‘expecting he is coming’ and hence, again, the question of the sameness of the red image and the red sample. This complements the moral of the discussion of meaning and understanding, earlier in the "Philosophical Investigations", which suggests that it would be hard to find free-standing mental objects with the right normative properties to explain the link between mental state and what fulfils it.

Wittgenstein himself at least considers the possibility that the expression of a propositional attitude forges or constructs the connection to what would satisfy it.

“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? (Wittgenstein [1953] 2009, §682)

I draw a head. You ask, “Whom is that supposed to represent?” – I: “It’s supposed to be N.” – You: “But it doesn’t look like him; if anything, it’s rather like M.” – When I said it represented N., was I making a connection, or reporting one? And what connection was there? (Wittgenstein [1953] 2009, §683)

But the idea that an avowal constructs the content of a propositional attitude also faces difficulty. For one thing, it matters what one means or intends by the words used for the avowal. As the discussion from §139 highlights, such meaning or intending looks to be just as puzzling an idea as the propositional attitudes that avowal is supposed to underpin and hence clarify.

The balance between the temptation to answer ‘description’ and ‘construction’ can itself be illuminated by looking away from Wittgenstein’s discussion to a surprisingly similar set of ideas albeit articulated within a different philosophical context: Donald Davidson’s account of propositional attitudes.

Davidson on Propositional Attitudes and Their Objects

Davidson’s roughly parallel account is set out in his paper ‘What is present to the mind?’ (Davidson 1989). It starts with a qualification about what is meant by something coming before or being present to the mind which, also, has echoes of Wittgenstein. Davidson suggests that there are two senses.

There is a sense in which anything we think about is, while we are thinking about it, before the mind. But there is another sense, well known to philosophers, in which only some of the things we can think about are said to be before the mind. (Davidson 2001, 53)

The latter are objects of thought postulated in philosophical accounts or explanations of thought. They are supposed to play an explanatory role shedding light on the former ordinary sense in which thoughts can be said to be before the mind. Davidson suggests that objects of thought of this second, philosophically explanatory sort are supposed to play two roles.

They identify a thought by fixing its content; and they constitute an essential aspect of the psychology of the thought by being grasped or otherwise known by the person with the thought. (Davidson 2001, 54)

But, he argues, these two roles cannot be reconciled. The problem is that, whatever its precise nature and limits, subjects typically have authority over the content of their own mental states. If, however, their content is fixed by an object which has to be known to the thinker, then to know the content of one’s own thoughts requires that one knows which object is before the mind. Davidson’s objection concerns the potential mismatch between first person authority and the fallibility of object recognition.
The trouble is that ignorance of even one property of an object can, under appropriate circumstances, count as not knowing which object it is. (Davidson 2001, 54)

If the mind can think only by getting into the right relation to some object which it can for certain distinguish from all others, then thought is impossible. If a mind can know what it thinks only by flawlessly identifying the objects before it, then we must very often not know what we think. (Davidson 2001, 56)

Davidson attempts to bolster this line of argument through a discussion of de re beliefs, rigid designation and singular thought offering the following diagnosis:

What lies behind some of these attempts to characterize the special relation between the mind and its objects is, of course, the Cartesian drive to identify a sort of knowledge which is guaranteed against failure. If this search is combined with the assumption that all knowledge consists in the mind being in psychological contact with an object, then objects must be found about which error is impossible—objects that must be what they seem and seem what they are. (Davidson 2001, 56)

This seems to me to be a misstep in the context of Evans and McDowell’s accounts of singular thought, at least. Rejecting the Cartesian divide, the objects they invoke to play a role in constituting singular thoughts are everyday worldly objects which could thus be misrecognised.

A typical visual experience of, say, a cat situates its object for the perceiver: in the first instance egocentrically, but, granting the perceiver a general capacity to locate himself, and the objects he can locate egocentrically, in a non-egocentrically conceived world, we can see how the experience’s placing of the cat equips the perceiver with knowledge of where in the world it is (even if the only answer he can give to the question where it is, is “There”). In view of the kind of object a cat is, there is nothing epistemologically problematic in suggesting that this locating perceptual knowledge of it suffices for knowledge of which object it is (again, even if the only answer the perceiver can give to the question is “That one”). So those visual experiences of objects that situate their objects can be made out to fit the account I suggested of the notion of acquaintance: abandoning Russell’s sense-datum epistemology, we can say that such objects are immediately present to the mind ….(McDowell 1998a, 231)

By rejecting Russell’s Cartesian assumptions about restrictions that apply to object-dependent thoughts (limiting them to sense data and some I-thoughts), McDowell, following Evans, accepts the possibility that the misrecognition of objects can lead to the failure of a subject to think the thought attempted. Perception is fallible and when it fails it can undermine the first person authority about the nature of the thoughts thought (in addition to their truth). But, McDowell argues, the fact that it can fail does not undermine its deliverances when it does not fail. (As in epistemic disjunctivism, the nature of the thought entertained is asymmetrically accounted for in good and bad cases.)
But the possibility of a plausible neo-Fregean account of object-dependent singular thoughts would not itself save the approach Davidson targets, which postulates explanatory objects before the mind’s eye for the whole of a propositional attitude, and not just the referential element of a singular thought. Thus it faces the objections Wittgenstein highlights in his discussion of meaning and understanding. The reason for this is that Davidson’s second role for propositional objects is supposed to explain the ordinary sense of having something before the mind’s eye. Hence any such object would have to explain the normative connection between it and the intentional object of the propositional attitude as pre-philosophically understood. But as Wittgenstein argues, nothing potentially explanatory that could come to mind would have the right normative properties to match meaning and understanding. Similarly, an explanation of how a propositional attitude determines what satisfies it has to capture the normativity of the link. But short of an account that presupposes a link to the description of the fulfilment condition, it is not clear how that can come about. Davidson stresses a worry about how to identify internal mental objects. Wittgenstein argues that free-standing mental objects could not sustain the normativity of thought. The objections are different but both target the second of Davidson’s roles for propositional objects.

Davidson’s response to his own critical arguments is to accept the first role of mental or propositional objects and reject the second. The argument for retaining the first element is that the analysis of propositional attitude ascriptions looks to be relational. A subject is related to a propositional object. Attempts, for example, at adverbial accounts seem implausible. But, Davidson suggests, this point can float free of the idea that a subject stands in an explanatory psychological relation to some sort of mental object.

It does not follow, from the facts that a thinker knows what he thinks and that what he thinks can be fixed by relating him to a certain object, that the thinker is acquainted with, or indeed knows anything at all about the object. It does not even follow that the thinker knows about any object at all. Someone who attributes a thought to another must … relate that other to some object, and so the attributer must, of course, identify an appropriate object, either by pointing to it or describing it. But there is no reason why the attributer must stand in any special relation to the identifying object; all he has to do is refer to it in the way he refers to anything else. We specify the subjective state of the thinker by relating him to an object, but there is no reason to say that this object itself has a subjective status, that it is “known” by the thinker, or is “before the mind” of the thinker. (Davidson 2001, 58)

An analogy with the ascription of weights is deployed to clarify the idea. The ascription of propositional attitudes to people functions like the ascription of weights to objects. Objects stand in various relations of the form: weighing more than, weighing less than, weighing twice as much as. For simplicity, these relations and ratios can be represented by the use of a standard weight. This
enables weights to be ascribed to objects directly using numbers. Thus one can say of an object that it weighs 5 kg. But this does not require the addition of kilograms into ontology in addition to weighty objects. On this picture, numbers are in no sense intrinsic to the objects that have weight nor are they part of them. Rather, numerical weights are used to keep track of the more basic relations between objects.

In accord with the analogy, attributing propositional attitudes to subjects is a matter of relating them to “suitable entities” which have the right relations to enable one “to keep track of the relevant properties of and relations among the various psychological states” (Davidson 2001, 60). As a student of Quine, Davidson expresses some concerns about thinking of these entities as sentences or propositions – though he reports these as “only marginal objections” – and suggests instead utterances. These have the right properties because:

Utterances are related to each other in much the same way beliefs are: by relations of entailment and evidential support. Utterances, like beliefs, are true and false. Aside from complications due to indexical elements, we identify a belief by uttering a sentence that has the same truth conditions as the belief it is used to identify. (Davidson 2001, 64)

But his concerns about invoking propositions are only marginal provided that, in accord with rejecting the second role for mental objects, subjects are not assumed to have to stand in explanatory psychological relations to them. This explains what looks to be too strong a summary: “in thinking and talking about the beliefs of people we needn’t suppose there are such entities as beliefs” (Davidson 2011, 60). There are such entities as beliefs. It is rather that they are not constituted by inner mental objects. (Similarly, the moral from Wittgenstein’s discussion of meaning is not that meanings cannot come to mind but that they do not come to mind via free-standing inner mental objects [McDowell 1984].)

As I hope this summary has shown, there are similarities between Wittgenstein’s and Davidson’s accounts of the linguistic mediation of propositional attitudes and their fulfilment conditions. Translating from Wittgenstein to Davidson helps to address the question left hanging at the end of the previous section, however. Does a particular avowal construct or report a propositional attitude? Is the general linguistic mediation between attitude and fulfilment constructive or descriptive?

Davidson’s stress on relations helps to show why ‘description’ is the right answer. The account in ‘What is present to the mind?’ builds on his principled account elsewhere of what constitutes states as mental states (Davidson 1984). The nature of mental states, together with the meaning of utterances, are both approached through the thought experiment of radical interpretation: interpretation from scratch. A radical interpreter is faced with two unknowns: what a speaker thinks and what she means. These are connected in that given
knowledge of what she thinks, her words can be interpreted and given knowl-
edge of her meanings, her beliefs can be identified. But neither can be assumed
prior to radical interpretation.

Davidson argues that it is necessary to restrain the degrees of freedom of
possible beliefs in order to interpret linguistic meaning. The interpreter must
impose his or her own standards of truth and coherence on ascriptions of beliefs
and meanings. There must be a presumption that many utterances or beliefs
held true are true, that beliefs are structured in accordance with basic logic,
probability theory and decision theory and that they are often caused by
features of the external environment (Davidson 1990, 326–328).

This complex of related assumptions governing the rationality imputed –
genGenerally briskly labelled the “Principle of Charity” – enables interpretation to
get off the ground. If utterances and underlying beliefs are assumed by the
interpreter to be generally true, rationally structured and to concern the worldly
states of affairs that prompt them, then they can be correlated with those
observed states of affairs and their meaning determined.

But Davidson goes further. He argues that the facts available to radical
interpretation are the only facts about meaning there are. After all, the only
justification that can be offered for knowledge of a first language – the only
potentially contrasting case to the second language explored by ethnographic
anthropologists – depends on facts available from the radical interpretation of
the contextually located utterances of kith and kin. Since access to the only
facts there are about meaning has to be mediated by the Principle of Charity,
this is not merely an epistemological shortcut. It reflects an ontological feature
of belief content and linguistic meaning itself. Both are governed by
a constitutive principle: the “Constitutive Ideal of Rationality” (Davidson
1970, 223). Belief and meaning (both facets of human intentionality) are
essentially governed by rationality.

The pattern in speech and action that, according to the account in ‘What is
present to the mind?’, is tracked by relating subjects to utterances (or, possibly,
propositions) is the inherent rational structure of thoughts and meanings. It,
according to Davidson’s account of radical interpretation, is an essential feature
of mentality. This suggests the following combination of necessities and con-
tingencies. Meaning and mental content have an essential rational structure.
Propositional attitudes are essentially tied to linguistic descriptions. But it is
a contingency that speakers are able to self-ascribe attitudes on the basis of no
evidence. This is not the result of an explanatory psychological relation to
a mental object but a contingent primitive ability to relate oneself, and one’s
subsequent actions, to the rational structure on which Radical Interpretation
rests. The lack of inner epistemology suggests the potential attraction of
a constructive, rather than descriptive view, but the broader picture shows
why it should be rejected.
The harmony of Thought and Reality: Wittgenstein vs McDowell

The detour via Davidson’s similar account of propositional attitudes suggests the following account implicit in Wittgenstein’s more fragmentary remarks.

1. The words that comprise the characteristic expression of a mental state can be converted into a description of its fulfilment condition.
2. Mental states are individuated by their fulfilment conditions.
3. The connection between a mental state and its fulfilment is reflected in the linguistic connection between the characteristic expression of a state, the description of the state and the description of its fulfilment condition.
4. The ability to avow propositional attitudes is a primitive contingent ability, not based on the recognition of an inner object with the right normative properties.

The difficulties of comparing mental objects and actual events are side-stepped by the connection between propositional attitudes and the language used to individuate them. A state, for example of expectation, is the state it is in virtue of its having the fulfilment condition it has. But as Wittgenstein points out, this applies more broadly: for example, to orders.

“An order orders its own execution.” So it knows its execution, then, even before it is there? – But that was a grammatical proposition and it means: If an order runs “Do such-and-such” then executing the order is called “doing such-and-such.” (Wittgenstein [1953] 2009, §458)

The connection between propositional attitudes and language allows for an account which is symmetric between truth and falsity. Since nothing has to play Davidson’s second role, there is no puzzle as to how propositional attitudes can concern events that do, or do not, occur. And hence the initially surprising claim that 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” highlights the fact that way thoughts are identified is independent of whether or not they are fulfilled or made true by events. As P.M.S. Hacker summarises this symmetry:

The agreement or harmony between thought and reality is not an agreement of truth (satisfaction, obedience). It is an agreement that necessarily obtains (is pre-established) between thought and reality no matter whether the thought is true or false, the wish fulfilled or not fulfilled, the order obeyed or disobeyed. (Hacker 2013, 171)

A related discussion in Zettel makes the nature of the account more explicit:

Like everything metaphysical the harmony between thought and reality is to be found in the grammar of language. (Wittgenstein 1981, §55)
This remark highlights just how different Wittgenstein’s and McDowell’s accounts of the harmony of thought and reality are. Using Hacker’s phrase, McDowell’s account *is* an agreement of truth, satisfaction and obedience. True thoughts agree with reality in virtue of the shared content from judgements, via experiences (at least in the *Mind and World* account), to the world, construed as the totality of true thoughts. The shared content links thought and reality. The price of this account is that the world has to be assumed to be thought-like: to possess conceptual structure. Hence McDowell’s appeal to Absolute or German Idealism. The revisionary aspects of this assumption are not the only difficulties with the picture. As I briefly described above, in response to criticism, McDowell has changed his account of experience moving away from the initial assumption that it can be closely modelled on utterance. The resulting picture is thus less truistic and therapeutic.

By contrast, Wittgenstein’s account concerns a harmony that exists whether or not thoughts are true. It depends, instead, on the way they are individuated. This may seem, therefore, not to dig deep enough. Wittgenstein *assumes* the representational powers of language to set out the connection between propositional attitudes and what fulfils them. But by drawing out the parallels between Wittgenstein and Davidson, this worry can be addressed. Radical Interpretation offers a plausible account of the worldly basis of language. (The *Philosophical Investigations* also contains materials for addressing the question of the connection between language and the world in for example the discussion of primitive language games (e.g., Wittgenstein [1953] 2009, §§2, 8, 21, 41, 42).)

Given the possibility of a therapeutic approach to the harmony of thought and reality in Wittgenstein’s own work, it is worth pinpointing the origins of the disagreement between them. I outlined four key assumptions behind McDowell’s *Mind and World*: disjunctivism, the “rational constraint constraint”, transcendental empiricism and Sellars’s Myth of Jones. McDowell’s epistemic disjunctivism is based on a plausible reading of Wittgenstein’s discussion and hence there is no reason to think that this marks a difference. McDowell and Wittgenstein also share an emphasis on the importance of conceptual relations mediating thoughts and what would fulfil them. Whilst the relaxed attitude to theorising exemplified by Sellars’ Myth of Jones is foreign to Wittgenstein (though not hypothetical and fictitious examples), the connection between thought and utterance finds parallels in the latter’s work. So the key difference is McDowell’s assumption that empiricism has to play a part in any account of thought’s bearing on the world. It is this, refracted through epistemic disjunctivism, that gives rise to a kind of harmony disjunctivism in which, in the good disjunct, the world impresses itself on thought through experience at the cost of making the world itself thought-like or linguistic.
That cost can be avoided, however, by rejecting empiricism. The comparison of Wittgenstein and Davidson helps point to a different, more therapeutic response to understanding the harmony of thought and reality. Harmony of thought and reality is not a success of thought but is guaranteed by the very idea of thought in the first place.

References


11 Davidson, Russell, and Wittgenstein on the Problem of Predication

José L. Zalabardo

1 Introduction

The last four chapters of Davidson’s posthumously published book, *Truth and Predication*, are devoted to what he calls the problem of predication. The problem encompasses a metaphysical question – “How are particulars related to universals?” (Davidson 2005, 77), and a semantic one – “How are names or other singular terms related to predicates?” (Davidson 2005, 77). Davidson follows the recent tendency of concentrating on the semantic problem.

After discussing the nature of the problem, Davidson goes on to provide a historical survey of attempts to solve it. He ends by presenting his own solution. Davidson’s historical survey has a big gap, jumping from Russell directly to Strawson, thereby omitting Wittgenstein’s contribution to the debate. The reason why Wittgenstein is left out is not that Davidson doesn’t find his ideas worth discussing. He explains the omission in a note on the manuscript added by the editors as a footnote:

My decision not to talk about Wittgenstein’s view needs a comment. The reason is simply that try as I may I cannot satisfy myself that I have a sufficiently justified opinion what his views on predication were. I lament my failure here (as no doubt elsewhere) to fill in an important piece of the picture. (Davidson 2005, 140)

In fact, the gap in Davidson’s coverage starts just before Wittgenstein, leaving out, apart from some passing remarks, Russell’s contribution after he abandoned the dual-relation theory of judgment in favour of the multiple-relation theory. Since then, Davidson tells us, “Russell did not return to the question of the unity of the proposition, since on the new theory there were no propositions left” (Davidson 2005, 107).

This is an unfortunate gap. I believe, and I hope to show here, that the exchanges between Russell and Wittgenstein, in the two years that Wittgenstein spent in Cambridge working with Russell, constitute one of the

I am grateful to Fraser MacBride for his comments on this material.
most interesting episodes of the history of the problem. My main goal in this chapter is to outline the main ideas to emerge from these exchanges regarding the problem of predication, in its metaphysical and semantic forms. I will close by discussing the nature of Davidson’s own solution to the problem.

2 The Metaphysical Problem

The bulk of Davidson’s discussion of the problem of predication deals with the semantic version, but he also devotes some attention to its metaphysical form. He presents it in the following passage:

Theaetetus sits. Theaetetus, we agree, is an entity, a person. He is sitting. The property of Sitting is another entity, this time a universal that can be instantiated by many particular entities. In the present case, Theaetetus is one of those entities. In other words, we explain what it is for it to be the case that Theaetetus sits by saying that Theaetetus instantiates the property of Sitting. The fact itself doesn’t, then, consist merely of the two entities, Theaetetus and the property of Sitting. It is a fact because those two entities stand in a certain relation to each other, the relation of Instantiation […] Now we see that a fact we can describe in just two words apparently involves three entities. But can ‘Theaetetus sits’ consist of just the three entities, Theaetetus, Sitting, and Instantiation? Surely not. The fact requires as well that these three entities stand in a certain relation to one another. To explain this fact we need to mention this fourth entity, which, unlike instantiation, is a three-place relation. We are clearly off on an infinite regress. (Davidson 2005, 85)

As Tyler Burge has pointed out, Davidson often blurs the distinction between the metaphysical and the semantical problem (Burge 2007, 589–90). This passage, for example, is presented as an attempt ‘to understand the structure and nature of judgments’ (Davidson 2005, 85). I agree with Burge that the two problems should be treated separately. I want to start in this section by considering the metaphysical problem, treating the quoted passage as expressing Davidson’s views on this.

In Davidson’s presentation, the problem concerns an explanation of ‘what it is for it to be the case that Theaetetus sits’. On this explanation, the fact is a compound, with a particular, Theaetetus, and a universal, the property of Sitting, as its constituents. The problem arises for the contention that these are the only constituents of the fact. The mere coexistence of Theaetetus and the property of Sitting doesn’t produce the fact. The fact that Theaetetus sits could fail to obtain even if both Theaetetus and the property of Sitting exist. In order for the fact to obtain, the particular has to be connected with the universal in the way that we call instantiation – Theaetetus needs to instantiate the property of Sitting. But this seems to amount to the admission that the fact that Theaetetus sits has, not two constituents, but three. In addition to Theaetetus and the property of Sitting, we also need to count the (binary) relation of Monadic Instantiation as one of its constituents.
But if this is right, the same line of reasoning can be applied to the current hypothesis concerning the composition of the fact that Theaetetus sits. The mere coexistence of Theaetetus, the property of Sitting and the relation of Monadic Instantiation doesn’t produce the fact. The three items need to be connected with one another in the way that we call binary instantiation – Theaetetus and the property of Sitting (in that order) need to instantiate Monadic Instantiation. But this brings in an additional constituent to our analysis of the fact that Theaetetus sits. In addition to Theaetetus, the property of Sitting, and the relation of Monadic Instantiation, we also need to count the (ternary) relation of Dyadic Instantiation as a constituent of the fact. And since the reasoning can be repeated at each stage, we never reach a satisfactory account of the composition of the fact that Theaetetus sits.¹

On this construal, then, the target of the argument is the application to the fact that Theaetetus sits of an account of facts according to which they are composite items, produced by the combination of particulars and universals. I am going to refer to this as the Compositional Account of Facts (CAF). The regress argument, on the construal that we are extracting from Davidson’s passage, would refute any account of the structure of the fact that Theaetetus sits that an advocate of CAF might put forward.

Let’s consider in some detail how the argument would proceed, on this construal, focusing in the first instance on the most natural application of CAF to the fact that Theaetetus sits:

A1. The fact that Theaetetus sits is a compound with exactly two components: Theaetetus and the property of Sitting.

We will have a refutation of A1 if we can show that it entails its negation. For this purpose we assume A1 and try to show that Not-A1 follows from this assumption. A1 entails that Theaetetus and the property of Sitting are among the components of the fact that Theaetetus sits. But for the fact to be produced, these two items have to be combined in a certain way – the former has to instantiate the latter, and it follows from this that Monadic Instantiation also has to be a constituent of the fact. Hence Theaetetus and the property of Sitting are not the only two components of the fact that Theaetetus sits. Since A1 entails its negation, A1 stands refuted, as desired. The reasoning from A1 to Not-A1 would have the following structure:

A1. (assumption)
B1. The fact that Theaetetus sits is a compound with Theaetetus and the property of Sitting among its components. (from A1)
C1. The fact that Theaetetus sits wouldn’t obtain unless Theaetetus instantiated the property of Sitting. (from B1)

¹ A similar difficulty would arise if we construed Instantiation as a single, multigrade relation.
D1. Monadic Instantiation is also a component of the fact that Theaetetus sits.  
(from C1)

Not-A1. (from D1)
This inference form A1 to Not-A1 is supposed to refute A1 – a specific account of the structure of the fact that Theaetetus sits. The advocate of CAF might at this point replace A1 with a different account of the composition of the fact that Theaetetus sits:

A2. The fact that Theaetetus sits is a compound with exactly three components: Theaetetus, the property of Sitting and the relation of Monadic Instantiation.

But another application of the same line of reasoning would show that A2 entails its negation:

B2. The fact that Theaetetus sits is a compound with Theaetetus, the property of Sitting and the relation of Monadic Instantiation among its components.  
(from A2)

C2. The fact that Theaetetus sits wouldn’t obtain unless Theaetetus and the property of Sitting instantiated the relation of Monadic Instantiation.  
(from B2)

D2. Dyadic instantiation is also a component of the fact that Theaetetus sits.  
(from C2)

Not-A2. (from D2)
And the same strategy can be used to refute any subsequent proposal as to how CAF can construe the fact that Theaetetus sits. It follows that there is no viable account of the fact that Theaetetus sits along the lines of CAF. In general, CAF cannot provide a viable account of facts.\(^2\)

Is this line of reasoning sound? Let’s focus on the purported refutation of A1. B1 is a logical consequence of A1. The inference from B1 to C1 strikes me as unassailable. If we construe the fact that Theaetetus sits as a compound, the production of the fact will require that its constituents are combined with one another in the right sort of way.

The point at which the argument is vulnerable is the transition from C1 to D1 – from the claim that Theaetetus would need to instantiate the property of Sitting in order for the fact to be produced to the claim that Monadic Instantiation has to be a constituent of the fact. In order to underwrite this inference, we need a principle along the following lines:

CONS. If an entity E is implicated in the constitution of a fact F, then E is a constituent of F.

\[^2\] Provided, that is, that we rule out a construal of facts as compounds with infinitely many constituents.
In the presence of CONS, the inference from C1 to D1 is unstoppable, but if CONS can be rejected, the refutation of A1 can be resisted at this point.

This is the line taken by Bertrand Russell, just at the point at which Davidson interrupts his historical survey. According to Russell, every complex must have what he calls a form, which is the way in which the constituents of the complex are put together. Forms, for Russell, are real entities, not mere abstractions. However, they are not constituents of the complexes whose forms they are, and this, precisely, in order to avoid the regress we are discussing here. He expresses the point in the following passage of the manuscript ‘What is Logic?’, of 1912:

In a complex, there must be something, which we may call the form, which is not a constituent, but the way the constituents are put together. If we made this a constituent, it would have to be somehow related to the other constituents, and the way in which it was related would really be the form; hence an endless regress. Thus the form is not a constituent. (Russell 1992, 55)³

Since Russell’s complexes include facts, Russell is here rejecting CONS. He accepts that forms are implicated in the production of complexes, but he refuses to treat them as their constituents.

Once we reject CONS, the argument against CAF can be easily blocked. We will be able to recognise the involvement of Monadic Instantiation in the constitution of the fact that Theaetetus sits without contradicting A1. Three entities will be involved in the constitution of the fact – its two constituents and the way in which they are put together, Monadic Instantiation, which is not a constituent. Advocates of CAF who are prepared to follow Russell in rejecting CONS, have nothing to fear from this version of the regress argument.

Now, if the fact that Theaetetus sits obtains, there will also be the fact that Theaetetus instantiates the property of Sitting. This fact will have Theaetetus, the property of Sitting and the relation of Monadic Instantiation as its constituents. These constituents will be put together by Dyadic Instantiation, which will not be a constituent of this fact. And in the same way we can describe an infinite sequence of ever more complex facts, all of which will obtain just in case the fact that Theaetetus sits obtains:

\[
\begin{align*}
F1. \; & S(t) \\
F2. \; & \in_1(S,t) \\
F3. \; & \in_2(\in_1,S,t) \\
F4. \; & \in_3(\in_2,\in_1,S,t) \\
\vdots
\end{align*}
\]

But all of these facts will be different from one another, not conflicting accounts of the structure of one and the same fact, as required by the argument.

³ See also Russell (1984, 98).
against CAF. Furthermore, by accepting that on her account every fact brings an infinite sequence of facts of this kind in its train, the advocate of CAF wouldn’t be accepting that her position is invalidated by an infinite regress. Burge puts the point well: “Such an infinity is profusion, not circularity or regress” (Burge 2007, 592).

So far I have argued that the version of the regress argument that I have extracted from Davidson’s passage fails to put genuine pressure on CAF, unless it is supplemented by a cogent defence of CONS. I want to explore next a different version of the argument that doesn’t make it depend on CONS.

A clue to the construal that I want to consider is given by Burge, in the continuation of the passage I’ve just quoted: “Circularity and regress are properties of explanations—of certain representational structures—not of ordinary facts or subject matters” (Burge 2007, 592). The point concerns the sequence of facts, F₁, F₂, … I have just argued that CAF is not undermined by the admission that the obtaining of F₁ entails the obtaining of the infinitely many facts in the sequence. Using this sequence in a cogent regress argument against CAF would require showing that there is a phenomenon that CAF needs to explain but any attempt to provide this explanation produces a never-ending chain involving this sequence. This is the line that I want to explore next.

Suppose that, as a matter of fact, Theaetetus sits. Now, the existence of Theaetetus and the property of Sitting doesn’t entail that Theaetetus sits. Consider, then, a possible state of the world in which both Theaetetus and the property of Sitting exist, but Theaetetus doesn’t sit. Call this the ET-alternative (for Erect Theaetetus). Which difference between actuality and the ET-alternative explains the fact that Theaetetus sits in the former but not in the latter? According to CAF, the fact that Theaetetus sits is a compound, with Theaetetus and the property of Sitting as its constituents. The constituents exist in the ET-alternative, but the fact doesn’t obtain. Why is this?

Clearly the answer that CAF would give to this question is that in actuality, but not in the ET-alternative, Theaetetus and the property of Sitting are combined in the requisite way – the former instantiates the latter. So according to CAF, the reason why the fact that Theaetetus sits obtains in actuality but not in the ET-alternative is that Theaetetus instantiates the property of Sitting in actuality and not in the ET-alternative. This explanans invokes a fact – the fact that Theaetetus instantiates the property of Sitting – that obtains in actuality but not in the ET-alternative. This fact, as we have learnt in our discussion of the previous version of the argument, should not be treated as identical with the fact in the explanandum. The relation of Monadic Instantiation is a constituent of the fact that Theaetetus instantiates the property of Sitting, but not of the fact that Theaetetus sits. What is happening is that the first fact in our infinite sequence, F₁, is being explained in terms of the second fact in the sequence, F₂. The difference between actuality and the ET-alternative that explains why
F1 is present in the former but not in the latter is that another fact, F2, is present in the former but not in the latter.

This can’t be the end of the story. CAF can’t treat the fact that F2 is present in actuality but not in the ET-alternative as a brute difference between the two situations. F2 is a fact. Hence, according to CAF, it is a compound. But its constituents – Theaetetus, the property of Sitting and the relation of Monadic Instantiation – are also present in the ET-alternative, in which F2 doesn’t obtain. So far, we’ve been told that the reason why the F1-compound is present in actuality but not in the ET-alternative is that the F2-compound is present in actuality but not in the ET-alternative. Until we have an explanation of why the F2-compound is present in actuality but not in the ET-alternative, we won’t have a satisfactory account of the difference between actuality and the ET-alternative that explains why F1 is present in the former but not in the latter.

But clearly this line of reasoning launches us on an infinite regress. The reason why F2 is present in actuality but not in the ET-alternative is that its constituents are combined in the right sort of way – that Theaetetus and the property of Sitting instantiate Monadic Instantiation. This amounts to explaining why F2 is present in actuality and not in the ET-alternative in terms of the fact that F3 is present in the former but not in the latter … At no point in this sequence of explanations will CAF offer a satisfactory account of why F1 is present in actuality and not in the ET-alternative. CAF has failed to identify the difference between these situations that explains why one contains F1 and the other one doesn’t.

Needless to say, this brief sketch doesn’t amount to a cogent refutation of CAF, but it does suggest that the difficulty is not removed by Russell’s idea that what combines the constituents of a fact into a unit is not itself a constituent. This is not the place to continue this dialectic. What I’d like to do before leaving the metaphysical version of the problem of predication is to outline briefly two alternatives to CAF for which the problem might not arise.

One alternative metaphysical picture is the Aristotle-inspired view that the fundamental constituents of the world are not particulars or universals, but a type of entity combining the two – not Theaetetus, on the one hand, and the properties that he instantiates, on the other, but a single entity incorporating Theaetetus and all his properties. On this view, the fact that Theaetetus sits wouldn’t be a compound of a multiplicity of components, but an aspect of this unitary whole.4 I don’t propose to discuss this approach in any detail here. I only want to point out that even if we accept that the view provides

---

4 Davidson ascribes to Aristotle the view that ‘if the forms are to serve as universals, then they cannot be separate from the entities of which they are properties’ (Davidson 2005, 89), but not the ontology of ‘thick particulars’ that I have sketched here. I am not claiming that the ‘Aristotle-inspired view’ was actually held by Aristotle.
a satisfactory account of subject-predicate facts, relational facts might still pose a problem. Suppose that Theaetetus with all his properties is a single fundamental unit, and Socrates with all his properties another. It’s hard to see what account we can provide with these materials of, say, the fact that Theaetetus criticises Socrates. For relational facts we seem forced to return to the conception of facts as compounds, facing once again the metaphysical version of the problem of predication.

The second alternative to CAF that I want to consider is a metaphysical picture that takes as fundamental items, not particulars and universals, like CAF, or the particular-cum-universal units of the Aristotelian view, but facts. Facts on this approach are the ultimate, simple constituents of reality, with particulars and universals construed as abstractions from these – as what we see the fact that Theaetetus sits as having in common with the fact that Theaetetus eats, or with the fact that Socrates sits. The approach has been eloquently characterised by Kenneth Olson as

the view that—contrary to what language would lead us to believe—the sort of thing expressed by a whole sentence is metaphysically more basic than some of the things expressed by the words making it up. (Olson 1987, 40)

And, as I have argued elsewhere (Zalabardo 2015, Ch. 4), this is the metaphysical approach put forward in Wittgenstein’s *Tractatus* (Wittgenstein 1974). This fact-based ontology is, precisely, what the opening sentences of the book express:

1 The world is all that is the case.
1.1 The world is the totality of facts, not of things.

This is not the standard interpretation of the metaphysics of the *Tractatus*, but it’s certainly not without advocates. It received an early endorsement from Brian Skyrms:

Wittgenstein’s truly daring idea was that the ontology of the subject (nominalism) and the ontology of the predicate (Platonism) were both equally wrong and one-sided; and that they should give way to the ontology of the assertion. We may conceive of the world not as a world of individuals or as a world of properties and relations, but as a world of facts—with individuals and relations being equally abstractions from the facts. John would be an abstraction from all facts-about-john; Red an abstraction from being-red-facts; etc. (Skyrms 1981, 199)

The point is not simply that particulars and universals cannot occur in isolation, but only in the context of states of affairs. The claim is rather that, for Wittgenstein, particulars and universals are ultimately nothing but features of states of affairs – John is what the states of affairs that we characterise as involving John have in common, and redness what the states of affairs that we

---

characterise as involving redness have in common. I think this is a very plausible reading of the Tractatus, but a defence of this point would require a more extensive discussion of the interpretation of the book than we can undertake in this chapter. Hence we will have to restrict ourselves to a conditional conclusion: if this reading is right, as I think it is, then Wittgenstein doesn’t face the metaphysical version of the problem of predication. Wittgenstein doesn’t have to provide an explanation of how facts arise from the combination of their components because he doesn’t think that facts are produced by a process of combination. On the contrary, facts are ultimate, indivisible units of his ontology, in which what we think of as constituents of facts are treated as mere abstractions.

3 The Semantic Problem

I want to turn now to the semantic version of the problem of predication. The sentence “Theaetetus sits” represents things as being a certain way, which may or may not coincide with the way things are. How does the sentence achieve this? The most promising strategy for answering this question is to explain the power of the sentence to represent the world in terms of the semantic properties of its constituents – the singular term ‘Theaetetus’ and the predicate ‘sits’. The semantic version of the problem of predication is to understand how the semantic properties of its constituents enable a sentence to represent the world as being a certain way. On the assumption that the role of singular terms is relatively clear, the problem is seen as concerning, specifically, how the semantic properties of predicates can give rise to the representational power of the sentences in which they figure.

Russell faced a version of the problem in his theories of judgment. Judgments, for Russell, are mental episodes in which we represent the world in consciousness as being a certain way. With his theories of judgment, Russell sought to explain how the mind is related to the world when we judge. A crucial challenge for this enterprise is posed by the fact that judgments can be false as well as true. This circumstance blocks an account according to which in a judgment the mind is related to a fact. The proposal would not work for false judgments. I can judge that Theaetetus flies, but my doing so cannot consist in a relation between me and the fact that Theaetetus flies, since there is no such fact.

Russell’s first strategy for overcoming this obstacle was to postulate a range of entities to act as what I am related to when I judge falsely, in the same way in which I am related to a fact when I judge truly. He referred to these items as

---

6 The position I am attributing to Wittgenstein is akin to versions of trope theory that construe particulars and universals as bundles of tropes. See, e.g., Bacon (1995).
objective non-facts, and to facts and objective non-facts together as propositions. The view that judgment is a relation between the mind and a proposition, in this sense, is, in a nutshell, Russell’s dual-relation theory of judgment.

But by 1910 Russell had abandoned this strategy and had replaced it with an alternative approach that takes advantage of the resources of CAF. On the new approach, there isn’t a single item to which my mind is related when I judge. On the contrary, my mind is related to a multiplicity of items – the constituents of the fact that would have to obtain in order for my judgment to be true. When I judge that Theaetetus sits, my mind is related to Theaetetus and to the property of Sitting; when I judge that Theaetetus flies, my mind is related to Theaetetus and to the property of Flying. My judgment that Theaetetus sits is true because there is a fact with Theaetetus and the property of Sitting as its constituents. My judgment that Theaetetus flies is false because there isn’t a fact with Theaetetus and the property of Flying as its constituents. This is the central thought of Russell’s multiple-relation theory of judgment.

This basic idea had to be refined in several ways before it could provide a satisfactory theory, and between 1910 and 1913 Russell endorsed, in quick succession, several versions of the view. One of the challenges that precipitated these changes was the problem of order, or direction. On the multiple-relation theory, when Othello judges that Desdemona loves Casio, his mind is related to Desdemona, Casio, and the relation of Love. But we can’t say that Othello’s judgment is true just in case there is a fact with Desdemona, Casio, and the relation of Love as its constituents. Casio loving Desdemona would be a fact with these constituents, but we wouldn’t want to say that the obtaining of this fact would suffice for making Othello’s judgment true.

What matters for our purposes is a different challenge faced by the theory, one that Russell didn’t fully grasp until the 1913 manuscript entitled Theory of Knowledge. The problem arises from the fact that bringing to consciousness Theaetetus and the property of Sitting does not suffice for judging that Theaetetus sits. A mental episode in which I bring these two items to consciousness might fail to represent things as being a certain way. In particular, it might fail to represent Theaetetus as sitting.

In Theory of Knowledge, the issue is discussed in its application, not to judgment, but to understanding – in our example, to the understanding of what someone says when they say that Theaetetus sits. Russell states repeatedly what would be missing from an account according to which this understanding consists in bringing to consciousness Theaetetus and the property of Sitting. In order to understand what someone says when they say that Socrates precedes Plato, Russell tells us, it is necessary to understand ‘how Socrates and Plato and

---

7 See Candlish (1996) for an account of Russell’s growing awareness of this problem as distinct from the problem of order.
“precedes” are to be combined’ (Russell 1984, 99). Understanding the statement ‘a is similar to b’ ‘would not be possible unless we knew how they [a, b and similarity] are to be put together’ (Russell 1984, 101). In order to understand the proposition ‘A precedes B’, in addition to knowing what is meant by the words that occur in it, ‘it is also necessary to know how these three terms [A, B and preceding] are to be combined’ (Russell 1984, 111). And in order to understand ‘A and B are similar’ ‘we must know what is supposed to be done with A and B and similarity, i.e. what it is for two terms to have a relation’ (Russell 1984, 116).

Returning to our example, in order to understand the proposition that Theaetetus sits, in addition to bringing to consciousness Theaetetus and the property of Sitting, I must grasp how these two items are represented as combined by the proposition – how they would have to be combined in order for the proposition to be true, in order for things to be as the proposition represents them as being. I am going to refer to this as the mode-of-combination problem. It is the problem of explaining how, in an episode of understanding, the subject is aware of the way in which the items the proposition is about are represented by the proposition as combined with one another – of how these items would have to be combined with one another in order for the proposition to be true.

There’s one approach to the mode-of-combination problem that generates an infinite regress parallel to the regress that we face in the metaphysical version of the problem of predication. On this approach, modes of combination are entities that we are capable of bringing to consciousness. We can bring to consciousness, for example, the relation of Monadic Instantiation, and the mode-of-combination problem is solved for subject-predicate propositions by reference to awareness of Monadic Instantiation. On this approach, I understand the proposition that Theaetetus sits by bringing to consciousness not only Theaetetus and the property of Sitting, but also the relation of Monadic Instantiation. By grasping this relation, I become aware of how Theaetetus and the property of Sitting are represented as combined with one another by the proposition – of how they have to be combined in order for the proposition to be true.

But this proposal doesn’t remove the problem. Bringing to consciousness Theaetetus, the property of Sitting and the relation of Monadic Instantiation doesn’t suffice for representing in mind Theaetetus as sitting, and hence for understanding the proposition that Theaetetus sits. The additional object of awareness doesn’t improve matters, as a subject who has brought to consciousness these three items might not grasp the way in which they have to be combined in order for Theaetetus to sit – in order for the proposition to be true: the pair consisting of Theaetetus and the property of Sitting has to instantiate the relation of Monadic Instantiation.
We could try to remedy the situation by introducing an additional object of awareness. What needs to happen to Theaetetus, the property of Sitting and the relation of Monadic Instantiation in order for Theaetetus to sit is that the particular and the property need to bear the relation of Dyadic Instantiation to the relation of Monadic Instantiation. This might suggest that the problem would be solved if we construed understanding of the proposition that Theaetetus sits as involving awareness of the three items already considered and the relation of Dyadic Instantiation. But this leaves us in the same situation that we faced before. We are off on a regress. Adding objects of awareness that embody modes of combination does not solve the problem. A different approach is needed.

Wittgenstein arrived in Cambridge in the autumn of 1911, with the intention of studying with Russell. He stayed in Cambridge until October 1913. In this period, Russell and Wittgenstein developed a famously intense personal and philosophical relationship. We know that Russell discussed with Wittgenstein the views that he was developing in May 1913 in the *Theory of Knowledge* manuscript. And Russell’s ideas had a profound influence on Wittgenstein’s conception of the problems. Wittgenstein encountered the semantic version of the problem of predication in the form that Russell had given to it in May 1913, i.e. as the mode-of-combination problem. I am going to argue that a central component of the position advanced in Wittgenstein’s *Tractatus* should be seen as Wittgenstein’s proposal for dealing with the problem.

In *Theory of Knowledge*, Russell sought to solve the mode-of-combination problem by introducing forms in his analysis of understanding. But Wittgenstein’s pre-Tractarian manuscripts show that, for Wittgenstein, Russell’s appeal to forms was fraught with difficulties. Wittgenstein thought that the items that Russell identified with forms (full existential generalisations) couldn’t play the role that Russell had assigned to them in his theory of understanding, and, more generally, no entity could play this role. As early as the ‘Notes on Logic’, dictated in October 1913, Wittgenstein’s verdict on Russell’s appeal to forms is unambiguously negative: ‘There is no thing which is the form of a proposition … This goes against Russell’s theory of judgment’ (Potter 2009, 282). Wittgenstein’s discussion of these ideas clearly indicates, first, that Wittgenstein was concerned with the problem that Russell had tried to solve with forms and, second, that Wittgenstein was convinced that forms could not be invoked in the solution to the problem.

What is needed is a way of construing the subject’s awareness of how, in our example, Theaetetus and the property of Sitting are represented as combined by the proposition that Theaetetus sits. As we’ve just seen, this cannot be achieved

---

8 Notice, however, that Russell’s proposal was not the regress-generating strategy that we have just considered. See Zalabardo (2015, §1.7) for details.
by thinking of the mode of combination as an object and including awareness of this object in our account of understanding. Another approach is required.

We get a suggestion of how this can be achieved if we go back to the unworkable account of judgment with which we started, according to which judgment is a relation between the mind and a fact. If the mind has the power to bring to consciousness the fact that Theaetetus sits and its structure, understanding of the proposition that Theaetetus sits can be explained in this way. By bringing to consciousness the fact that Theaetetus sits, the subject will grasp, not only the constituents of the fact, but also the way in which they are combined with one another in the fact. And this grasp doesn’t lead to an infinite regress. We are not invoking a third item, whose connection to the previous two would then have to be grasped. The subject doesn’t grasp a mode of combination in isolation. She grasps an actual combination and abstracts from this the way in which its constituents are combined.

This model of understanding offers a solution to the mode-of-combination problem, but, as we have seen, it cannot serve as our general account of understanding, since it’s only applicable to the understanding of true propositions. The proposition that Theaetetus flies cannot be understood in this way. This would require bringing to consciousness the fact that Theaetetus flies, but since the proposition is false there is no such fact. The item that would need to be brought to consciousness in order to understand the proposition in this way simply doesn’t exist.

Nevertheless, a modification of this approach will give rise to an account of understanding that is equally applicable to true and false propositions. Notice that the source of the problem with false propositions is that the fact grasp of which is supposed to produce understanding is the same as the fact that would make the proposition true. But this is not an essential component of the solution to the mode-of-combination problem. For the purpose of solving this problem, understanding can be taken to consist in grasp of any other fact, provided that its constituents are combined with one another in the way in which objects would have to be combined in order for the proposition to be true. Thus, for example, by grasping the fact that Socrates stutters and its structure, we grasp the way in which Theaetetus and the property of Flying would have to be combined in order for the proposition that Theaetetus flies to be true – they would have to be combined in the same way in which the constituents of the fact that Socrates stutters are actually combined with one another in that fact. The crucial consequence of this move is that the resulting account is applicable to the understanding of false propositions as well as true propositions. The fact that I grasp when I understand the proposition has to obtain, but the fact on whose obtaining the truth value of the proposition depends may or may not obtain.

Hence we have reached an account according to which we understand a proposition by grasping a fact whose constituents are combined with one
another in the same way in which objects in the world (the referents of the terms in the proposition) would have to be combined with one another in order for the proposition to be true. This account of understanding is, I submit, the central idea of Wittgenstein’s picture theory of representation. The pictures, thoughts and propositions of the Tractatus are the facts that we grasp, on this account, in episodes of understanding.9 The claim that propositions represent pictorially is Wittgenstein’s solution to the semantic version of the problem of predication.

Wittgenstein states very clearly that what he is calling pictures are facts (TLP 2.141), and that the advantage of using facts to represent the world as being a certain way is that the constituents of a fact are combined with one another in a certain way (TLP 2.14, 2.031), i.e., that the fact exemplifies a mode of combination. This circumstance enables us to use a fact to represent things as combined in a certain way:

2.15 The fact that the elements of a picture are related to one another in a determinate way represents that things are related to one another in the same way.

Hence, by using a fact to represent things as being a certain way, we bring a mode of combination into the representational episode – we represent things in the world as combined with one another in the same way in which the constituents of the depicting fact are combined with one another. And this is achieved without generating an infinite regress. We don’t answer the question of how we are representing objects in the world as combined with one another by identifying an additional object that plays the role of a mode of combination, whose combination with the objects it’s supposed to combine would then need to be grasped. We are answering the question by singling out an instance of the mode of combination in question: we represent objects in the world as combined with one another like that – as the constituents of the depicting fact are combined with one another.

For pictorial representation, then, we can provide a satisfactory explanation of how the way in which objects are represented as combined with one another is brought into representational episodes. Hence, if mental and linguistic representation followed the pictorial model, we would have at our disposal a solution to the mode-of-combination problem. And this is precisely what Wittgenstein goes on to claim. Thoughts are pictures of a certain kind (TLP 3), and propositions give expression to thoughts (TLP 3.1). Propositional representation is then directly characterised as following the pictorial model:

3.14 What constitutes a propositional sign is that in it its elements (the words) stand in a determinate relation to one another. A propositional sign is a fact.

9 The main ideas of this account of the relationship between Wittgenstein’s picture theory and Russell’s theory of judgment were first presented by David Pears. See Pears (1977).
Propositions, like all pictures, are facts whose constituents (the words) are combined with one another in a certain way, and they represent objects in the world (the referents of the words) as combined with one another in that same way. Commentators have often been puzzled by how the pictorial model could apply to linguistic representation. If we think of ‘Theaetetus criticises Socrates’ as a fact, whose constituents are combined with one another in a certain way, it is natural to think of it as the fact, say, that ‘Theaetetus’, ‘criticises’, and ‘Socrates’ are concatenated with each other in this order. But then it seems that ‘Theaetetus criticises Socrates’ cannot represent Theaetetus as criticising Socrates. All it can do is represent Theaetetus, the relation of Criticising and Socrates as concatenated with one another in that order. However, this characterisation of the situation doesn’t take account of two important points.

First, if a proposition is to represent the referents of its constituents as combined with one another in the same way as the constituents are combined in the proposition, then the constituents of the proposition have to have the same combinatorial possibilities as their respective referents. Hence in a proposition that represents Theaetetus as criticising Socrates, the representatives of Theaetetus and Socrates have to be particulars, e.g., the words ‘Theaetetus’ and ‘Socrates’, but the representative of the relation of Criticising has to be a binary relation – e.g., the relation $\rho$ that $x$ bears to $y$ when $x$ is to the left of $y$ with the word ‘criticises’ between them. Then the fact with which ‘Theaetetus criticises Socrates’ has to be identified is not the fact that ‘Theaetetus’, ‘criticises’, and ‘Socrates’ are concatenated with each other in this order, but rather the fact that ‘Theaetetus’ $\rho$ ‘Socrates’.

This might not seem to be much progress, since all we seem to be able to depict with this fact is Theaetetus standing to the left of Socrates with the word ‘criticises’ between them. But here we would be assuming that the constituents of ‘Theaetetus’ $\rho$ ‘Socrates’ are ‘Theaetetus’ and ‘Socrates’, while $\rho$ is the way in which they are combined. We avoid this pitfall if we take account of the second point: thoughts and propositions are pictures of a special kind – they are logical pictures. This means, in a nutshell, that the way in which its constituents are combined with one another is a logical mode of combination, with every non-logical aspect of the proposition treated as a constituent. On this construal, the constituents of ‘Theaetetus’ $\rho$ ‘Socrates’ are ‘Theaetetus’, ‘Socrates’ and $\rho$, put together by the logical relation of binary instantiation – i.e., ‘Theaetetus’ bearing $\rho$ to ‘Socrates’. Now, if ‘Theaetetus’ stands for Theaetetus, ‘Socrates’ for Socrates and $\rho$ for the relation of Criticising, ‘Theaetetus’ $\rho$ ‘Socrates’ will represent Theaetetus, Socrates and the relation of Criticising as combined with one another in the same way in which the constituents of the proposition are combined with one another. In this way, it will represent Theaetetus as criticising Socrates, as desired. I take this to be the
account of how propositions represent that Wittgenstein offers in the following section of the *Tractatus*:

3.1432 Instead of, ‘The complex sign “aRb” says that a stands to b in the relation R’ we ought to put, ‘That “a” stands to “b” in a certain relation says that aRb.’

I want to highlight one controversial aspect of this construal of Wittgenstein’s views. If Wittgenstein was trying to solve the mode-of-combination problem, his goal was to explain how to bring into representational episodes awareness of how objects in the world are represented as combined – of how they would have to be combined for things to be as they are represented as being. Read in this way, Wittgenstein’s position is that we bring the mode of combination to consciousness by grasping the fact that we use as a picture and its structure – the way in which its constituents are combined with one another. But what form is this grasp supposed to take? One thing seems clear: if this grasp could only take the form of pictorial representation (a picture of a picture), we would be off on a new regress. The proposal is only viable if it presupposes a more basic form of grasp of facts and their structure – one that isn’t pictorial in nature. I want to close my discussion of Wittgenstein by suggesting that this more basic type of grasp is provided in Wittgenstein’s framework by the notion of showing. A picture, he tells us at TLP 2.172, ‘displays’ (aufweisen) its pictorial form. Our access to logical form is described in similar terms at TLP 4.121: Logical form is ‘mirrored’ (spiegeln) in propositions, it ‘expresses itself’ (sich ausdrücken) in language. Propositions ‘show’ (zeigen), ‘display’ (aufweisen) the form of reality. And the same mode of access is invoked to explain our grasp of the pairings of the constituents of the picturing fact with the objects they stand for: ‘one proposition “fa” shows that the object a occurs in its sense’ (TLP 4.1211).

On the reading that I have sketched, I would grasp the fact that I use as a picture, its structure, and the pairings of its constituents with their referents, with the faculty that enables me to be a recipient of what is shown. This type of grasp is not explained, but presupposed in Wittgenstein’s proposal. It should be clear however that presupposing this kind of grasp doesn’t make picturing redundant. What we are presupposing is a faculty of grasping actually obtaining facts and their structure, whereas what we are explaining in terms of this faculty is grasp of combinations that may or may not obtain – the kind of grasp of Theaetetus as flying that enables us to understand the proposition that

---

10 For this construal of Wittgenstein’s notion of logical picturing, see Zalabardo (2015, §2.5). See also Peter Long’s illuminating exegesis of TLP 3.1432 (Long 1969).

11 And, in any case, Wittgenstein doesn’t think that the way in which the constituents of a proposition are combined with one another can be represented with propositions. See TLP 4.12.
Theaetetus flies, or judge that he does, even if, as a matter of fact, he’s never left the surface of the Earth.\textsuperscript{12}

\section*{4 Burge’s Fregean Solution}

Before we turn to Davidson’s own approach to the problem, I’d like to consider briefly a solution that Tyler Burge claims to find in Frege’s work. In a highly critical discussion of Davidson’s ideas on the problem of predication, Burge has claimed that Frege’s semantics offer a satisfactory solution of the semantic version of the problem. Burge disagrees with Davidson’s claim that “associating predicates with objects such as universals, properties, relations, or sets … will always lead to an infinite regress” (Davidson 2005, 143). Burge agrees that pairing predicates with universals cannot be the whole solution to the problem, but he thinks that pairing predicates with universals can be part of a satisfactory solution. What generates the regress, for Burge, is not the appeal to universals, but the assimilation of the relationship between predicates and universals to the relationship between singular terms and their referents: ‘The regress gets started if the syntactic and semantical roles of a predicate are assimilated to those of a singular term. It is absolutely essential to any solution that this assimilation be rejected’ (Burge 2007, 590). We avoid the regress if we recognise that a predicate doesn’t simply denote a property. In addition, the predicate of a subject-predicate sentence attributes the property to an individual:

The semantical relation between predicate and property is only an aspect of predication. Let us say that the predicate indicates (or predicatively denotes) the property or relation. Beyond indication, predication also constitutively involves a capacity for application or attribution of the property to the individual. Thus predication involves both indication and application of what is indicated to something further. (Burge 2007, 593)\textsuperscript{13}

Now, the Russellian construal of the problem of predication that I have been discussing here (the mode-of-combination problem) is significantly different from the construal of the problem that Burge considers, and it may well be that on Burge’s construal of the problem, his proposal offers a satisfactory solution. But the question I want to ask is whether Burge’s proposal offers a promising avenue for dealing with the mode-of-combination problem.

We can easily adapt Burge’s proposal to provide an account of Russell’s phenomenon of understanding. An episode of understanding of the proposition that Theaetetus sits would involve bringing to consciousness two items – Theaetetus and the property of Sitting – but each of these items would be brought

\textsuperscript{12} On this reading, the faculty of grasping what’s shown would be playing for Wittgenstein the role that the notion of acquaintance played for Russell. See Zalabardo (2015, §2.11) for details.

\textsuperscript{13} See also Wright (1998); MacBride (2006, 2011); Liebesman (2015).
to consciousness in a different way. Theaetetus would be brought to consciousness in a purely denotational mode, while the property of Sitting would be brought to consciousness in a mixed mode, which includes both denotation and, crucially, attribution of what is denoted to something else – to the item (Theaetetus) brought to consciousness in the episode of understanding in a purely denotational mode.

Now, if this proposal is going to offer a satisfactory solution to the mode-of-combination problem, the attributive character of the subject’s awareness of the property of Sitting would have to allow her to bring to consciousness the way in which the property and the individual are represented as combined – the way in which they would have to be combined in order for the proposition to be true. Attribution, in other words, would have to make the subject aware of the relevant mode of combination. Hence the proposal would have to be supplemented with an account of how attribution achieves this. One possibility would be to contend that the attributive aspect of predicative awareness ensures that when a property is brought to consciousness in this mode, the relation of Monadic Instantiation is simultaneously present to the mind. It should be clear, however, that this approach would take us back to the regress-generating proposal considered above. To provide a promising solution to the mode-of-combination problem, the Burgean proposal would have to be supplemented with an alternative account of how the attributive aspect of predicative denotation furnishes the subject with the requisite awareness. I don’t find in Burge’s discussion of the problem of predication any suggestion as to how this could be achieved.

5 Davidson’s Theory of Judgment

In the final chapter of *Truth and Predication*, Davidson puts forward a solution to the semantic version of the problem. Davidson’s solution is based on a lesson we can learn from Tarski’s work on truth:

for a language with anything like the expressive power of a natural language, the class of true sentences cannot be characterized without introducing a relation like satisfaction, which connects words (singular terms, predicates) with objects. (Davidson 2005, 30)

This indispensable role of the semantic properties of predicates and singular terms in the specification of the truth conditions of the sentences of a language is supposed to explain their contribution to the power of sentences to represent the world as being a certain way:

If we can show that our account of the role of predicates is part of an explanation of the fact that sentences containing a given predicate are true or false, then we have incorporated our account of predicates into an explanation of the most obvious sense in which sentences are unified, and so we can understand how, by using a sentence, we can make assertions and perform speech acts. (Davidson 2005, 155)
It is not at all obvious how Davidson’s solution is to be understood. The answer to this question can be expected to depend on the specific construal of the problem we focus on, and I think we can cast some light on Davidson’s proposal in relationship to Russell’s project of explicating the phenomenon of judgment and the version of the problem of predication that it faces.

As we’ve seen, the goal of Russell’s theories of judgment was to explain what has to be the case in order for a subject S to judge that p or, later, to understand the proposition that p. Specifically, his goal was to explain how S has to be related to the world in order for her judgment to have the content that p or for her to understand the proposition as having the content that p. From 1910, Russell thought that in order to achieve this the relevant mental episodes would have to be construed as involving a relation (acquaintance) to predicates and universals – to the items that would need to be combined into a unit in order for things to be as represented by the judgment or the proposition being understood. Davidson is unquestionably opposed to this strategy. What I want to consider is what alternative approach to Russell’s problem Davidson could be seen as advocating.

Throughout his discussion of the semantic version of the problem of predication, Davidson blames the difficulty on the appeal to universals as referents of predicates. This suggests that the aspect of Russell’s multiple-relation approach that Davidson would take exception to is the presence of universals among the worldly relata of episodes of judgment or understanding. If this is so, then a modification of Russell’s approach that avoided appealing to universals would meet with Davidson’s approval. This can be achieved with the help of Tarski’s notion of satisfaction. If we focus on subject-predicate sentences, obviating the need to treat infinite sequences of entities as satisfiers, satisfaction is a relation that particulars bear to predicates: a particular a satisfies a predicate P just in case P is true of a – i.e. just in case a sentence in which P is ascribed to a singular term referring to a would be true. Using this notion we can give the semantics of ‘Theaetetus sits’ without invoking a universal as the referent of ‘sits’. The worldly relata of the sentence would be the referent of ‘Theaetetus’ and those individuals that satisfy ‘sits’. The sentence would be true just in case the referent of the singular term is among the individuals that satisfy the predicate. On this account, there are no universals among the worldly relata of the sentence.

We can try to use these ideas to produce a version of Russell’s approach that doesn’t postulate universal relata for episodes of judgment or understanding. On this position, a predicative judgment would be connected to the world by two relations. On the one hand, it would be connected to an individual by the singular-reference relation. On the other hand, it would be connected by the satisfaction relation to a manifold of individuals – those that satisfy the predicate that would figure in a sentence expressing the judgment.
It is not clear that this move would produce a promising theory of judgment. Notice that the relations of singular reference and satisfaction would have to enable the subject to bring to consciousness their world-side relata. For example, when a subject judges that Theaetetus sits, she would have to have present to the mind Theaetetus and, instead of the property of sitting, the manifold of sitting individuals. But the new requirement is highly implausible. Surely judging that Theaetetus sits cannot require bringing to consciousness every sitting individual.

However, even leaving this difficulty to one side, it should be clear that Davidson would not endorse this approach. This would involve explaining the truth conditions of sentences in terms of the satisfaction conditions of predicates, but it is clear that, for Davidson, the order of explanation proceeds in the opposite direction:

Which of the two semantic concepts, satisfaction or truth, we take as basic is, from a formal point of view, open to choice. Truth, as Tarski showed, is easily defined on the basis of satisfaction; but, alternatively, satisfaction can be taken to be whatever relation yields a correct account of truth … The second is, I think, the right view. (Davidson 2005, 34–35)

If, as this passage and others indicate, the satisfaction relation is defined as whatever individual-predicate relation produces the right truth conditions for sentences, which truth conditions a sentence has cannot be defined in terms of the satisfaction conditions of the predicates that figure in it, on pain of circularity.

It should be equally clear that Davidson would not support either a return to Russell’s dual-relation approach. He certainly doesn’t see his endorsement of Tarski’s theory of truth as carrying a commitment to this approach:

Tarski provides no entities at all to which sentences correspond or which sentences name, picture, or otherwise represent. No facts appear in the official apparatus, nor do propositions, either as the meanings of sentences or as half-extensional, half-intensional entities in the world as in Russell’s early theory. (Davidson 2005, 155)

Davidson’s solution doesn’t postulate entities that could play the role of worldly relata in a dual-relation theory.

It seems, then, that Davidson would oppose all the strategies with which Russell attempted to explain how a subject has to be related to the world in order for her judgment to have the content that p or for her to understand a proposition as having the content that p. And I want to suggest that he is not offering an alternative strategy for explaining these subject-world relations. Davidson doesn’t think that such an explanation is to be had. Notice that if we had an account of the semantic relations between sentences or judgments and

---

bits of the world, we should be able to derive from it a pairing of true sentences or judgments with facts. However, in Truth and Predication Davidson argues forcefully against the idea of a correspondence of this kind: ‘there is nothing interesting or instructive to which true sentences correspond’ (Davidson 2005, 39), since ‘if true sentences correspond to anything, they all correspond to the same thing’ (Davidson 2005, 40). So, for Davidson, it’s not possible to provide an informative pairing of true sentences or judgments with the facts that make them true. But if this is not possible, then, a fortiori, it’s not possible to have an informative account of the relations between sentences or judgments (true or false) and the world from which the former obtain their content. We have to conclude that, for Davidson, it’s not possible to explain these relations between mind or language and the world. His ‘solution’ to the problem of predication is not a strategy for overcoming an obstacle to the provision of this explanation.

It follows that, according to Davidson, linguistic meaning and mental content cannot be explained in terms of relations between linguistic items or mental episodes and the world. How then is he proposing to explain these phenomena? I want to suggest that, in the case of linguistic meaning on which Davidson concentrates, his proposal is to explicate the phenomenon in terms of linguistic use.15

He expresses the point clearly in the following explanation of what is missing in the Tarskian truth theory with which he proposes to specify the semantics of a language:

What is missing is the connection with the users of language. Nothing would count as a sentence, and the concept of truth would have no application, if there were no creatures who used sentences by uttering or inscribing tokens of them. Any complete account of the concept of truth must relate it to actual linguistic intercourse. (Davidson 2005, 36)

Elsewhere he states explicitly that his views on the connection between truth and meaning are not in conflict with a use-based semantics:

What is clear is that someone who knows under what conditions a sentence would be true understands that sentence, and if the sentence has a truth value (true, false or perhaps neither) then someone who does not know under what conditions it would be true doesn’t understand it. This simple claim doesn’t rule out an account of meaning which holds that sentences mean what they do because of how they are used; it may be that they are used as they are because of their truth conditions, and they have the truth conditions they do because of how they are used. (Davidson 2000, 70)

However, the idea that meaning should be explained in terms of use is often seen as enjoining an epistemic view of truth, according to which we can specify in terms of beliefs or other cognitive attitudes what has to be the case in order for a sentence to be true. And Davidson’s rejection of this view is unambiguous.

15 For an application of these ideas to mental content see Davidson (2001b).
Here is his verdict on the version of the view according to which truth is warranted assertibility:

I respect this idea for the same reason I respect closely related pragmatic theories, because it relates truth to human attitudes like belief, intention and desire, and I believe any complete account of truth must do this. But theirs cannot be the right way to express the relation. For either the conditions of warranted assertibility are made so strong that they include truth itself, in which case the account is circular, or circularity is avoided by making the conditions explicit, and it then becomes clear that a fully warranted assertion may be false. (Davidson 2000, 67–68)

But his rejection of epistemic accounts of truth doesn’t lead Davidson to give up on the project of explicating truth and meaning in terms of use. His hope is that there is a way of achieving this that doesn’t take the form of an epistemic view of truth:

there must be some connection [between belief and truth] if we are to relate the truth of utterances to their use. The question is what that connection can be. (Davidson 2005, 42)

Davidson is certainly not alone in thinking that we can explain truth in terms of use without identifying truth with warranted assertibility. The view has been eloquently defended by Huw Price:

pragmatists have often ignored the resources of their own theoretical standpoint—even, in a sense, their own principles—in seeking to equate truth with something like warranted assertibility. A better alternative, in my view, is to seek to explain in pragmatic terms why our notion of truth does not line up neatly with warranted assertibility—in other words, to explain what practical use we have for a stronger notion. (Price 2011a, 16)

Advocates of epistemic accounts of truth seek to replace the analysis of truth as correspondence with an alternative analysis in terms of use. According to Price, their mistake is to think that an analysis is needed. Rather than trying to provide an analysis of the notion, the pragmatist should simply “investigate its function and genealogy” (Price 2011b, 167). Robert Brandom characterises the pragmatist methodology in similar terms, as “beginning with a story about the practice or activity of applying concepts, and elaborating on that basis an understanding of conceptual content” (Brandom 2000, 4).

I want to end by suggesting that Davidson’s position fits nicely in this methodological mould. The account of the practice of interpretation offered in Chapter 3 of *Truth and Predication*, and the related account he offers

---

16 In a note on the manuscript that the editors have included as a footnote at the end of Chapter 3, Davidson writes: “I want to make clear that my ‘solution’ isn’t a basic one. It is an alternative to deflationary, epistemological correspondence theories not in proposing a better definition (or short summary) but in suggesting a different approach which relates the concept of truth to other concepts” (Davidson 2005, 75).
elsewhere of the emergence of thought and linguistic meaning, is precisely what is required for an account of semantic phenomena along the lines of the brand of pragmatism advocated by Price and Brandom, among others. If Davidson was indeed pursuing this pragmatist methodology, his explanation of semantic phenomena would not require specifying relations between linguistic or mental items and the world that endow the former with their semantic properties. Since the problem of predication arises within this now redundant enterprise, Davidson’s pragmatism would have a good claim to count as his solution to the problem.

References


19 Richard Rorty has long advocated a reading of Davidson’s ideas along pragmatist lines. See, e.g., Rorty (1986). In Truth and Predication, Davidson expresses a partial acceptance of this reading: ‘Rorty has compared my views on the nature of truth with Dewey’s. I find much of what he has to say on this topic congenial, and I think he is right in a general way. I share Dewey’s attitude toward truth. …as I read him, Dewey thought that once truth was brought down to earth there were philosophically important and instructive things to say about its connections with human attitudes, connections partly constitutive of the concept of truth. This is also my view …’ (Davidson 2005, 9–10). Concerning the role that he assigns to Tarski’s work, he tells us that Rorty ‘sees clearly that for me this is related to the rejection of a representational picture of language and the idea that truth consists in the accurate mirroring of facts’ (Davidson 2005, 10).


Index

action
  agential knowledge of, 10, 25–6
  causal theory of, 2, 8–10, 13, 15–18, 23, 25–6
  continent vs incontinent, 14–17
  intentional vs unintentional, 9–16, 24–5
akrasia. See weakness of the will
Allen, Colin, 179
animal minds, 160, 162–3, 167, 173
  holodoxastic vs holophrastic beliefs, 169–70, 181
Anscombe, Elizabeth, 2, 8–10, 17–18, 25–6, 164
Arrington, Robert L., 215
assimilationism, 159, 162, See also
differentialism
Augustinian conception of meaning, 54, 100, 139, 145, 148–9, 151–3
Austin, J.L., 15, 46, 168

Ayer, A.J., 162

Bacon, John, 234
Baker, Gordon P., 151, 175
behaviourism, 125, 138, 159, 176
Bekoff, Marc, 179
Blackburn, Simon, 99, 118
Boghossian, Paul, 76, 99
Brandom, Robert, 159, 206, 208, 247–8
Bratman, Michael, 15
Brewer, Bill, 208
Bridges, Jason, 148, 156
Burge, Tyler, 227, 231, 242–3
Byrne, Alex, 76
Byrne, Richard W., 166, 172, 176

Call, Josep, 176
Candlish, Stewart, 235
cause
  initiating vs monitoring, 11–14, 25–6
  typical. See triangulation
Cavell, Stanley, 142
Chater, Nick, 180
Child, William, 189, 191
cohortentism, 208
combinatorial account of facts, 240
compositionality, 3, 28–9, 32, 36–7, 39–42, 44–5, 87
conditions of correct application. See
correctness conditions
constitutive ideal of rationality, 221
convention, 3, 46–7, 49–58, 60, 62–3, 65–6, 99, 130, 135
correctness conditions, 3, 52, 56–63, 72, 85, 89, 99–108, 112–18, 125, 133, 135, 146, 191–2, See also meaning,
determination of
correct applications, infinity of, 4, 71–3
correct applications, objectivity of, 4, 72–3, 83–9
decision, 135, 196, 198–9, 203
deflationism, 3, 28–9, 36, 38–44
DeGrazia, David, 162, 173–4, 179
detectable similarity, 87–93
deviant causal chains, 2, 8, 10–14, 17, 23
differentialism, 159, 162–4, 167, See also
asassimilationism
disjunctivism, 206–13, 218, 223
dispositional properties, 189–90, 192
Dodd, Julian, 213
Dretske, Fred, 152
Dummett, Michael, 129–30, 134–5, 174
Dupré, John, 176, 179

essence of language, 5, 100, 139–41, 148–54
Evans, Gareth, 210, 218
experience, 32, 188, 201–02, 205–13, 218, 223

first-person authority, 6, 161, 167
Fodor, Jerry, 177
Fogelin, Robert, 101
Index

Myth of Jones, 206, 208–10, 212, 223
Myth of the Given, 208

natural kinds, 66, 144
natural properties, 70, 89–92
neo-Fregeanism, 210–11, 219
non-reductionism, 4–5, 70, 75, 93, 97–8, 102, 106, 110, 114, 120, 131, 155, See also semantic naturalism
normative beliefs, 9–10, 16–20, 22, 24
normative judgments, 8, 10, 14, 16–17, 24
normative properties, 22–4, 214, 216, 219, 222

Olson, Kenneth, 233
ostension, 4, 63–4, 101, 109, 113, 115, 120, 124–8, 131, 133, 191, 215
ostensive definition. See ostension
ostensive teaching. See ostension
other minds, knowledge of, 187, 206
Pagin, Peter, 70, 72–3, 75–6, 83, 86
Pears, David, 107, 239
pictures, 239–40
Potter, Michael, 237
Price, H.H., 174
Price, Huw, 247–8
principle of charity, 3–4, 29–30, 32, 36, 42–3, 45, 48, 61–4, 70, 79–88, 92, 156, 221
pro-attitudes
holistic views of, 2, 9–10, 12, 17–21, 23, 25–6
standard views of, 9, 11–12, 14–23
systemic aim of, 9, 16, 20–5
quietism, 4, 6, 75, 97–8, 106–7, 114, 120
Quine, Willard Van Orman, 33, 37, 55, 69, 82, 88, 123, 180, 220
role of the radical interpreter, 3, 48, 61, 78, 80–9, 115–18, 220–1
reasons
motivating, 8, 16–17
normative, 8, 18
primary, 8, 11–15, 17
reference, 5, 28, 43, 90, 148–52, 154, 244
Rodriguez, Ángel García, 164
Rollin, Bernard, 162
Rorty, Richard, 248
paradox. See sceptical paradox
Rundle, Bede, 170
Russell’s theory of descriptions, 210
dissolution of, 1, 4, 6, 98, 108
solution of, 69, 76, 98, 108
sceptical problem, 97–8, 107–8, 110–20, See also sceptical paradox
Schwarz, Wolfgang, 90
secondary qualities, 23
Sellars, Wilfrid, 167, 206, 208, 210, 223
semantic context, 97, 106, 120
semantic externalism, 88, 113
interpersonal, 22
perceptual, 110, 113
social, 22, 113, 131
semantic internalism, 113–14
semantic naturalism, 93, 144, 155, See also non-reductionism
Singer, Peter, 162
singular thought, 210, 218–19
Skyrms, Brian, 233
Smith, Barry, 191
Smith, Michael, 9
Sorabji, Richard, 173
Stalnaker, Robert, 47
Stich, Stephen, 177, 180
Stoecker, Ralf, 14
Strawson, P. F., 46–7, 49–50, 65–6, 85, 187, 190, 193, 226
Stroud, Barry, 107, 136, 143
Stroud, Sarah, 15
surveyable representation, 150, 194
Tarskian truth-theory, 33, 41, 43, 48, 77, 81, 117, 155, 243–6, 248, See also T-sentence
Thöle, Bernhard, 191
Thornton, Tim, 107
Tissaw, Michael, 173
Tomassello, Michael, 171, 176
training, 61, 64, 110, 120, 124–7, 131, 176
transcendental empiricism, 206, 208, 223
translation, 28, 33, 38, 45, 69
Travis, Charles, 211
triangulation, 2, 4, 9, 18, 21–3, 76, 111–20, 126–7, 156, 170–1, 190
aspect problem, 22, 84, 101–02, 106, 111–13, 117–18
Index

distance problem, 111–12, 116

typical cause, 111, 113–14

T-sentence, 79, 117, 127–8

Velleman, David, 10, 16
Verheggen, Claudine, 10, 17, 20, 22, 24, 61, 63, 70–2, 76, 93, 100, 102, 106, 109, 112, 114–15, 118, 120, 171

Von Savigny, Eike, 165–6

weakness of the will, 2, 14, 17, 23

Weatherston, Brian, 90

Whiting, Daniel, 100

Wikforss, Åsa, 53, 55, 57–8, 60, 62, 70, 72–3, 99–100

Williams, J. Robert G., 90

Williams, Meredith, 46–7, 56–7, 60–2, 64–5, 101–02, 110

Williams, Michael J., 43

Williamson, Timothy, 76

Wilson, George, 101, 106–7

Wright, Crispin, 69, 76, 93, 99, 128, 242

Zalabardo, José L., 233, 237, 241–2, 248