

SEMANTIC EXTERNALISM

Jesper Kallestrup

 **Routledge**
Taylor & Francis Group
LONDON AND NEW YORK

CONTENTS

Acknowledgements	ix
Introduction	1
Chapter 1 Descriptivism	10
1.1 Descriptivism expounded	10
1.2 The identity argument	15
1.3 Puzzles about belief	21
1.4 Sense, linguistic meaning and communication	26
Chapter summary	32
Annotated further reading	34
Chapter 2 Referentialism	35
2.1 Rigidity and direct reference	35
2.2 Kripke's modal argument	40
2.3 Rigidification	45
2.4 Referentialist belief ascriptions	51
Chapter summary	56
Annotated further reading	57
Chapter 3 From language to thought	58
3.1 Putnam's Twin Earth argument	58
3.2 Internalist rejoinders to the Twin Earth argument	64
3.3 Burge's arthritis argument	69
3.4 Internalist rejoinders to the arthritis argument	74
3.5 Davidson's Swampman argument	79
3.6 Externalist rejoinders to the Swampman argument	83

Chapter summary	89
Annotated further reading	91
Chapter 4 Varieties of narrow and wide content	93
4.1 Object-dependent thoughts	93
4.2 Indexicality and egocentric thoughts	100
4.3 Two-factor theories of content	105
4.4 Natural kind concepts revisited	112
4.5 The metaphysics of content properties	117
Chapter summary	122
Annotated further reading	124
Chapter 5 Self-knowledge	125
5.1 Introducing self-knowledge	125
5.2 Entitlement to self-knowledge	132
5.3 Incompatibilism	139
5.4 Slow switching	143
5.5 Reasoning	149
Chapter summary	154
Annotated further reading	156
Chapter 6 Scepticism	157
6.1 Scepticism about self-knowledge	157
6.2 External world scepticism	162
6.3 Putnam's proof	166
6.4 McKinsey's recipe	173
Chapter summary	183
Annotated further reading	184
Chapter 7 Mental causation	185
7.1 The varieties of mental causation	185
7.2 The modal argument for narrow content	190
7.3 The doppelgänger challenge	197
7.4 The explanatory value of wide content	204
Chapter summary	210
Annotated further reading	211
Glossary	213
Notes	224
Bibliography	249
Index	265

INTRODUCTION

To see that Jane's jumper is red and to believe that apples are wholesome are both mental states. These two states are also representational. To say that a mental state is representational is to say that it serves the function of being about something in the world, or that it takes the world to be a certain way. Perceptions and beliefs are representational states. To be in such states is to represent the world as being a certain way. The way a state represents the world as being is its representational content. In particular, being in a belief state involves being in a state that can be true or false, depending on whether the world is as the belief represents it as being. Some mental states also have phenomenal characters. To say that a perceptual experience has a phenomenal character is to say that there is something it is like for the subject who undergoes that experience. Think of the qualitative feel associated with a visual experience of a red jumper. While some philosophers take phenomenal character to be fully determined by the representational content of the perceptual experience in question, others maintain that phenomenal character cannot be exhaustively understood in terms of the representational content of that experience. That is not a dispute we shall probe into here. In fact, we shall henceforth focus on beliefs and other so-called propositional attitudes. To say that Anna believes that apples are wholesome is to say that Anna bears the attitude of belief towards the proposition that apples are wholesome. Propositions are abstract entities to

which one can be belief related. They are composed of concepts and capable of being true or false. The kind of representational content that shall occupy us is propositional content. Unless explicitly stated otherwise, we shall use ‘meaning’ in the sense of propositional content. Some philosophers maintain that some representational content is non-conceptual. They believe that we can represent the world in ways that outstrip our conceptual resources. Anna’s visual experience of shade red_{29} is phenomenally different from that of red_{31} but she has no such colour concepts. Hence, the representational content of her visual experience is more fine-grained than its propositional content. Others contend that we can fully specify the way Anna represents the world just by using concepts that she possesses, e.g. the propositional content of her experience is that that shade (demonstrating red_{29}) is thus and so. Again that is not a controversy we shall delve into in the following.

The nature of meaning, understood as propositional content, is a multifaceted, vexed issue. This book introduces and assesses what philosophers call the problem of semantic externalism. Given that most of the literature on the subject concerns the debate between semantic externalists and semantic internalists, much of the book will be taken up with examining the various arguments and positions at stake in that debate. These two camps disagree on how meaning is individuated. While those in the latter camp say that meaning is fully determined by features that are internal to the speaker, those in the former say that meaning is determined at least in part by features that are external to the speaker. Here is an example. Suppose the man fumbling for his car keys in the bush in front of you is John. While pointing at the bush you utter the sentence ‘that man is drunk’. What you have said is that John is drunk. Now suppose instead that John’s identical twin James is in the bush. As far as you know everything else is the same, indeed apart from the fact that the seemingly identical individuals John and James are in fact distinct, everything else is the same. While pointing at the bush you utter the same sentence. What you have said is now that James is drunk. You are internally the same in the two situations. For instance, the visual experience you have in the first situation is indistinguishable from the one you have in the second situation. The proposition you expressed by those two utterances cannot therefore be a function of what is in your head, but must rather depend on what the external world is like. The upshot is that meaning is determined at least in part by external factors possibly beyond speakers’ ken. Or so semantic externalists contend. That is not terribly controversial in the case of demonstratives such as ‘that’

or indexical expressions such as 'I'. The mere fact that John and James are distinct individuals entails that what John expresses when he utters 'I feel elated' is different from what James expresses when he utters 'I feel elated'. Still, there also seems to be an important sense in which John and James spoke alike. After all, each of them feels elated. Similarly, there is some significant sense in which I said the same thing in the two situations, i.e. that the demonstratively identified individual is drunk. Semantic internalists intend to pinpoint the claim that meaning, or at least an aspect thereof, depends solely on what speakers are like internally, and so is insensitive to the relevant kinds of variation in the external environment. By contrast, semantic externalists are keen to stress that the meaning of natural kind terms, e.g. 'tiger' or 'water', and even social kind terms, e.g. 'carburettor' or 'sofa', are also externally individuated. The externalist strategy is, again, to let two internally identical speakers inhabit two environments identical apart from some imperceptible microphysical or sociolinguistic difference. Unbeknownst to them, their usage of these terms will thus pick out different natural or social kinds, and so what they express will be externally fixed in different ways, possibly behind their backs. But of course if what these speakers believe is determined by the propositions expressed by the sentences they use to express their beliefs, then not only is linguistic meaning externally individuated, so also are the contents of their beliefs. Semantic externalists conclude that belief content is determined externally, and therefore also the belief states themselves, provided that such states are individuated by their contents. Semantic internalists typically respond by pointing out that since the relevant environments are identical with respect to all manifest or superficial features, the best sense is made of these speakers' linguistic and physical behaviour if they are instead attributed beliefs with some common content. The contents of speakers' beliefs represent how they take the world to be, and internally identical speakers who are embedded in environments which for all they know are identical, take the world to be the same way. Importantly, semantic internalists submit that such speakers behave alike for the purposes of psychological explanation because their conceptions of how things are, are the same.

Semantic externalism gained prominence in the 1970s and 1980s, in part as a result of Saul Kripke and Hilary Putnam's revolutionary attack on the descriptive theory of reference, or simply descriptivism, in the philosophy of language. Going back to Gottlob Frege, this view says that there are sets of descriptive properties associated by competent speakers with singular and general terms which both give those terms their meaning and

determine their reference. Instead these philosophers advocated the direct reference theory, or simply referentialism, according to which such terms pick out their referents directly, i.e. unmediated by any descriptive properties. On this view, the meaning of such terms is exhausted by their referents, but what determines their reference is the baptismal way in which they were initially introduced into the language and then passed on through a causal-historical chain. While descriptivism and semantic internalism, and referentialism and semantic externalism, are distinct doctrines, they nevertheless in some versions often go hand in hand. For instance, referentialism entails a version of semantic externalism. If the meaning of a referring term is given by the external object to which it refers, then such meaning is obviously externally individuated. But the converse is false. Some semantic externalists hold that the meaning of a referring term consists in a possibly descriptive way of thinking of its referent which would not exist had that term lacked a referent. To properly appreciate the case for semantic externalism as a view in philosophy of mind or thought, we need to examine the objections levelled against descriptivism as a view in philosophy of language. Similarly, in order to deepen understanding of the recent counter-revolution, led by David Lewis, Frank Jackson, David Chalmers and others, aiming to revive some brand of semantic internalism, we need to determine whether these philosophers are right that the referentialist objections can be met in a satisfactory way. The first two chapters are devoted to these issues.

Chapter 1 begins by presenting descriptivism as the view that the meaning of singular and general terms consists in descriptive content which is both what determines their reference, and is what competent speakers know when they understand them. For instance, the meaning of 'Aristotle' could be given by the definite description 'the teacher of Alexander the Great', and the meaning of 'water' could be given by 'the clear, potable liquid that fills the oceans, falls from the sky, and is called "water"'. This view has been defended by John Searle, Michael Dummett, Frank Jackson, David Lewis and others, but can be traced back to Frege and Bertrand Russell. Then Frege's case for the existence of what he called 'sense' as distinct from reference is laid out. Here sense is taken to be a mode of presentation, or a way of thinking, of the referent. This famous identity argument in its strengthened version trades on intuitions about the behaviour of referring terms in propositional attitude contexts. A propositional attitude is roughly an individual bearing an attitude towards a proposition, e.g. John believes that Mary is off work. The argument shows that sense

must be what is cognitively significant to speakers: if an individual can take different cognitive attitudes towards two propositions, then they differ in sense. Finally, deploying a well-known example of Frege's, the question of whether sense might constitute a notion of linguistic meaning shared across a speech community is discussed. In this context, the role that knowledge of meaning plays in communication is examined. The pressing worry is that if sense varies hugely from speaker to speaker, then intersubjective transmission of knowledge in communication is jeopardized.

Chapter 2 starts out by introducing direct reference and rigidity, as these notions are key to understanding referentialism. Thus rigidity is the idea that a term picks out the same object at all possible worlds. A possible world is a way our world might have been. On this view, referring terms are rigid, because they are directly referential. Sentences containing referring terms thus express singular propositions, which have as their constituents the referents of those terms. Kripke's modal argument against descriptivism is then presented. This celebrated argument turns on intuitions about the behaviour of referring terms in modal contexts. The descriptivist's response is either to invoke two notions of content, or avail herself of rigidified definite descriptions. Gareth Evans defended a hybrid view according to which one aspect of content determines reference at the actual world, while another aspect of content determines reference at possible worlds. And Lewis and Jackson propose that a proper name such as 'Aristotle' be short for something like 'the actual teacher of Alexander the Great', so that 'Aristotle' refers to Aristotle even at possible worlds where someone other than Aristotle taught Alexander the Great. However, neither of these rejoinders is entirely unproblematic. Scott Soames has pointed out that speakers in a possible world perfectly resembling the actual world need have no beliefs about the actual world in order to have beliefs about Aristotle. Finally, a response due to Nathan Salmon and Scott Soames on behalf of referentialism to the belief argument from Chapter 1 is detailed. This says that singular propositional content is grasped under semantically insignificant modes of presentation.

Chapter 3 covers the transition from philosophy of language to philosophy of thought and mind. First Putnam's Twin Earth argument is presented in some detail followed by critical discussion. This argument purports to show that the meaning of natural kind terms is partially determined by underlying physical facts about the natural kinds that are picked out by those terms even if those facts are unknown to the speakers in question. The reasoning seems compelling, but some philosophers, e.g. Tim Crane, Gabriel Segal

and Frank Jackson, have put forward challenging lines of discontent. Then Tyler Burge's three-way extension of Putnam's conclusion is spelled out. Firstly, if linguistic content is externally individuated, then so is mental content. Indeed if mental states themselves are partially individuated by their contents, being in a belief state with such content is also externally individuated. Secondly, mental content is not only dependent on physical facts about the external environment, it also depends on social facts about the way language is used conventionally in the speech community. Thirdly, not only is the content of natural kind terms externally individuated, so is the content of non-natural kind terms. Burge's arthritis argument is then presented, followed by critical discussion of various objections that have been levelled against this social externalist view. Here the phenomenon of semantic deference plays a pivotal role – the idea that speakers who incompletely understand an expression can nevertheless competently use that expression and have beliefs about its referent by deferring to those expert speakers who possess the relevant knowledge. Finally, Donald Davidson's Swampman example is scrutinized. This thought experiment poses a challenge not only to Davidson's own historical account of representational content but also to teleosemantics according to which the selectional histories of individuals individuate the content of their representational states. Several responses are discussed, including one which highlights a deep tension in Davidson's work on meaning and radical interpretation.

Chapter 4 begins by scrutinizing Dry Earth cases where the relevant external facts go missing. They purport to show that if content is individuated by external circumstances, then such content is also dependent for its existence on the obtaining of those circumstances. Examples of object-dependent content are given by David Kaplan's account of perceptual-demonstrative thoughts, and by Gareth Evans and John McDowell's idea of Fregean *de re* senses. Then the proposal that natural kind terms are short for rigidified definite descriptions is revisited. Given that rigidification involves an actuality operator, and that 'actually' is an indexical expression, thoughts involving the concept of water become egocentric in nature. The question is then whether the content of egocentric thoughts should be viewed as truth-conditional, or else, as David Lewis has it, consists in the self-ascription of certain properties. A natural proposal is to acknowledge the existence of two distinct components of mental content: narrow content, as internally individuated, and wide content, as externally individuated. Two such hybrid views, due to Colin McGinn and David Chalmers, are discussed in some detail. While they disagree about the semantic import of

narrow content, they both assign such content the role of causally explaining behaviour. At this juncture, the semantics of natural kind concepts is revisited. In particular, Putnam and Chalmers' claim that such concepts have an indexical component is critically examined. Instead a referentialist semantics is explored. Finally, care is needed when characterizing narrow content. To be in a state with narrow content is to be in a state that supervenes on intrinsic properties, but narrow content is not intrinsic. Drawing on work by David Lewis, Frank Jackson and Robert Stalnaker, it is argued that narrow content is best seen as intra-world narrow, content that is shared by internal duplicates only within the same possible world, not across different possible worlds.

Chapter 5 is mostly devoted to the dispute between compatibilism and incompatibilism. Compatibilists claim that semantic externalism and self-knowledge are compatible, while incompatibilists deny that claim. Competent speakers are standardly credited with privileged access to the contents of their own occurrent mental states, which gives rise to a priori knowledge of these contents. But this would seem to be impossible if those contents depend for their individuation on external circumstances that such speakers have no special access to, and indeed need know nothing about. There are different ways of setting up this incompatibilist problem. Paul Boghossian argued that if a speaker is slowly but unwittingly switched back and forth between Earth and Twin Earth, she will fail to know a priori the wide contents of her occurrent thoughts, as she is unable to rule out the possibility that she is in fact having a thought with a relevantly different content. Relying on Burge's notion of self-verifying thoughts, some semantic externalists reply that a speaker can know a priori what she is thinking without having a priori knowledge whether that content is identical to, or different from, some other content that she is thinking. But Burge did not take self-verification to be what bestows warrant on second-order judgements about thoughts. Instead, speakers are a priori entitled to such judgements in virtue of the role these judgements play in critical thinking. Slow switching gives rise to another kind of problem for semantic externalism. If a competent speaker is unwittingly transported back and forth between Earth and Twin Earth, she will unawares think different wide contents when she utters sentences containing 'water'. Consequently, if she were to go through an intuitively valid argument while undergoing such transportations, occurrences of 'water' in the premises would equivocate between these distinct contents. Her reasoning would then be rendered invalid.

Chapter 6 continues to examine the epistemological implications of semantic externalism, and in particular whether this view swaps a problem about knowledge of the external world with one about knowledge of the internal world. Anthony Brueckner confronts the semantic externalist with a line of reasoning to the effect that a competent speaker's introspective belief that she is thinking that water is wet cannot constitute a priori knowledge. If this speaker knows a priori that she is thinking that water is wet, then she also knows a priori that she is not thinking that twin-water is wet. Since she does not know a priori that she is not thinking that twin-water is wet, she does not know a priori that she is thinking that water is wet. In response, some question the underlying epistemological principles, and others maintain that the content sceptical argument is self-refuting. Brueckner contends that while semantic externalism provides the basis for scepticism about self-knowledge, it offers no resources when it comes to scepticism about the external world. Putnam argued that one successfully thinks one is not a brain in a vat (BIV) only if one is not a BIV. So, if one knows a priori that one thinks that one is not a BIV, one knows a priori that one is not a BIV. Relatedly, Michael McKinsey and Paul Boghossian argue that combining a strong version of semantic externalism with self-knowledge leads to implausible a priori knowledge of ordinary contingent external world propositions. Respondents include Jessica Brown, Bill Brewer, Jim Pryor, Brian McLaughlin and Michael Tye. This last argument might be regarded as proof that anti-sceptical arguments supported by semantic externalism over-reaches by way of delivering easy knowledge of such common or garden-variety propositions.

Chapter 7 begins by emphasizing the importance of mental states causing physical states or other mental states. In particular, mental states cause their effects in virtue of their contents. Basically, what causes John to go into the pub is his desire for a beer and his belief that he can satisfy that desire by going into the pub. Had his belief or desire had a different content, John would have acted differently. But if mental content is externally individuated and causation is local, mental states with such wide content seem causally idle. John's physiological and neurological properties are causally efficacious with respect to his physical behaviour, but the features of his external environment that individuate the content of his mental states are not. This suggests that only the narrow contents of John's mental states are causally efficacious. Different ways of sharpening this challenge to semantic externalism have been proposed, most prominently by Jerry Fodor and Harold Noonan. Their arguments are presented in some detail, and the question of whether they sustain any viable notion of narrow content is

assessed. In response, one might distinguish between causation and causal explanation, or between different ways of describing the effect. Thus Timothy Williamson argues that knowledge states play an irreducible role in causal explanation of behaviour, Frank Jackson and Philip Pettit argue that causal explanations can cite features that program without actually producing anything, and Fred Dretske's dual-explanandum strategy has it that the triggering physical properties of mental states are responsible for mere bodily movement, while the structuring content properties of such states are responsible for behaviour.

At the end of each chapter is an annotated list of further reading, which offers details on relevant further literature pertaining to that chapter. Additionally, each chapter is appended with a chapter summary, which allows for a swift recap of the key points in that chapter. Difficult philosophical terminology or relevant technical terms are briefly explained in the glossary of philosophical terms at the end of the book.

Semantic externalism is a vexed issue in contemporary philosophy involving a vast literature. To bring the exposition into focus, the following chapters ignore many otherwise important approaches to this intriguing debate. For instance, Fred Dretske, Michael Tye and others have argued that the contents of experiential states are equally externally individuated. As mentioned, an experiential state has phenomenal character in that there is something it is like to be in that state. We set such phenomenal externalism aside, and focus entirely on linguistic contents and the contents of propositional attitudes. Another increasingly popular branch of externalism is so-called active externalism, which maintains that the external environment plays an active role in constituting cognitive processes. Recent advocates include Andy Clark and David Chalmers. Whereas semantic externalism says that some mental contents are externally individuated, active externalism holds that the vehicles of these contents are externally located. On the face of it, semantic externalism is compatible with active internalism, and semantic internalism is compatible with active externalism. To properly assess the merits of active externalism and these intriguing compatibility claims is beyond the scope of this monograph. In order to ease exposition the chapters are also deliberately selective when it comes to particular arguments and objections. Rather than covering too much ground superficially, key territory is expounded as thoroughly as possible. While it is fair to say that a majority of professional philosophers nowadays incline towards some version of semantic externalism, the aim is throughout for the exposition to be as even-handed as possible, or at least to ensure that both sides to a given dispute are represented in the text.