

ON THE STRUCTURE OF COMMUNICATIVE UNDERSTANDING

by

Endre Begby

Cand. mag., Universitetet i Oslo, 1998

Cand. philol., Universitetet i Oslo, 2000

Submitted to the Graduate Faculty of
Arts & Sciences in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy

University of Pittsburgh

2008

UNIVERSITY OF PITTSBURGH

Arts & Sciences

This dissertation was presented

by

Endre Begby

It was defended on

05/01/2008

and approved by

Robert B. Brandom, Distinguished Service Professor, Department of Philosophy

J.E. McGuire, Professor, Department of History and Philosophy of Science

Thomas Ricketts, Professor, Department of Philosophy

Kieran Setiya, Associate Professor, Department of Philosophy

Dissertation Advisor: John McDowell, University Professor, Department of Philosophy

Copyright © by Endre Begby

2008

ON THE STRUCTURE OF COMMUNICATIVE UNDERSTANDING

Endre Begby, PhD

University of Pittsburgh, 2008

Meaningfulness in human affairs manifests itself in at least three ways: in thought, in speech, and in intentional action. In our day-to-day communicative interactions, we clearly presume to draw substantive, explanatory connections between these three: we *hear* what people say, we *understand* this as the expression of cognitive activity, and we *see* them engaging in intentional actions which, when all goes well, corroborate our attributions of specific cognitive states to them.

I take this to embody rudimentary, though essentially correct, model of communicative understanding. My dissertation seeks to contribute to our understanding of this model and the constraints it imposes on a philosophical theory of meaning. A comprehensive theory of meaning, I argue, should provide a satisfactory account of all three elements of the model and of the conceptual and epistemic relations between them. Minimally, any sketch of a theory of meaning, however partial, should show it capable of serving in implementations of such a model.

The bulk of my dissertation takes the form a critical exploration of Tyler Burge's anti-individualism. Anti-individualism trades on the idea that normatively infused conventionally established concepts, shared among speakers of a language, must play a crucial role in underwriting cognition and communication. My dissertation constitutes a sustained argument to the effect that this strategy will not work. Its *prima facie* plausibility notwithstanding, I argue that theories taking this idea as their starting point will fail to satisfy the constraints laid out in the basic model, and will, as a consequence, end up misrepresenting the basic pattern of our communicative interactions.

Table of Contents

Preface	vii
Introduction	1
1. Introducing anti-individualism		
1.1	Introductory	6
1.2	Internalism, externalism, and anti-individualism.....	8
1.3	The ‘arthritis’ thought-experiment.....	11
1.4	From anti-psychologism to anti-individualism (I): Frege	16
1.5	From anti-psychologism to anti-individualism (II): Putnam.....	24
1.6	Burgean anti-individualism	29
2. Concepts: their alleged role in cognition and communication		
2.1	Introductory	37
2.2	A new look at the ‘arthritis’ thought-experiment.....	42
2.3	Two vital specifications	45
2.4	Further observations on concept attribution.....	50
2.5	The acquisition problem.....	54
2.6	Interlingual communication: a counter-thought-experiment.....	58
2.7	Conventional concepts do not explain communicative possibility	62
	Addendum.....	67
3. Incomplete concept mastery		
3.1	Introductory	81
3.2	Incomplete concept mastery: the basics	85
3.3	Objection from <i>de re</i> attitudes.....	87
3.4	Objection from indefiniteness of content.....	90
3.5.0	Objection from variant concepts	92
3.5.1	No-retroactive-reinterpretation argument.....	93
3.5.2	Shared belief argument.....	96
3.5.3	<i>Ad hoc</i> argument.....	97
3.5.4	Not-true-to-the-evidence argument	103
3.5.5	His-beliefs-aren’t-true-anyway argument.....	107
3.6	Metalinguistic reinterpretation.....	113
3.7	Philosophical arguments for reinterpretation	122

4. Cognitive perspective and self-knowledge

4.1	Introductory	133
4.2	Burgean anti-individualism and cognitive perspective introduced	138
4.3.0	The generic problem of externalism and self-knowledge	143
4.3.1	The challenge of incompatibilism	146
4.3.2	Resolving the achievement problem	149
4.3.3	Slow-switching	150
4.3.4	<i>Cogito</i> -thoughts and basic self-knowledge	152
4.3.5	Philosophical limitations of the <i>cogito</i> -argument	158
4.4	The idea of critical reasoning and its connections with self-knowledge	160
4.5	Deflationary and substantive self-knowledge	165
4.6	Critical reasoning requires substantive self-knowledge	169
4.7	Two kinds of cases	173
4.8	Closing remarks	178

5. Normativity

5.1	Introductory	181
5.2	Putnam on the division of linguistic labor	183
5.3	Normativity in “Individualism and the Mental”	185
5.4	Elms and beeches redux	190
5.5	What is the explanatory force of linguistic norms?	193
5.6	Can a reinterpretivist account for the subject’s error?	205

References	215
------------------	-----

PREFACE

I must start by thanking the members of my dissertation committee, John McDowell, Bob Brandom, Ted McGuire, Tom Ricketts, and Kieran Setiya. Special thanks are, of course, due to John McDowell, who generously agreed to supervise my work at a time when he was already overbooked and was free to decline. Throughout my work on this dissertation, John has guided me in a way that encouraged exploration and articulation of my own perspective of the philosophical matters at hand. He has consistently returned my (frequently overlong) chapter drafts quickly, often within days, yet his comments have never failed to be incisive and to the point, concerned more than anything with making my arguments clearer and more decisive. In these and many other regards, I could not have asked for a better advisor. In particular, John's ability to spot (and gently warn of) argumentative dead ends weeks and months before I would even have gotten there has saved me more time and effort than I dare imagine.

Ted McGuire has been a close friend and source of inspiration since I first started attending his seminars in 2003. Among many other things, I am particularly grateful to Ted for helping keep my philosophical spirits up during a time when I found graduate school particularly hard going. Tom Ricketts joined the committee late, but was still able to help shape several key arguments in the dissertation. Along with Steve Engstrom, Tom also deserves thanks for help and encouragement in preparing me for the hardships of the job market.

Of other friends in and around graduate school, my greatest debts are no doubt to Chris Frey and Jo-Jo Koo. Both have been ideal philosophical conversation partners over the last six years.

Chris has given generously of his extensive knowledge of the matters that I have written about, and must shoulder a large part of the blame for the project taking the particular shape that it has, however much he would disagree with many of its conclusions. Jo-Jo has also offered deep and rewarding friendship, alongside philosophical and professional advice that has helped me stay on course through times both good and bad. Further thanks are due to Jim Bogen, Henrik Forssell, Graham Hubbs, and Christine Young, as well as Ian Blecher, Brett Caloia, Jack Fortune, Jennifer Frey, Anjana Jacob, Cynthia Swinehart, Hao Tang, and Vanessa Wills. Finally, although he was never involved in forming this dissertation project, I would also like to thank Nicholas Rescher for giving generous help and encouragement at other crucial times during my graduate school career.

Some of my greatest debts go farther back. Bjørn Ramberg was my advisor at the University of Oslo and did more than anyone to ignite my interest in philosophy, help me find my bearings, and to give me a sense of the kind of philosopher I wanted to become. While conflicting duties prevented Bjørn from taking as much part in this project as both he and I would have liked, I gratefully acknowledge his support throughout my six years in Pittsburgh, and look forward to opportunities to work with him in the future.

After completing my degree at the University of Oslo, I was incredibly lucky to be invited to work with Henrik Syse and Greg Reichberg at the International Peace Research Institute, Oslo, where I stayed for almost two years. Besides being two of the finest human beings I have ever met, Henrik and Greg generously granted me scholarly opportunities far beyond what my humble station would warrant. I really cannot thank them enough for what they have taught me and the possibilities they have given me. Many other enduring friendships were forged at PRIO as well. In particular, I must thank Håvard Strand and J. Peter Burgess, who continue to challenge me with collaborative work far beyond the confines of what I learned in graduate school.

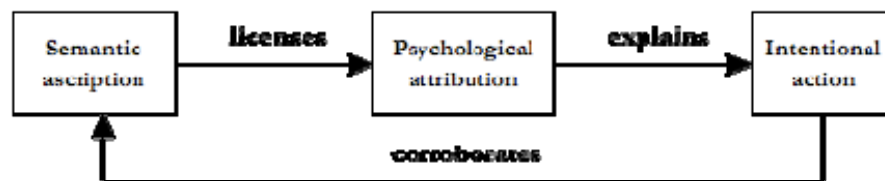
Farther back still, special thanks are due to my family and friends back home, who have never once failed to be supportive and encouraging during my long periods of absence, and who have always made returning home such a splendid experience.

Men min største taknemlighetsgjeld er selvsagt til Holly Andersen, som har gått fra å være kollega, til å være venn, til å være kjæreste, til å være samboer, til å være min kone i løpet av seks korte år. *Det* er noe å være stolt av.

I hereby acknowledge the financial support I have received from the US-Norway Fulbright Foundation (2002-03) and the Andrew W. Mellon Foundation (2003-4, 2007-08).

INTRODUCTION

Meaningfulness in human affairs manifests itself in at least three ways: in thought, in speech, and in intentional action. In our day-to-day communicative interactions, we clearly presume to draw substantive, explanatory connections between these three: we *hear* what people say, we *understand* this as the expression of cognitive activity, and we *see* them engaging in intentional actions which, when all goes well, corroborate our attributions of specific cognitive states to them. Schematically:



My dissertation seeks to contribute to our understanding of this model and the constraints it imposes on a philosophical theory of meaning. A comprehensive theory of meaning, I argue, should provide a satisfactory account of all three elements of the model and of the conceptual and epistemic relations between them. Minimally, any sketch of a theory of meaning, however partial, should show it capable of serving in implementations of such a model.

This view, with its way of mingling philosophy of language with philosophy of mind and action theory, is certainly not without its detractors. Thus, for instance, Nathan Salmon has been a vocal opponent of any view that takes our assessment of the meaning of a speaker's utterance to be in any way interestingly connected with our assessment of what belief that speaker is giving expression to. He writes: "an expression is used correctly *from the point of view of pure semantics* if and

only if it is used with its literal meaning” (Salmon, “Two Conceptions of Semantics,” p. 321). This talk of “pure semantics” might lead one to speculate about whether there could be an applied semantics, and if so, what its working principles might be. Salmon discourages this idea:

To conceive of semantics as concerned with speaker assertion (i.e., with what the speaker who uses the sentence thereby asserts) is not merely to blur the distinction between semantics and pragmatics. It is to *misidentify* semantics altogether, and to do so sufficiently badly that those who conceive semantics in this way, when using semantic expressions like ‘denote’, ‘content’, ‘true’, are often fruitfully interpreted as not speaking about the notions of denotation, content, or (semantic) truth at all, but other notions entirely – specifically various pragmatic notions. (Salmon, “Two Conceptions of Semantics,” p. 321.)

In their widely cited book *Insensitive Semantics*, Herman Cappelen and Ernest Lepore present an argument to much the same effect. Thus, they identify an assumption which they claim (plausibly) has underlain much work in 20th century philosophy of language, according to which “[a] theory of semantic content is adequate just in case it accounts for all or most of the intuitions that speakers have about [...] what speakers say, assert, claim, and state by uttering sentences” (Cappelen and Lepore, *Insensitive Semantics*, p. 53). Throughout their book, they refer to this simply as the “Mistaken Assumption.”

If this is what semantics is, they can have it. Nothing I will say in the following constitutes an argument against the viability of a pure semantics in this sense.¹ But this conception of semantics still leaves us without even the beginnings of an account of the apparently systematic relations between our assessments of what people say, what they think, and how they act, and with no reason to think that this is not a task that the philosopher can or ought to apply herself to. This is this task that I hope to contribute to here.

¹ Though I will retain the right to use terms such as ‘true,’ ‘denotes,’ ‘refers,’ and ‘semantic content,’ and leave it to the reader to determine how they should be understood in the context of the argument.

The bulk of my dissertation takes the form a critical exploration of Tyler Burge's anti-individualism. Anti-individualism is uniquely suited to serve as the starting point of my investigation in that it constitutes an attempt to combine, on the one hand, the view of the workload of a theory of meaning that I laid out above, with, on the other hand, the pure semanticist's predilection toward literal meaning. In short, anti-individualism trades on the idea that normatively infused conventionally established concepts, shared among speakers of a language, must play a crucial role in underwriting cognition and communication. My dissertation constitutes a sustained argument to the effect that this strategy will not work. Its *prima facie* plausibility notwithstanding, I argue that theories taking this idea as their starting point will fail to satisfy the constraints laid out in the hermeneutic schema above, and will, as a consequence, end up misrepresenting the basic pattern of our communicative interactions.

Thus, in Burge's famous 'arthritis thought-experiment,' we are offered the conclusion that a counterfactual subject who lacked the concept *arthritis* could not think thoughts about "arthritis as such." I show in ch. 2 of my dissertation that this is false: the counterfactual subject can perfectly well come to share in a reflective agreement with the actual subject concerning arthritis, and thereby shows himself capable of thinking the thought embodied in this agreement. He comes to share in this agreement, moreover, without undergoing any kind of change that could plausibly be described as his coming to acquire a new concept. It follows that the two subjects of Burge's thought-experiment were not, in the sense required for its conclusion to come through, *incapable* of thinking the same kinds of thoughts.

In ch. 3, I go on to show that so far from being incapable of thinking the same kinds of thoughts, Burge's argument fails even to demonstrate that the two subjects must be thinking different thoughts on the occasion in question. What we have, I argue, should rather be seen as two

competing descriptions of a single thought-content. This allows me to concede that there is a sense in which anti-individualist thought-attributions – those that involve ‘incompletely understood concepts’ – can be true. Nonetheless, as I go on to show in ch. 4, these kinds of attributions are *inappropriate*, unfit for the tasks of capturing and conveying the force of the subject’s reasoning and explaining his actions. I conclude that the explanatory power of any theory which privileges such attributions is severely constrained, in that it cannot account for the systematic unity of content in speech, thought, and action. In this sense, such attributions do not yield *understanding*, however much they may yield true descriptions of the subject’s thought contents on some occasion.

These conclusions leave public, conventional concepts with no decisive role to play in supporting or structuring our communicative interactions. In my final chapter, I examine lingering concerns that my theory leaves us unable to account for our sense that there is *error* involved in these putative cases of conceptual confusion. On this view, what favors attributions of thoughts involving public concepts even where these are not fully mastered by the subject in question is the peculiar prerogative that such concepts have vis-à-vis the cognitive possibilities of individual subjects. Thus, it is only by reference to the norms inherent in such concepts that we can say that the subject is in error. I show that my theory is quite capable of ascribing error even without invoking such concepts, and that this invocation of norms has none of the explanatory power that the anti-individualist claims for it.

The position that will emerge from this critical examination of anti-individualism does not signal a return to any form of individualism. I might rather label it, with a tip of the hat to Clifford Geertz, ‘anti anti-individualism.’² But this too could only constitute a provisional characterization. For despite the critical thrust of most of what I will have to say here, my overarching ambition is

² Clifford Geertz, “Anti Anti-Relativism.”

positive. I hope, at the end of the day, to have helped clear the ground for a form of externalism that gives due weight to the role of language and various social institutions in shaping and supporting our cognitive possibilities as human beings, but without tying these possibilities to static linguistic conventions.

1

BURGE'S ANTI-INDIVIDUALISM

1.1 Introductory

We start with an exposition of Burge's particular form of externalism, variously referred to as 'social externalism' and 'anti-individualism.' In trying to pinpoint the original contribution of Burge's arguments, we will pay particular attention to how they develop out of a certain historical and dialectical context.

As I shall be using these terms, we pass from *internalism* to *externalism* by tying the contents of the subject's thoughts and utterances to factors in that subject's natural environment. We pass from *externalism* to *social externalism* when we also bring in factors in the subject's social environment. In my view, *anti-individualism* should be thought of as a particular form of social externalism, marked by a particular view of how the social factors in:¹ anti-individualism suggests that we bring the social to bear on the theory of thought attribution through a consideration of certain practices of inquiry and classification. These practices are encoded in linguistic conventions. The hallmark of these linguistic conventions is that they carry a certain kind of normative force, such that even when a subject is unfamiliar with the norm, the norm still applies. This particular view of the way the subject's social

¹ We note that sometimes, especially in more recent writings, Burge will speak of anti-individualism also in presenting arguments that have nothing to do with a speaker's social environment. In these contexts, Burgean anti-individualism becomes practically indistinguishable from the kind of natural kind externalism usually associated with the work of Hilary Putnam. More on Putnam's view shortly.

environment bears on the proper interpretation of his words and his thoughts is what I shall be criticizing in the following.

As I explained in the preface, we are looking for a unitary notion of content that can cover semantics, psychology, and action explanation in one fell swoop. In my view, it is among the signal contributions of Burge to insist that an externalist theory should fulfill this requirement as surely as an internalist theory does. The broad theme of this dissertation is that this is a task that we cannot fulfill if we bring the social to bear on content attributions in the way that Burge proposes, that is, through the normatively infused conventions that govern our language. The social setting in which we think our thoughts, utter our words, and perform our actions forms an essential background context to the exercise of these capacities. But it does not, as I shall argue, impact on particular attributions in the way that Burge supposes. If it did, then we could not reliably use these attributions to form a picture of each other as rational thinkers and agents.

This first chapter seeks to reconstruct the emergence of anti-individualism as a philosophical position, and to chart the ideas that motivate it. In this sense, it is less an argument than a narrative reconstruction of an argument. The key pivots of this narrative reconstruction will be Gottlob Frege's "On *Sinn* and *Bedeutung*" and Hilary Putnam's "The Meaning of 'Meaning'."²

² If my aim were a full history of the developments leading up to anti-individualism, then Saul Kripke's *Naming and Necessity* and perhaps also Keith Donnellan's "Reference and Definite Descriptions" would also have to be mentioned. However, the philosophical problematic that I am trying to isolate is largely untouched by these contributions, which I will therefore pass over in relative silence. We will touch on Kripke's work mostly to note Burge's important reservations against the view that proper names and natural kind terms refer directly, that is, without the mediation of anything like Fregean *Sinne* (cf. Burge, Introduction to *Foundations of Mind*, pp. 9-13, "Belief *de re*," pp. 59-61, "Sinning against Frege," pp. 222-224). Burge himself has sought to cast the net of historical progenitors to anti-individualism progressively farther back. Hegel is mentioned in brief already in "Individualism and the Mental," but the suggestion is never followed up in detail. More lately Aristotle is added (*Foundations of Mind*, Introduction, p. 5), whereas Leibniz and Kant are treated as borderline cases (*ibid.*). Most surprisingly, Descartes is included among the ranks of anti-individualists, mainly in virtue of his conclusion that the demon hypothesis is incoherent (cf. "Philosophy of Language and Mind, 1950-2000," p. 455/26, "Descartes on Anti-Individualism").

1.2 Internalism, externalism, and anti-individualism

Burge markets his anti-individualism as a form of externalism. Yet at first glance it is not entirely clear how the divide between individualism and anti-individualism lines up with that between internalism and externalism. We receive a first hint that the two distinctions do not precisely overlap when we see that there may be externalist theories that yet remain individualistic in a damaging way: Burge's anti-individualist argument is applied in turn both to Cartesianism (presumably internalist) and various forms of behaviorism (presumably externalist).³

The crucial factor common to Cartesianism and behaviorism is a neglect of the role of the social in connection with the cognitive lives of human beings. This is their individualism: in order to get a proper fix on the cognitive states of a human being, we need cast our net no wider than the individual herself. Here it matters little whether the cognitive states of the individual are assumed to be graspable primarily or only by the individual herself (internalism)⁴ or are held to be publicly available (externalism).⁵ This individualistic trend in philosophy – whether internalist or externalist – stands in sharp contrast to “the Hegelian preoccupation with the role of social institutions in shaping the individual and the content of his thought” (Burge, “Individualism and the Mental,” p. 100). Thus, insofar as Burge is about to present us with a single, generalized argument, we may presume that the target is not specifically – or at least not only – the internalist character of the

³ In a lengthy discussion, Burge also includes functionalism (“Individualism and the Mental,” pp. 136-140). Since I am here only making the point that a philosophical position can be externalist yet individualist, I focus on the simpler case of behaviorism. Note also that ‘Cartesianism’ is here intended in the generic sense of the idea that the contents of a subject’s thoughts can be fixed without taking into account that subject’s external reality. Burge has recently argued that Descartes himself was not, in this sense, a Cartesian (cf. “Descartes on Anti-individualism”).

⁴ Cartesianism’s “secret cogitations, [...] innate cognitive structures, [...] private perceptions and introspections, [...] grasping of ideas, concepts, or forms” (Burge, “Individualism and the Mental,” p. 100).

⁵ Behaviorism’s “input-output relations, and the dispositions, states, or events that mediate them” (Burge, “Individualism and the Mental,” p. 100).

Cartesian picture of mind and meaning, but rather its tendency “to feature the individual subject,” a tendency it shares with its decidedly externalist successors such as behaviorism.

The initial impression, then, is that ‘individualism’ applies more widely than ‘internalism,’ so that anti-individualism should not be simply conflated with externalism. An explicit statement in “Individualism and the Mental” seems to confirm this: ‘individualism’ applies to

philosophical treatments that seek to see a person’s intentional mental phenomena ultimately and purely in terms of what happens to the person, what occurs within him, and how he responds to his physical environment, without any essential reference to the social context in which he or the interpreter of his mental phenomena are situated. (“Individualism and the Mental,” pp. 132-133)

Two things are immediately notable about this rough definition: i) individualism, as noted, cuts across the distinction between internalism and externalism,⁶ and ii) it is crucial for anti-individualism that the external relations considered be not just physical relations, but social relations as well. There is, that is to say, no strain in imagining a position that is externalist, yet individualist in the relevant sense.⁷

This focus on the role of the social environment is presented upfront in Burge’s article, and is widely regarded as the most important contribution of anti-individualism. Hilary Putnam had already developed explicitly externalistic ideas based on arguments tracing relations between the individual and her physical environment. Burge now adds considerations concerning the relation

⁶ “... what occurs within him, and how he responds to his physical environment” – I take it that the ‘and’ here should be understood as the inclusive ‘or.’

⁷ We note right away that this leaves untouched a third option, which is roughly speaking the one that I will be defending. In this view, the social plays an essential role in explaining the emergence of the broad set of abilities the exercise of which we are here trying to understand. Moreover, reference to the specific form that social institutions have taken in some local context can certainly play an important role in explaining the thoughts and actions of the individuals that find themselves in this context. But it cannot play an *essential* such role, and this is what is at stake in the argument. Burge argues that one who was not a member of this social practice *could not* be thinking the kinds of thoughts that we are considering. On the other hand, if you *are* a member of this social practice, then the norms and conventions that govern it have a certain kind of power over you, even in cases where you might be unaware of such a norm. These are the ideas that will concern us here.

between the individual and her social environment. Broadly speaking, then, we pass from internalism to externalism by bringing in considerations involving the individual's physical environment; we pass from externalism to anti-individualism by bringing in considerations regarding the individual's social environment.⁸ But as we shall see, with Burge's wider and more ambitious form of externalism comes also another move: unlike Putnam's externalism, Burge's anti-individualism concerns not only the meaning of our words, but also and primarily, the content of our mental states.

There are two key elements of Burge's position, then, that we will be trying to isolate in this chapter: (i) the idea that our externalism should encompass the philosophy of language and the philosophy of mind – semantics and psychology – in the same breath;⁹ and (ii) the idea that social institutions of a certain kind have an essential role to play not only in giving meaning to our words, but also in providing content to our psychological states.

The next section (1.3) outlines the argument Burge uses to establish his anti-individualism, i.e., the famous 'arthritis' thought-experiment. The following sections present a small historical sketch of some philosophical developments leading up to Burge's anti-individualism, focusing on Frege (1.4) and Putnam (1.5). Together, these two sections will equip us to home in on some of the key questions that will occupy us in this dissertation, which we will first try to articulate in section 1.6: the transition from a merely semantic externalism to an externalism that encompasses the philosophy of language and the philosophy of mind in the same breath; and the transition from

⁸ Burge adds in a recent Postscript that at the time of writing "Individualism and the Mental," he still regarded physical relations as "more fundamental" in "determining the natures of mental states," and that he focused "first on the social environment because [he] thought that its role was less close to the surface, less easily recognized" ("Postscript to 'Individualism and the Mental,'" p. 153). This comports well with Burge's more recent focus on emphasizing commonalities in the cognitive lives of human beings and non-social animals (cf. "Perceptual Entitlement" as well as his John Dewey lectures *Self and Self Understanding*, delivered at Columbia University, December 2007). The view that our physical environment has a certain priority over our social environment in "determining the natures of mental states" and certainly in anchoring our communicative possibilities is quite congenial to me, and central to the argument that I develop in ch. 2.

⁹ I leave out any explicit mention of action explanation here, on the assumption that it falls naturally under the heading of philosophy of mind. Our mental states are what explain our actions.

physical externalism to social externalism. In both cases, we will find reasons to agree with the motivations for Burge's move. But we will also find that there are deep, unresolved flaws in Burge's particular way of executing these transitions.

1.3 The 'arthritis' thought-experiment

Burge frequently presents the philosophical issue between individualism and anti-individualism in metaphysical terms, in terms of the type-individuation of mental states and events. A particularly crisp example of this approach can be found in "Individualism and Psychology":

According to individualism about the mind, the mental natures of all a person's or animal's mental states (and events) are such that there is no necessary or deep individuating relation between the individual's being in states of those kinds and the nature of the individual's physical or social environment. (Burge, "Individualism and Psychology," pp. 221)

Anti-individualism, then, is the claim that there are "deep or necessary individuating relation[s]" between (at least some of) our mental states and our physical and social environment. Although this way of putting the point was never upfront in "Individualism and the Mental," similar turns of phrase crop up in Burge's recent Postscript to that article. The argument, he says, is "about how *having* certain thoughts constitutively depends on relations to the environment;" further, "anti-individualism is about the conditions under which mental states and events can have the representational contents that they have" (Burge, "Postscript to 'Individualism and the Mental,'" p. 156).¹⁰

¹⁰ See, further, Burge, "Postscript to 'Individualism and the Mental,'" p. 162, Burge, "Phenomenality and Reference: Reply to Loar," p. 438.

We should understand Burge's inquiry, then, in terms of the following question: what must be true of the individual's environment (physical or social) in order for the individual to be able to count as instantiating some propositional attitude p ? The internalist answer, according to Burge, is 'nothing': quite generally, the individual's environment introduces no specific constraints on the kinds of mental states which that person might be in. We may conceive of any manner of changes to the individual's environment without this affecting her mental contents in the least. An externalist, by contrast, will insist that certain things must be true of the subject's physical environment for her to be in this or that mental state.¹¹ As I have been arguing, however, an externalist position can remain individualistic: this would be a matter of denying that there are similarly essential connections between the subject's mental contents and his social environment.

Since our critical focus in the following will mostly be on the role of the social environment, we can take the claim concerning the crucial role of the physical environment as demonstrated by Hilary Putnam's "The Meaning of 'Meaning'" (reviewed in section 1.5 below), with crucial emendations by Burge.¹² The success of anti-individualism will hinge on finding a counterexample to the claim that the subject's social environment plays no essential role in individuating her mental states.

It is Burge's famous 'arthritis' thought-experiment which is supposed to provide this counterexample and thereby establish the truth of anti-individualism. The thought-experiment

¹¹ This statement masks an important distinction that will crop up occasionally in this dissertation (cf. section 4.3.4), namely that between a *generic* form of externalism and what we might call a *locally-specific* form of externalism. According to the former, a subject's having any particular propositional attitude does impose constraints on the environment that he finds himself in, but no *specific* constraint: what is required is merely that the subject finds himself in an environment of a suitable *generic complexity*, as one might say. On Putnam's and Burge's views, by contrast, there appear to be *local and specific* constraining relations between the kinds of propositional attitudes a subject is capable of tokening and the kind of environment he finds himself in. Thus, it seems, a subject could not think water-thoughts in water-less environments, could not think arthritis-thoughts in the absence of a social practice of classifying rheumatoid ailments in a certain way, and so on. As we shall see, Burge will tend to moderate such statements under pressure. It is doubtful, however, that the metaphysically laden programmatic statement of anti-individualism will survive such moderation.

¹² See, specifically, Burge, "Other Bodies" and the Introduction to *Foundations of Mind*, pp. 9-14.

proceeds in three steps. In the first step, we are to imagine a speaker who, in spite of a generally good grasp of his language, makes a conceptual error of a certain kind, for instance in uttering the words “I have arthritis in my thigh.” The error here is conceptual, not merely empirical, because in his language, the concept *arthritis* means ‘rheumatoid ailment of the joints,’ and so does not apply to ailments elsewhere, quite irrespective of whether they have striking etiological and phenomenal similarities to arthritis. We are to ask, what thought is this person giving expression to? Our initial inclination might be to protest that it could not be a thought involving our communal concept *arthritis*, since that concept just means ‘rheumatoid ailment of the joints.’ However, this knee-jerk reaction is supposed to crumble in the face of deeper reflection on the case. In fact, our initial inclination showcases a peculiar blind spot in our thinking about these matters, which it is crucial for Burge’s argument that we isolate and contain. This is the idea that *we can only think with concepts we completely understand*.¹³ Burge’s doctrine of incomplete concept mastery – that we can think with concepts we do not fully master – is tremendously exciting and perhaps the most controversial aspect of Burge’s argument. It is the topic of chapters 3 and 4 of this dissertation.

In the thought-experiment’s second step, we are to consider a physical replica of the original subject. This person is qualitatively identical to the first in all respects but one: she is a member of a linguistic community in which the term ‘arthritis’ applies also to rheumatoid ailments outside the joints.¹⁴ In this counterfactual scenario, an utterance of the syntactic string “I have arthritis in my thigh” would not signal a conceptual error, but a proposition which may or may not be true.

¹³ See, in particular, “Individualism and the Mental,” p. 131. To forestall confusion, I ask the reader to note that where I speak of ‘concepts’, “Individualism and the Mental” frequently speaks of ‘notions’ (defined in “Individualism and the Mental,” p. 102). Burge later switches back to the more widespread nomenclature.

¹⁴ Burge presents the thought-experiment in terms of a single subject embedded in two different social contexts, whereas I present it here in terms of two different individuals. The difference has no bearing on the conclusion, and I do it this way merely in order to distinguish the questions that arise from the original thought-experiment from those that arise in connection with the question of ‘slow-switching,’ which we will review in chapter 4.

We must observe that what we have here is not a single concept *arthritis*, which happens to have a different extension in the two communities. Rather, the difference in extension shows that we are in fact speaking about two different concepts.¹⁵ Thus, we might say that in Community₁, the syntactic string “arthritis” expresses the concept *arthritis* (extending to rheumatoid ailments of the joints only), whereas in Community₂, the syntactic string “arthritis” expresses the concept *tharthritis* (extending to all rheumatoid ailments).

The following diagnosis suggests itself. In the actual situation, a correct interpretation of the words “I have arthritis in my thigh” (as spoken by the subject in question) yields the proposition *I have arthritis in my thigh*, whereas in the counterfactual situation, a correct interpretation yields the proposition *I have tharthritis in my thigh*. Step 3 of the thought-experiment draws a comparative judgment over the two cases: assuming a tight connection between semantic ascription and psychological attitude attribution (a correct interpretation of a sincere first person report yields the attitude expressed in that report),¹⁶ the thought experiment suggests that the two persons entertain different psychological attitudes in the two cases. And this in spite of the fact that everything else about them (including their speech patterns and verbal dispositions) is the same. Burge draws the following startling conclusion:

In the counterfactual situation, the patient lacks some – probably *all* – of the attitudes commonly [i.e., actually] attributed with content clauses containing ‘arthritis’ in oblique occurrence. He lacks the occurrent thoughts or beliefs that he has arthritis in the thigh, that he has had arthritis for years, that

¹⁵ See Burge, “Individualism and the Mental,” p. 106.

¹⁶ This assumption can certainly be doubted, and some might indeed take the kind of arguments offered by Burge to be precisely the arguments we need to substantiate the doubt. For more on this, consult Brian Loar, “Social Content and Psychological Content.” The tight connection between semantic ascription and psychological attribution is, however, a foundational assumption that Burge and I share, and so I will not go into further details on this matter here.

stiffening joints and various sorts of aches are symptoms of arthritis, that his father had arthritis, and so on. (Burge, “Individualism and the Mental,” p. 106)¹⁷

If the inference to this conclusion is warranted, then it provides Burge with the counterexample which refutes individualism. For what it would show is that there is at least one kind of thought – arthritis-thoughts – which the individual could not have unless certain things were true of her environment. One specific thing which must be true of her environment is that the members of her speech community use ‘arthritis’ to refer to rheumatoid ailment of the joints only. Since the counterfactual community does not have the *arthritis* concept, subjects in the counterfactual community have no *arthritis*-thoughts, and, moreover, since the argument is couched in terms of conditions of possibility, *could not have* any such thoughts. This conclusion is taken up for critical scrutiny in chapter 2 of this dissertation.

Arthritis may seem like an esoteric example, but according to Burge, the underlying principles have an extremely wide application. The kind of incomplete concept mastery that drives the thought-experiment, he says, may be “*common or even normal in the case of a large number of expressions in our vocabularies*” (Burge, “Individualism and the Mental,” p. 112). In particular:

[the argument] does not depend [...] on the kind of word ‘arthritis’ is. We could have used an artifact term, an ordinary natural kind word, a color adjective, a social role term, a term for a historical style, an abstract noun, an action verb, a physical movement verb, or any of various other sorts of words. (Burge, “Individualism and the Mental,” pp. 106-7)

¹⁷ As we shall see, the wording of this conclusion (“the patient lacks some – probably all” of these thoughts) is significantly weaker than what Burge’s argument, if true, would entail, and weaker than what it is explicitly taken by Burge to entail elsewhere. In ch. 2, I will argue that anti-individualism can only be sustained on a modally inflected reading of the connection between concept possession and cognitive ability: if our subject did not have this concept, he *could not* be thinking any of these thoughts.

This should suffice to give us a sense of how widely the argument is supposed to apply. Any attempt at a more precise delineation may prove pointless at this juncture.

Thus, the type-individuation of a certain kinds of psychological attitude (of which arthritis-thoughts serve just as a convenient example) is sensitive to factors external to the individual agent, specifically to social factors such as the linguistic habits of the community in which the agent finds himself.¹⁸ The argument can thereby claim to have secured both externalism and anti-individualism in one fell swoop. We must conclude, according to Burge, that “the patient’s mental contents differ while his entire physical and non-intentional mental histories, considered in isolation from their social context, remain the same” (“Individualism and the Mental,” p. 106). Thus, individualism is false, and anti-individualism is established as true.

1.4 From anti-psychologism to anti-individualism (I): Frege

In order to fully appreciate the unique and radical nature of Burge’s proposal, it will be helpful to consider a small historical sketch. A natural starting point in telling the story of the emergence of anti-individualism is, not surprisingly, the work of Gottlob Frege. The characteristic Fregean move that set the tone for much subsequent work in the philosophy of language is what is often characterized as his *anti-psychologism*.¹⁹ Frege targets the time-honored idea that meaning, or meaningfulness, has its source in psychology, in the mental operations of individual subjects. According to the traditional picture, the subject quite literally brings meaning into existence through

¹⁸ See for instance the following statement of the argument of “Individualism and the Mental” from the paper “Other Bodies” (p. 84): “In ‘Individualism and the Mental’ I presented a thought experiment in which one fixed non-intentional, individualistic descriptions of the physical, behavioral, phenomenalistic, and (on most formulations) functional histories of an individual. By varying the *uses of words* in his linguistic community, I found that the contents of his propositional attitudes varied.”

¹⁹ This phrase is almost universally employed in the secondary literature, although I do not know of a single textual source where Frege himself employs it. However, locations are legion where he expresses opinions which can justifiably be summed up in that phrase, e.g., “Thought,” pp. 325-326, 337nE; “Letters to Husserl, 1906,” p. 302.

the activity of thinking, meaning which may subsequently find its expression in overt speech, if the subject is so moved. Therefore, it would seem that in order to understand linguistic expressions, we need first to understand the psychological stirrings in the individual which led to these expressions in the first place.

Frege, quite rightly, finds this idea to be shot through with problems. It bears remarking that it matters little here whether the thought-processes in question are considered to be unique to the individual (as, for instance, in Romantic hermeneutics)²⁰ or whether they are thought to fall under empirical patterns describable by a scientific psychology.²¹ The key point is that both Romantic hermeneutics and scientific psychology must fail to capture what is for Frege the true hallmark of meaning – namely its *normativity*. Thus, Frege will speak of the *laws of thought*, by which he means not empirically discernible regularities in how people actually think, but how they *should* think. The aspect of Frege’s project which will concern us here consists, then, in sharply distinguishing the normative discipline of semantics (or logic) from the descriptive discipline of psychology.²²

In spite of its critical thrust, the terminology which Frege adopts to make his point sounds at first blush very much like the very traditional one. The relationship that the theory of meaning must focus on, he argues, is that between *Sinn* and *Bedeutung*. However, the essential difference from the traditional picture soon comes to the fore. The *Sinn-Bedeutung* distinction applies both at the level of (sub-sentential) terms or expressions and at the level of sentences or judgments. In what follows, I will focus on the application of this distinction to singular terms and sentences, thereby setting aside the complications that arise from Frege’s treatment of general terms and predicates.²³ Singular terms

²⁰ For instance, Friedrich Schleiermacher, *Hermeneutics and Criticism*. See also Hans-Georg Gadamer’s illuminating discussion in *Truth and Method*, pp. 184-197.

²¹ William James’s *Principles of Psychology* could serve as an example of this. I mention James’s work specifically because it synthesizes an impressive array of mostly German empirical psychology, roughly contemporaneous with Frege.

²² See, e.g., Frege, “Thought,” pp. 325-326.

²³ Most famously, perhaps, in “On Concept and Object.” Thus, the following terminological clarification is in order. I am primarily interested in Frege’s theory of *Sinn* as a contribution to the theory of *concepts*, in the current sense of that

first: *Sinne* – senses – are *modes of presentation of objects* (Frege, “On *Sinn* and *Bedeutung*,” p. 152). In Frege’s celebrated example, ‘the Morning Star’ and ‘the Evening Star’ are two *Sinne* for the same object, Venus, which thus serves as the *Bedeutung* of both. That either expression is *Sinnvoll* at all is due to the fact that it denotes an object. Yet if both expressions denote the same object, why do they mark different senses? To see this, we must consider how the expressions fare when embedded in judgments. That they are different senses even as they denote the same object shows itself in the fact that someone could fail to acknowledge that the Morning Star *is* the Evening Star, without this in any way impugning her rationality. Doubting this identity statement is a matter of failing to know an empirical fact, not a failure of rationality. ‘The Morning Star’ and ‘the Evening Star’ are different senses for the same object in that they each present that object in a different light. Thus the *differential (rational) dubitability* of the judgments into which these terms are embedded is the test of whether these terms express different senses.

As finite thinkers, we can never apprehend all aspects of an object at the same time: our understanding is always perspectival. A sense is just such a perspective, a way in which a finite mind can connect with objects. Yet the observation that our cognition is irreducibly perspectival should not be thought to entail that our manner of connecting to objects is irreducibly subjective.²⁴ Senses

term. Thus, depending on how circumspect one wants to be, one might say that Fregean *Sinne* express concepts, or embody concepts, or even just *are* concepts. We need to note, however, that all of these are quite at odds with Frege’s own usage, particularly clear in his treatment of general terms and predicates, according to which *Begriffe* are located on the side of *Bedeutungen*, not on the side of *Sinne*. Thus, predicates refer to (*bedeuten*) *Begriffe*, somewhat as singular terms refer to objects and sentences to truth values (cf. Frege, “On Concept and Object,” pp. 182-183; “Letter to Husserl, 25.5.1891,” pp. 149-150; Michael Dummett, *Frege: Philosophy of Language*, pp. 173-174; Michael Beaney, *Frege: Making Sense*, pp. 186-188). However, Frege’s choice of introducing concepts as the referents of predicates is quite idiosyncratic and entirely motivated by theoretical concerns that I do not share. Thus, I follow a host of contemporary writers in treating Fregean *Sinne* as concepts, or at least as standing in a one-to-one relation with concepts. Pertinently, Putnam’s discussion in “The Meaning of ‘Meaning’” reflects this view, as does Burge’s usage, except on occasions when he is explicitly discussing Fregean *Begriffe*. Christopher Peacocke’s *A Study of Concepts*, pp. 2-3, nicely lays out the rationale for treating Fregean *Sinne* in this fashion. Moreover, this usage seems also to have become commonplace philosophy of mind; see, e.g., Stephen Laurence and Eric Margolis, “Concepts and Cognitive Science,” pp. 5-8, Jesse Prinz, *Furnishing the Mind: Concepts and their Perceptual Basis*, pp. 6-8, 281-282, Jerry Fodor, *Concepts: Where Cognitive Science went wrong*, pp. 15-21.

²⁴ Rather as, say, Leibniz argued that each monad is a unique and distorted representation of the world. See, e.g., Leibniz, *Monadology*, §§ 56-57, *Discourse on Metaphysics*, § 9.

are objective too. Frege is clear that we must distinguish the sense of an expression from any subjective associations – ideas – a thinker may have with it:

A painter, a horseman, and a zoologist will probably connect different ideas with the name ‘Bucephalus.’ This constitutes an essential distinction between the idea and the sign’s sense, which may be the common property of many people, and so is not a part or a mode of the individual mind. For one can hardly deny that mankind has a common store of thoughts which is transmitted from one generation to another. (Frege, “On *Sinn* and *Bedeutung*,” p. 154)

That we are able to communicate or to transmit knowledge at all would be inexplicable, Frege suggests, unless we could draw a distinction between objective, shareable senses and the subjective associations that these senses may evoke in the minds of each thinker. The idea that communication would be impossible were it not for shared *Sinne* is one of Frege’s most lasting and powerful ideas.²⁵ It is clearly instrumental in shaping Burge’s project, which, if I am right, can be read precisely as an attempt to satisfy this Fregean publicity constraint without relying on assumptions about a mysterious third realm of Platonic entities,²⁶ but rather by simply tying each *Sinn* – now understood as concepts – to linguistic conventions and the social practices that they embody.

The application of the *Sinn-Bedeutung* distinction to thoughts is a more tangled affair, and one on which Frege’s proposal has engendered a whole lot more controversy. For instance, one would readily assume that “the Morning Star is visible in the morning” and “Venus is visible in the

²⁵ I should say that it is also in my view a quite misguided idea, which has had a hugely detrimental impact on philosophy of mind and language over the last 100 years or so. Ch. 2 of this dissertation is devoted to showing that shared concepts are not required for sharing thoughts, and so that such concepts do not have the explanatory power with respect to communication that the Fregean tradition assumes that they must have. It belongs to a larger project to show that we should not, in fact, think of communication in terms of jointly attending to a proposition at all, if a proposition is something that corresponds to the meaning of a sentence in the natural language of each communicant.

²⁶ Cf. Frege, “Thought,” pp. 336-337. As it happens, I am not particularly worried about positing a distinct ontological category of non spatio-temporal entities *per se*. I am with the skeptics, however, in holding that careful explanation is required to say what kind of role such entities can play in the lives of individual thinkers. Burge offers powerful and original ideas on this topic. See, e.g., his “Frege on Knowing the Third Realm” and the Introduction to *Truth, Thought, and Reason: Essays on Frege*, pp. 30-32.

morning” are senses sharing a *Bedeutung*. And so they are, but not in the way that one might think. *Bedeutungen* of thoughts, on Frege’s view, are truth-values (Frege, “On *Sinn* and *Bedeutung*,” pp. 157-158). And so “the Morning Star is visible in the morning” and “Venus is visible in the morning” are indeed senses sharing a *Bedeutung*, but so are “grass is green” and “snow is white.” I cannot here follow up the controversies surrounding this Fregean doctrine, but wish to field the suggestion – hopefully to be substantiated later – that this testifies to the awkwardness of maintaining a Fregean *Sinn-Bedeutung* distinction as we move from considering subsentential expressions to considering full-fledged judgments.²⁷

In sum, the *Bedeutung* of a singular term is the object denoted, the *Sinn*, the term’s mode of presenting the object. The *Bedeutung* of a proposition is its truth-value, the *Sinn*, the thought contained in it. Frege’s notion of *Sinn* differs from the candidate notions offered by his psychologistic predecessors in that it is not a subjective, unique tokening, but an objective content, capable of being shared by several speakers. Frege would thus be free to say that *Sinn* determines *Bedeutung* without falling into the form of subjectivistic psychologism that we discussed at the outset.²⁸

That Frege employs the word ‘thought’ [*Gedanke*] to denote the *Sinn* of a proposition is, of course, striking. With regard to the project of separating the normative study of semantics from the

²⁷ In particular, I will argue that this awkwardness impacts Burge’s argument in the following way. It would seem, on the one hand, that we understand what it would be to think with a concept that one incompletely masters, for we can model it simply enough in terms of using a word that one does not fully know the meaning of. But the doctrine of incompletely understood concepts has as its corollary the doctrine of incompletely understood thoughts. As I will show in ch. 4, we have no notion of a thought that is incompletely understood in the way that Burge depicts, but which can also play the role in rationalizations of that subject’s actions that our theory requires. If this is right, then we have reason to conclude that initial impressions were misleading – incompletely understood words do not in fact give us a good grasp of the notion of incompletely understood concepts, if concepts are, as Burge and the Fregean tradition suppose, the constituents of thoughts.

²⁸ For Frege’s claim that to every *Sinn* there corresponds a definite *Bedeutung*, see “On *Sinn* and *Bedeutung*,” p. 153. He observes, however, that natural language often frustrates us in this regard: in actual fact, there may be words without definite senses as well as senses that do not correspond to any object. So much the worse for the claim of natural language to be a fully adequate conveyor of meaning. (It is doubtful that Burge would want to follow Frege in this distrust of natural language.)

descriptive study of psychology, however, this nomenclature should not perturb us in the least. Indeed, there is a sense in which this usage might spur us toward an even clearer sense of our theoretical objective. For Frege's category of 'thought' contains no essential reference to a process of *thinking*. Instead, any given thought is a repeatable content-*type* of which a thinking may be a *token* instantiation. The thought *p* and a thinking that *p* stand to each other, if all goes well, as universal to particular.²⁹ And it is with the characteristics and laws that govern the relations between the universals, not the particulars, that Frege's theory is concerned. Thus, for instance: "Thoughts are not mental entities, and thinking is not an inner generation of such entities but the grasping of thoughts which are already present objectively" (Frege, "Letters to Husserl, 1906," p. 302).

What this amounts to is, in Michael Dummett's useful if controversial phrase, "the extrusion of thoughts from the mind" (Dummett, *Origins of Analytical Philosophy*, p. 23). As the above quote suggests, it is characteristic of Frege's philosophy that he tries to latch this distinction between universal and particular on to a distinction between objective and subjective. Thus, Dummett comments further: "Frege maintained a very stark dichotomy between the objective and the subjective, recognizing no intermediate category of the intersubjective. The subjective was for him

²⁹ Which is not to say that thinking – grasping a thought – is merely a matter of (passive) instantiation. The question of what exactly grasping a thought might consist in on a Fregean theory is a difficult one, much complicated by the rather mysterious character of the third realm and its denizens (cf. Michael Dummett, *Origins of Analytical Philosophy*, ch. 10). Putnam, as we shall see, seems to hold that in spite of all the fine talk of anti-psychologism, Frege simply doesn't have a theory of grasping which is not a just throwback to the old psychologistic theory (cf. Putnam, "The Meaning of 'Meaning'," p. 222). In an interesting discussion, Burge distinguishes two lines in Frege's work on what grasping a thought consists in. On the one line, grasping a thought is a simple matter – one merely thinks it. On the other line, grasping a thought is a cognitive achievement of a sort that may elude even the best among us (thus, Newton may have failed to fully grasp his own concept *limit*, and thereby failed to fully grasp his own thoughts). Burge acknowledges the sense that there is a tension between these two lines, but argues that Frege was eventually able to resolve it by appeal to what is, in effect, a variation on Burge's own doctrine of incomplete concept mastery: one can *think* a thought without fully *grasping* it, i.e., without fully mastering the concepts that are contained in it (Burge, "Frege on Sense and Linguistic Meaning," pp. 258-262). In my view, this utterly fails to resolve the tension in question, and merely provides a redescription of it. Our question now becomes what we might mean by saying that a subject may "think" a thought that he does not fully "grasp." More on these matters in ch. 4.

essentially private and incommunicable; he therefore held that the existence of what is common to all must be independent of any” (ibid.).³⁰

Dummett is right to highlight a certain danger involved in the Fregean project. The boon of gaining the prospect of an uncluttered and systematic theory of the relation of *Sinn* to *Bedeutung* may come at a certain price, namely our inability to say much that is useful regarding how meaning relates to the *actual* psychological processes of people, specifically, regarding the actual thought processes of individuals on occasion. Frege’s move might be unproblematic for anyone who is comfortable with holding their philosophy of language (semantics) at arm’s remove from their philosophy of mind (psychology). However, a distinct unease with this separation has come over recent work in philosophy.

One source of this unease is the emergence of the Chomskyan program in linguistics and the impressive body of work in cognitive science it has precipitated.³¹ Much of the research carried on under this banner is clearly empirical in nature, and features a distinctive focus on individual psychology. Thus, it would appear that sticking doggedly to the Fregean program will force philosophers to either set about showing why this empirical program is misguided, or find some way

³⁰ Again, it should be noted, Dummett’s interpretation is not uncontroversial, and has met with resistance, notably from Tyler Burge. It is, I believe, in the conjunction of these two Dummettian lines of interpretation that Burge finds cause for concern. First, the coinage “the extrusion of thoughts from the mind” suggests that the mind must somehow now traverse a great chasm in order to ‘grasp’ thoughts. This, to Burge’s thinking, is a completely mistaken way of looking at it. (See Burge, *Truth, Thought, and Reason*, “Introduction,” pp. 29-32, in particular p. 31.) Moreover, although Burge never says so explicitly, it would seem that if the anti-individualist line is sound, then the apparent problem of the missing intersubjective dimension falls away as soon as we rid ourselves of these confusions regarding individuals’ ability to grasp thoughts. According to Burge, it would seem that the objectivity of thoughts just is their manner of being intersubjectively available. Indeed, on Burge’s reconstruction of Frege, positing the objectivity of thought – the idea that a thought can be the common property of many thinkers, and so cannot be constitutively dependent on any of them – is our only way of getting even a first foothold with the task of explaining the very possibility of (intersubjective) communication. Whether this last claim is true, and whether, in general, this is a sustainable line of thought on the part of Burge’s Frege, are questions which will become crucial as we go on.

³¹ The *locus classicus* of Chomsky’s program is his 1957 work *Syntactic Structures*. A recent source of Chomsky’s opinions about current work in philosophy of language and mind is his *New Horizons in the Study of Language and Mind*. A good, if somewhat dated, overview of Chomsky’s impact on linguistics and cognitive science is Howard Gardner, *The Mind’s New Science: A History of the Cognitive Revolution*, ch. 7. For more recent discussions, see Louise M. Antony, “Rabbit-Pots and Supernovas: On the Relevance of Psychological Data to Linguistic Theory,” Stephen Laurence, “Is Linguistics a Branch of Psychology?” Michael Devitt, “Linguistics is not Psychology.”

by which to insist that there is room for two distinct programs for the study of language and meaning, one philosophical and one empirical.³²

This source of unease, while important, will not be much in focus in this dissertation. Instead, most of what I will have to say will constitute an attempt to come to grips with another sense of unease. This other sense of unease stems from the commonsensical idea that ascribing meaning to the words of others is our primary way of learning what they are thinking and that a large part of the reason why we speak at all is to communicate to others what is on our minds.³³ It should be clear why Frege's insistence on a strict separation between semantics and psychology appears, *prima facie* at least, to be in conflict with this commonsensical intuition. But it should also be clear that the commonsensical intuition mentioned above is not *just* about the relation between semantic ascription and psychological attribution. It also includes, as a natural and ineliminable element of this scheme, the explanation of action.³⁴ Briefly stated: the project of ascribing meaning to a person's words is central to our project of attributing content to a person's mental life. However, such content ascription is not typically an end in itself. Rather, a large part of the point of ascribing psychological content to a subject is that it helps us in explaining her actions.³⁵ What we

³² Chomsky is evidently extremely sceptical, as can be gleaned from many of the essays collected in his *New Horizons in the Study of Language and Mind*. Arguing for a pluralist line is always going to be difficult, but not ultimately impossible. I take it that Burge's own view would fall somewhere in this category. Indeed, it is tempting to speculate that Burge's insistence that Fregean theory is not ultimately about natural language and ordinary communication is partly motivated by the need to temper the sense of conflict between the two approaches. For although Burge occasionally also likes to don his science-hat (see, for instance, "Comprehension and Interpretation," pp. 231-232, "Perceptual Entitlement"), he has never, to my knowledge, confronted in writing the widespread sense that there must be a conflict between the two programs.

³³ Jennifer Hornsby's "Semantic Innocence and Psychological Understanding" is a good introduction to the motivations underlying philosophical defenses of this commonsense intuition.

³⁴ I know of no recent philosopher who defended and explored this idea in more detail than Donald Davidson, which is why his work is so important to my own. It also looms large on the horizon of Akeel Bilgrami's *Belief & Meaning*, from which I have learned much. Burge rarely addresses action explanation in any detail, but it is clear that he understands it as an integral part of the scheme of semantic ascription and psychological attribution. On this, see "Other Bodies," p. 84, 92, "Individualism and the Mental," p. 118.

³⁵ To say that this is part of 'the point' of content attribution is, I admit, somewhat tendentious, as if there needs to be any specific point, however complex, that such attribution subserves. For this reason, I choose to be rather non-committal about just how much of the point of content attribution resides in action explanation. Following the lead of Wilfrid Sellars's famous 'Myth of Jones' (*Empiricism and the Philosophy of Mind*, sections 48 onwards), Donald Davidson (cf.

learn in the effort at explaining her actions can in turn feed back on and correct our initial understanding of her words, and thus lead to a new, refined attribution of thought.

Much of Burge's work, both that which explicitly addresses Frege and much that does not, can be read as an attempt to show that the broad structure and ambition of a Fregean theory is after all compatible with this commonsensical intuition. After reviewing Hilary Putnam's contribution to the development of anti-individualism immediately below, we shall see more about how Burge hopes to effect this reconciliation.

The specific question that we will be addressing is thus whether Burge's particular manner of resolving the tension between the Fregean ambition and the commonsensical intuition is one that allows action explanation to be brought naturally into the fold. Our suspicion, of course, is that it does not. This will be explored further in chapter 4, where I ask whether anti-individualist thought-attributions can serve to capture a subject's cognitive perspective in the right kind of way.

1.5 From anti-psychologism to anti-individualism (II): Putnam

Putnam's contribution to the development of anti-individualism is vital in several ways. In a work that was to wield enormous influence, Saul Kripke had already explored the idea that what an expression refers to may be determined by the nature of the referent causally responsible for an individual subject's representation, rather than anything intrinsic to the specific descriptive content

"Mental Events," "Psychology as Philosophy") and, perhaps more strongly, Daniel Dennett ("Intentional Systems," "True Believers") have toyed with the idea that the *entire* point of semantic ascription and content attribution is the explanation of action. Criticism of this view is not hard to come by. In my view, one of the more interesting such criticisms can be found in Richard Moran's "Interpretation Theory and the First Person," where it is argued that proper attention to the unique features of self-knowledge forces the conclusion that the "primary and meaning constituting use of the concepts of commonsense psychology" is not the explanation and prediction of behavior (p. 172). There is much to be learned from Moran's argument, although I think that the important thrust behind Davidson's (and Dennett's) theories can easily survive a curbing of their initially instrumentalistic excesses. Moreover, I am worried that Moran's argument overlooks the fact that many of the factors that he rightly claims are neglected in the account of self-knowledge are also present in our understanding of others. But this must remain a topic for another occasion.

of the representation.³⁶ But it was Putnam who took these ideas farther and also designed the basic argument structure by means of which such externalism would continue to be explored and tested. In the same article, he hypothesized further that externalism could illuminate a rarely noted phenomenon which he called ‘the division of linguistic labor.’ This was to have an enormous influence on Burge’s program. Finally, and less often appreciated, Putnam’s article provided an important philosophical articulation of the two sources of unease with the Fregean program that we mentioned above, thereby giving the anti-individualist program much of its distinctive shape and focus.

It is noteworthy that Putnam opens “The Meaning of ‘Meaning’” by lamenting the current status of the study of semantics. In particular, he points to the scientific revolution in the study of syntax following the work of Noam Chomsky and Zelig Harris, and assigns a large part of the blame for the ‘prescientific’ state of the study of semantics on the continuing influence of Frege.³⁷ As a first step toward rectifying this situation, Putnam offers an argument which purports to show that two theses characteristic of the Fregean program in semantics cannot jointly be true:

- i) the thesis that knowing the meaning of a term is a matter of being in a certain psychological state, and
- ii) the thesis that meaning – i.e., *Sinn* or intension; what is known through this psychological state – determines the extension of the term.³⁸

³⁶ Cf. Saul Kripke, *Naming and Necessity*.

³⁷ He also mentions Rudolph Carnap as working from Fregean presuppositions. See Putnam, “The Meaning of ‘Meaning’,” pp. 218-219.

³⁸ Putnam, “The Meaning of ‘Meaning’,” p. 219.

Readers of Frege will perhaps be inclined to protest that the first of these theses cannot plausibly be said to be a *Fregean* thesis. Quite the contrary, this would seem to be precisely the kind of idea that his anti-psychologism is an argument against. Putnam can here counter, not without some plausibility, that we have misunderstood his (and Frege's) point. Frege's anti-psychologism concerns the nature or constitution of *Sinne*, or as I shall say henceforth, concepts. True, concepts are not individuated by the subjective capacities of individual speakers or thinkers. Rather, they are public entities, capable of being simultaneously entertained by several subjects. Putnam's concern, however, is not with the constitution of concepts, but precisely in what capacity subjects may be said to entertain, know, or grasp these concepts. And nothing in Frege's program appears to contradict the idea that successful such graspings on the part of an individual subject must consist in (or at least correspond to) a specific psychological state of that individual. Indeed, his theory seems very much to presuppose something along those lines. Otherwise, there would be no clear way in which his semantics could claim *any* tangible connection to the cognitive lives of individual subjects.³⁹

The argument by which Putnam hopes to show that no notion of meaning can jointly satisfy the two conditions presupposed by Fregean semantics is the famous Twin-Earth thought-experiment. Earth and Twin-Earth are exact physical duplicates in all but one respect: there is no water on Twin-Earth. Instead, what flows in its rivers and faucets is a substance whose chemical composition Putnam suggests that we abbreviate as XYZ. To say that Twin-Earth is an exact duplicate of Earth in all other respects is to say that there are exact duplicates of all the Earth's

³⁹ In "Sinning against Frege," p. 233/19, Burge takes issue with Putnam's attribution to Frege of the view that "grasping meaning (or sense) is just a matter of being in a psychological state." Certainly, Putnam's use of the word "just" in this connection is ill-advised, but he need hardly have introduced that emphasis to make his point: what is in question is whether it must *also* be a matter of being in a certain psychological state.

inhabitants on Twin-Earth as well.⁴⁰ For everything that happens on Earth, there will be a duplicate event on Twin-Earth.

What we want to know is how to understand the following two duplicate events: a subject on Earth utters “Water is good to drink” and a subject on Twin-Earth utters “Water is good to drink.” If an Earthling and a Twin-Earthling were to meet, they would each find it natural to translate the other’s word “water” as “water.” We may imagine such an encounter occurring in 1750, that is, before the chemical revolution and its counterpart on Twin Earth. In such a scenario, it is plausible to assume that nothing known to either of the interlocutors could reveal to them that these translations would be wrong. Nonetheless, the word has a different extension in the two settings and so means different things in the vocabularies of Earthlings and Twin-Earthlings. On Earth, the word “water” has as its extension H₂O, whereas the Twin-Earth word “water” has as its extension XYZ. It appears, though, that speaker on Earth and the speaker on Twin-Earth would be *exact psychological duplicates*, even though they are talking about – referring to – different kinds of things. Thus, Putnam concludes, “the extension of the term ‘water’ (and, in fact, its ‘meaning’ in the intuitive preanalytical usage of that term) is *not* a function of the psychological state of the speaker by itself” (Putnam, “The Meaning of ‘Meaning’,” p. 224).

What must be added in order to determine extension is consideration of what Putnam calls the *contribution of the environment* (Putnam, “The Meaning of ‘Meaning’,” p. 271). It is the contribution of the environment that differs in the two cases, and so that must be what explains the difference in meaning. Hence, consideration of factors external to the psychological domain of the speaker is crucial in determining what the speaker is speaking about. In Putnam’s controversial slogan, “meanings just ain’t in the *head!*” (Putnam, “The Meaning of ‘Meaning’,” p. 227). If this is true, it shows that the two theses considered above cannot jointly be true: it must either be the case that

⁴⁰ Except, of course, in the sense that their bodies do not contain water, but rather XYZ.

psychological state does not determine intension, or that intension does not determine extension. Faced with this choice, Putnam suggests that we opt to retain the connection between intension and extension, and rather sever the connection between psychological state and intension.

We note, however, that in spite of the explicitly anti-Fregean thrust of Putnam's argument, there is still an important sense in which his approach remains Fregean. From our current perspective, Putnam's criticisms amount to the claim that the Fregean project, as we have presented it, remains valid, but that Frege's execution of this project is flawed. It is flawed insofar as it is still unable to liberate his theory from the idea that knowing the meaning of a term is a psychological state of the individual. Thus, Putnam appears to take over the Fregean division of labor between semantics and psychology, saying that the theory of meaning properly conceived – conceived in a manner fit to do for semantics what transformational linguistics has done for syntax – must focus on what he now calls “wide content” – a relationship between conceptual *Sinne* and their external *Bedeutungen*. This leaves psychology outside the domain of the theory of meaning, now sailing under the heading of “narrow content.” This is a laudably frank proposal for how to construe the status of Frege's rather obscure notion of a “subjective idea”: ideas, composing narrow content, are to be understood as solipsistic and in this sense Cartesian.⁴¹ A notable consequence of the distinction between narrow content and wide content is that semantic ascription comes apart from psychological attribution and action explanation. Putnam's proposal thus appears to extend the Fregean project in at least this respect: in its effort to save our prospects for a systematic semantics, it is willing to let the

⁴¹ We should note that not all narrow content theories need assume that narrow content is solipsistic and Cartesian in this way. But the more one relaxes this constraint, the harder it will become to say exactly what the distinctive function of the appeal to such narrow content is; why, that is, narrow content specifications and wide content specifications are not, after all, competing specifications of the same content. I take it that exploring such questions is part of what Colin McGinn is up to in “The Structure of Content.”

psychological realm float. Mind and language, for the purposes of philosophical theory, are two entirely distinct matters.

1.6 Burgean anti-individualism

This impasse provides an excellent vantage from which to survey Burge's contribution. On the one hand, Burge is deeply appreciative of several philosophical theories that are commonly seen as containing important criticisms of the Fregean project (in particular Putnam's, as we have seen, as well as Kripke's analysis of proper names) and his own arguments draw much of their force precisely from these. On the other hand, Burge is clearly also trying to consolidate the Fregean project in the face of these criticisms.⁴² In doing this, he provides an entirely new perspective on the relation between Fregean semantics and the philosophy of mind, and this is what makes his work uniquely important. What Frege's move, pushed one step further by Putnam, leaves us without is a sense of how and why understanding what a person is saying could tell us anything about what that person is thinking. In pursuing the Fregean track, we stand, perhaps, to gain the prospect of a systematization of semantics with a scope and power that we had hitherto not seen the likes of, but we appear simultaneously to lose any sense of why an application of this systematized theory on occasion should provide a guide to a person's mental life. Whatever our ultimate philosophical predilections might incline us towards, this is not, it should go without saying, a loss that we should accept lightheartedly. For it will amount to giving up not only on a long held philosophical desideratum. We would be giving up what is clearly a deep-seated dictum of common sense.

⁴² It belongs to the story here that Burge thinks Kripke's and Putnam's arguments are, in an important sense, wrong. Not, that is, wrong as such, but wrong as criticisms of Frege. See Burge, Introduction to *Truth, Thought, and Reason*, pp. 40-52; "Sinning against Frege."

It appears, then, that we are left with a dilemma – either opt for a systematization of semantics which leaves us without a connection between semantics and the philosophy of mind, or preserve that connection but abandon the hope of a systematized normative semantics. It is instructive to think of Burge’s program as an attempt to show that this is not a true dilemma. According to Burge, what *would* jeopardize our hopes of a systematic normative semantics is not just any essential connection with the psychological domain *per se*, but the particular way in which, when such a connection is drawn, we are prone to think of the psychological domain, viz. in an *individualistic* manner. Moreover, once we see that what should bother us is not simply ‘psychologism’ in some yet unspecified sense of that idea, but an individualistic psychologism, we will see that the fundamental resources required to connect the philosophy of language and the philosophy of mind are there already in Frege. The problem, in Frege as well as in Putnam, is a fundamentally solipsistic, and in this sense individualistic, construal of the mind.⁴³

Among the questions which will remain at the forefront throughout this dissertation is precisely whether this is a feasible project; i.e., whether any theory which retains this broadly Fregean structure can give a plausible portrayal of the contents of the minds of individual agents. It will be my argument that it cannot. In ch. 4, I will present a detailed argument toward the conclusion that Burge’s anti-individualist theory of thought-attributions will not reliably allow us to capture the rationality of the subject’s processes of reasoning, and therefore will not reliably allow us to explain that subject’s actions in the right kind of way.

It is striking that Burge attempts to consolidate the Fregean project by confronting Putnam’s argument at exactly the point where Putnam seemed most Fregean: namely the strong distinction

⁴³ Recall Dummett’s claim that on Frege’s view, the subjective is “essentially private and incommunicable” (Dummett, *Origins of Analytical Philosophy*, p. 23).

between subjective ideas and objective thoughts, or in Putnam's parlance, between narrow content and wide content. Putnam, as we saw, introduced the distinction between narrow content and wide content as a way of characterizing the distinction between psychology and semantics. Briefly, Putnam argues that we should think of semantic content as bearing essentially on an (external) referent, whereas psychological content should be construed merely in terms of intrinsic features of the subject's representations, with no essential reference to any worldly objects.

Burge is highly critical of this idea:

[Putnam] remarks in effect that the subject's *thoughts* remain constant between his actual and his counterfactual situation. [...] he explicates the difference between actual and counterfactual cases in terms of a difference in the extensions of terms, not a difference in those aspects of their meaning that play a role in the cognitive life of the subject. (Burge, "Individualism and the Mental," p. 108*n*)

In this, he continues, "much of what Putnam says about psychological states (and implies about mental states) has a distinctly individualistic ring" (*ibid.*).

What would be individualistic about Putnam's proposal is its consignment of the content of the subject's mental states to a solipsistic bubble, so that the content of these states stays the same, even as the environment changes in important ways: a psychological internalism to go with one's semantic externalism. This is precisely the kind of Cartesian throwback that Burge's argument is designed to target. There are, in fact, two interconnected aspects of Putnam's thinking about this which Burge calls to attention. One is the idea, as we have just seen, that the contents of the subject's mental states stay constant over relevant environmental changes. The other important aspect follows directly: that semantic ascriptions do follow these environmental changes even where psychological attributions do not. Thus, the meanings of our words and the contents of our thoughts can come apart in important ways, and so semantic ascription cannot in general be used as a reliable guide in psychological attribution.

Adopting helpful terminology from Akeel Bilgrami,⁴⁴ we might thus classify Putnam's position as a form of *bifurcationism* about content. A bifurcationist about content holds that the commonsensically unitary notion of content which flows through our thought and speech must be split into two, for scientific and philosophical purposes: psychological content and semantic content. By contrast, Burge insists on the need for a unified notion of content.

In light of this, it is easy to see how someone might be tempted to summarize Burge's critique of Putnam's individualistic psychology in the following way: Burge holds that we should be externalists also about the mind, and so that we should treat psychological content as wide content too.⁴⁵ Nonetheless, this would constitute a grave misunderstanding of Burge's point, for he does not accept the narrow content—wide content distinction in the first place. Nor is this refusal to work from the wide-content category merely a conceptual splitting of the hairs on Burge's part. Looking more closely at his rebuttal of this idea will reveal something very important about how he thinks about content.

Burge attacks each element of the narrow content—wide content distinction in several ways. For our purposes, the most important line of thought is the following: the idea of psychological content as narrow content is the idea of a genuinely representational content that is, in a relevant sense, environmentally neutral. But there is no such content: “[T]he idea that there is a constant element that shifts its referent with environmental ‘context’ does not account for the nature of the mental states or their representational behavior” (Burge, Introduction, *Foundations of Mind*, p. 12).⁴⁶ It

⁴⁴ See Bilgrami, *Belief & Meaning*.

⁴⁵ This seems to be Bilgrami's reading in *Belief & Meaning*.

⁴⁶ That said, there are occasions recently where Burge has seemed willing to jeopardize this extremely important insight. Thus, addressing concerns about the fate of self-knowledge under anti-individualism, Burge writes: “The authority of authoritative self-knowledge does not extend to the *res in de re* judgments. The relevant self-attributions in authoritative self-knowledge are to be seen as about the intentional content and the attitude-type of the attributed attitude. [...] *The intentional or representational content does not even include the referents of the conceptual elements. It includes only the ‘senses’ or modes of presentation, or conceptual or applicational elements of the thought*” (Burge, “Mental Agency in Authoritative Self-Knowledge: Reply to Kobes,” p. 424, my emphasis).

is easy to see how Putnam's work fits this mold: Putnam invites us to see the Earthling and the Twin-Earthling as exact psychological duplicates, the only difference between them being the meaning of their respective words 'water.'

Likewise, wide content too fails to be genuinely representational content. Speaking in terms of wide content would be a way of connecting a cognitive subject with an object thought about. But wide-content specifications fail to tell us anything about *the manner in which* the subject thinks of that object – what I will call his 'cognitive perspective;' e.g., thinking of the disease in one's thigh *as* arthritis. Taking no account of the subject's cognitive perspective, wide-content specifications fail to differentiate between thinking of one's arthritis as arthritis and thinking of it as tharthritus. Herein lies a problem:

Individuals necessarily think about subject matters in certain ways or from certain conceptual perspectives. [...] It will not do to take the referent as capturing any sort of psychological content. Representational content needs to explain or indicate how referents are thought about. (Burge, Introduction, *Foundations of Mind*, p. 12)

The conceptual perspectives that Burge here speaks about are, of course, Fregean *Sinne* or some suitably refined successor. Thus, in a nutshell, the problem with narrow content is that it is all *Sinn* and no *Bedeutung*, whereas the problem with wide content is that it is all *Bedeutung* and no *Sinn*. Neither can fill the theoretical and explanatory role that the notion of representational content – mental content – is supposed to fill.⁴⁷

Burge is quite clear, then, that we would be wrong to try to appropriate Frege's *Sinn*—*Bedeutung* distinction in terms of two different kinds of content, one to be sorted out under the

⁴⁷ Burge mentions specifically "semantical, psychological-explanatory, and epistemic" such roles (Introduction, *Foundations of Mind*, p. 13). Later in this dissertation we will investigate the claim of Burge's notion of content to fill all of these roles. Special emphasis will be placed on the semantical and the psychological-explanatory.

heading of psychology, the other to be sorted out under the heading of semantics. We need, and can have, only one kind of content, a kind of content which spans over both these elements. With this insight, we move closer to a full characterization of the notion of content that Burge's anti-individualism operates with. We are looking for a notion of content which is *unitary*, in the sense that it gives us psychological content and semantic content in a single breath; which is *externalistic*, in that it makes essential reference to the worldly objects that the content is about but which yet relates thinkers to such objects by way of the particular *representational aspects* of their thoughts. Finally, the notion of content in question must be *anti-individualistic* in its insistence that these mediating representations are not unique to each individual, or each occasion of thought. Rather, we should think of these representations as contained in *socially constituted concepts*. This is why the actual subject and the counterfactual subject in Burge's thought-experiment are not thinking the same thought. Thoughts are individuated by the concepts that figure in them. The two subjects are members of different linguistic communities, which individuate rheumatoid ailments differently. As such, the two subjects are thinking with different concepts, and thus thinking different thoughts. As we have seen, it is Burge's contention that the fact that one subject appears to have an incomplete mastery of the relevant concept in his community does not matter to the question of whether that is nonetheless the concept that he is thinking with. Burge thinks that common sense confirms his conviction here. Yet neither a narrow-content specification nor a wide-content specification will succeed in showcasing this difference. On either specification, it will seem that the contents of the two subjects' thoughts are the same. These are therefore shown to be importantly deficient as ways of specifying mental content. For, according to Burge, there is a relevant difference: even as they may be thinking of the same worldly object, they are thinking of that worldly object deploying different socially constituted concepts.

At some level of abstraction, I agree with all these desiderata. In what follows, however, we will find fault with various aspects of Burge's execution of this project. We will be able to trace these faults back to the particular manner in which Burge presents the social dimension of cognition, namely in terms of shared, socially constituted concepts. Moreover, as should be clear, this is no merely peripheral commitment in Burge's project. Our next task is to investigate this notion directly. We will do this in two operations. In chapter 2, we will inquire into Burge's warrant for claiming that the actual and the counterfactual subject must be thinking different thoughts because they are operating with different (socially constituted) concepts. This, recall, was the conclusion that was supposed to establish anti-individualism as such. In chapter 3, we will investigate the theoretical gambit most characteristic of Burge's anti-individualism, namely the idea that we can correctly portray people as thinking with concepts they do not completely master. In chapter 4, I will show that attribution of such incompletely understood concepts will tend to undermine our ability to use such attributions to guide us in assessments of the subject's processes of reasoning and to explanations of his actions. Finally, in chapter 5, I take on the question of normativity: I argue that Burge fails to establish the relevance and plausibility of the view that linguistic conventions encode norms that bear on the meaning of our words and the contents of our thoughts in such a way that, even where a subject fails to understand such a norm, the norm still applies.

2

CONCEPTS: THEIR ALLEGED ROLE IN COGNITION AND COMMUNICATION

2.1 Introductory

There is a longstanding tradition in philosophy, with roots stretching at least back to Kant's first *Critique*, according to which we can delineate the range of thoughts that a person can entertain by reference to the concepts at that person's cognitive disposal.¹ From within a certain frame of mind, this thesis may even seem to enjoy an air of self-evidence: thinking some thought is to conceive a possibility, and such conceiving is an activity we engage in by employing concepts. That for which we have no concepts, we cannot conceive, and so cannot think.

More recently, something like a corollary of this idea has taken hold in the philosophy of language, namely the idea that we can explain the possibility of communication by reference to *shared concepts*. The relation to the Kantian ancestor-thesis runs deeper than a first glance might suggest: for on this model, we are to understand communication in terms of two or more people *sharing a thought*. Thus, if one's ability to entertain a certain thought hinges on one's having the appropriate set of concepts, it would seem like the very possibility of communication – sharing thoughts – hinges on one's sharing concepts.

¹ I say 'person,' although in this context, 'creature' might be more felicitous: Kant thought that the conceptual abilities in question were in important ways species-universal.

Perhaps the most influential statement of such an idea can be found in the works of Gottlob Frege. Indeed, many of the awkward moves that Frege makes in order to secure a notion of *Sinn* which is public and timeless, independent of the cognitive activities of any particular thinker or set of thinkers, can be seen as directed precisely toward securing the possibility of sharing thoughts by way of sharing concepts.² Only such materials, it would seem, could serve to build the foundations of a philosophical theory of communication and the transfer of knowledge.³

It is rare to see the thesis that communication requires shared concepts spelled out in so many words today, let alone defended with anything like Frege's ferocity. Nonetheless, I think the thesis still wields considerable influence, perhaps more as an unstated background assumption on the basis of which research programs are framed and against which particular claims are tested.⁴

This chapter will present an argument against a certain reading of these twin ideas: the broadly Kantian idea that the ability to think a certain thought requires one's having the right concepts at one's disposal, and the broadly Fregean idea that if two people are to be able to communicate a certain thought, they must share (i.e., both have) the requisite concepts. I say I will target 'a certain reading' of these theses, since I will not rule out that there may be innocuous things

² E.g., Frege, "Thought," pp. 336-337. See section 1.4 for more detail.

³ The famous 'Bucephalus-passage' from "On *Sinn* and *Bedeutung*" nicely conveys the basic idea: "A painter, a horseman, and a zoologist will probably connect different ideas with the name 'Bucephalus.' This constitutes an essential distinction between the idea and the sign's sense, which may be the common property of many people, and so is not a part or a mode of the individual mind. For one can hardly deny that mankind has a common store of thoughts which is transmitted from one generation to another" (Frege, "On *Sinn* and *Bedeutung*," p. 154). Frege adds (p. 155) that two people "are not prevented from grasping the same sense; but they cannot have the same idea." Thus, Frege draws a strong distinction between subjective and incommunicable ideas, on the one hand, and objective and shared *Sinne* / concepts on the other. His language strongly suggests that if we did not assume that concepts are capable of being shared in this sense, we could not explain communication, the shareability of thought, and the transfer of knowledge.

⁴ Consider, for instance, the ease with which Jerry Fodor and Ernest Lepore think they can defeat a certain kind of holism: holists think that the content of a concept – and thereby its identity – is a matter of its inferential connections with other concepts. But no two people are ever likely to agree on all the inferential connections of a concept. Therefore, no two people are ever likely to have the same set of concepts. Reductio. (Implicit assumption: Communication requires shared concepts. Explaining the possibility of communication was part of the aim of the holistic theory. To the extent that it fails in doing this, holism undermines its own purpose.) For an influential statement of this program, see Fodor and Lepore, *Holism: A Shopper's Guide*, and Fodor, *Psychosemantics*, ch. 3. A more recent exchange on these matters can be found in Robert Brandom, "Inferentialism and Some of its Challenges" (especially section 5) and Fodor and Lepore, "Brandom Beleaguered" (especially section 3).

one might mean by these words – innocuous, but correspondingly vacuous. The reading I will argue against is one which tries to mine the theses for local explanatory potential:⁵ that we can say that a certain thought is within the cognitive reach of a given subject *because* this subject possesses the requisite concepts, and ditto for communication. Such explanatory schemes could only work assuming the following readings of the two theses:

[Concept-cognition thesis]: in order to be able to think a certain thought, an individual must already be in possession of the required concepts.

[Concept-communication thesis]: in order for two people to be able to properly communicate a certain thought, they must both already be in possession of the required concepts.

But as I will argue in this chapter, these claims are not true, and so are unavailable to lend support to the explanatory project in question. What renders them false (rather than merely vacuous) is the temporal marker ‘already’. This is perhaps most readily seen in the case of communication: it cannot be that communicating a certain thought requires that we *already* share concepts, since it is only *in and through* communication that we can get our concepts into proper alignment, if they diverged in the first place. If the concept-communication thesis were true, miscommunication would prove irreparable whenever it occurs. The falsity is less apparent, perhaps, in the case of the role of concepts in cognition. But the basic idea is in any case simple: it seems strange to suppose that in encountering, say, a novel natural phenomenon, we need already possess the concept of that phenomenon, or else never be able to get our thoughts properly fixed on it.

⁵ By stressing the locality of the explanatory effort, I mean to offer at least a partial exculpation of Kant’s transcendental invocation of the basic idea of the concept-dependency of thought. Since Kant thought that human cognitive capacities were necessarily ensconced in this set of concepts and not another, his project does not license any comparative judgments about the ranges of thought contents available to different (human) individuals with different conceptual repertoires. These are the kinds of judgments that will provide my target here.

Rather, it seems more plausible to say simply that it is the thing itself which presents us with the concept required to think about it.⁶

It remains to show that these theses are actually operative in current philosophy, and that I am not just directing my arguments against a straw-man. Much of this chapter will be dedicated to showing that such assumptions are operative in Burgean anti-individualism, and in particular in “Individualism and the Mental.” We will not, of course, find Burge defending these claims in so many words. Instead, my argument is that the main conclusions of that paper would be available only in light of such background assumptions. Accordingly, they are as yet unsupported by good argument. In particular, Burge proposes to draw strong conclusions regarding the comparative ranges of thoughts that two individuals can think, conclusions based on observations of the conceptual repertoires made available to each through their membership in different linguistic communities. The explanatory drift of the argument is that different languages afford different conceptual repertoires to their speakers, and so, that different sets of beliefs are within the cognitive range of each group. I will show that shorn of support from the concept-cognition thesis and the concept-communication thesis, these kinds of conclusions lose their edge, and collapse into variations on a much more familiar and mundane theme.⁷

⁶ If, indeed, we should still be inclined to discuss these matters in terms of concepts at all. Michael Williams, in a short commentary article, argues that holistic externalism “does not go well with taking concept-talk seriously” (Williams, “Bilgrami on Belief and Meaning: Is Fregean Externalism Possible?” p. 610). I am strongly inclined to agree, but I cannot pursue the matter here. The compatibility of Fregeanism and externalism is further discussed in, *inter alia*, Åsa Wikforss, “Content Externalism and Fregean Sense” and Jessica Brown, “Externalism and the Fregean Tradition.” Throughout most of this dissertation, I will continue to speak in terms of concepts, but all the while with a clear critical intent. The bottom line is this: what we fundamentally want is to find ways of speaking about the mental lives of people – what they believe, fear, hope for, and so on. Focusing on concepts has the bearings of a vital first step toward executing that task. But as I will show, many of the things which philosophers have said about concepts threaten to undermine precisely our ability to talk about people’s mental lives in the way that we seek to do.

⁷ The mundane theme corresponds to the innocuous-but-vacuous employment of the theses which I gestured at above: we might find that invoking concepts is a useful way of describing the cognitive situation of a person as it obtains at a given time, and similarly for the communicative situation between two persons at a given time. Thus, for instance, we might say that two people fail to properly communicate about a certain possibility because unbeknownst to them, they mean different things by the same word. But this lends no support to modally inflected claims about what these persons are in a wider sense *capable* of thinking and communicating about. It is only such stronger claims that could support

Here is a brief overview of the argument to follow. In section 2.2, I present Burge's anti-individualism and outline the argument which is designed to establish that thesis, the famous 'arthritis' thought-experiment. In section 2.3, I introduce two crucial specifications of what is involved in that thought-experiment: (i) in order to establish its conclusion – the truth of anti-individualism – Burge's argument must be invested with a certain kind of modal force: we must understand it not only as pertaining to what thoughts our subjects are or are not thinking on occasion, but what thoughts each subject is in a wider sense *capable* of thinking; (ii) the counterfactual subject's alleged inability to think any of the thoughts "commonly attributed with content clauses containing 'arthritis' in oblique occurrence"⁸ has no bearing on his ability to think *de re* thoughts *about* arthritis.⁹ Section 2.4 runs through some of the standard critical questions asked of Burge's argument, allowing us to properly position our own inquiry within that dialectic. Section 2.5 compares Burge's arthritis thought-experiment with Hilary Putnam's Twin-Earth thought-experiment. I argue that the difference in relative plausibility between the conclusions of the two thought-experiments is much greater than is usually acknowledged. The difference turns on the ability that each subject has in Burge's thought-experiment (but not in Putnam's) of forming *de re* attitudes about the relevant feature of their shared environment. Section 2.6 introduces a thought-experiment of my own: we will examine a snippet of conversation between a speaker of the actual language and a speaker of the counterfactual language. In spite of the fact that, by hypothesis, the counterfactual speaker does not have the concept *arthritis*, it will turn out that he and the actual speaker are perfectly capable of reaching reflective agreement concerning arthritis-related matters of

Burge's anti-individualist conclusion. I defend this claim in section 3 of this chapter, and explore its consequences further in section 6.

⁸ Burge, "Individualism and the Mental," p. 106.

⁹ A small specification: I will be talking about *de dicto* and *de re* not as two different kinds of *belief*, but rather as two different kinds of *belief-attribution*. For a statement of this approach to the distinction, see Daniel Dennett, "Beyond Belief," in particular, p. 176, 181.

fact in conversation. This undercuts Burge's conclusion that they are incapable of thinking the same kinds of thoughts. Section 2.7 rebuts possible objections from Burge and offers the following conclusion: whenever two people of roughly equal cognitive ability share a physical environment, there can be no philosophical warrant for saying that there is a relevant range of thoughts that one is capable of thinking but the other is not; or a range of concepts that one can have but the other cannot.

2.2 A new look at the 'arthritis' thought-experiment

Burge's "Individualism and the Mental" presents an attempt to demonstrate the truth of anti-individualism: the idea that there obtain "necessary or deep individuating relations" between the kinds of thoughts a person can think – or types of representational states he can token – and his natural and social environment.¹⁰ For our purposes here, we can take the anti-individualist claim regarding the relevance of a thinker's natural environment as demonstrated by Hilary Putnam's "The Meaning of 'Meaning'" (with important corrections and emendations by Burge¹¹). "Individualism and the Mental" seeks to demonstrate the correlative claim regarding the relevance of the individual's social environment, in particular, the relevance of the conventions governing his language.

¹⁰ The language of type-individuation of mental states is actually not prevalent in "Individualism and the Mental," and the quoted phrase is from the slightly later essay "Individualism and Psychology" (p. 221). Nonetheless, it has become the standard way of presenting the argument of the earlier paper, both in the critical literature and in Burge's own writings (see, for instance, the recent "Postscript to 'Individualism and the Mental,'" p. 151, "Other Bodies," p. 90, "Two Thought Experiments Reviewed," p. 184, "Cartesian Error and the Objectivity of Perception," p. 192, "Wherein is Language Social?" pp. 276-277, *Truth, Thought, and Reason: Essays on Frege*, Introduction, p. 57, and legions more).

¹¹ Which is not to say that we should, for *all* philosophical purposes, see the doctrine as thus established. I allow myself to do so here, partly because I am interested in exploring the relative strength of Putnam's and Burge's arguments (see section 5 of this chapter). Burge's crucial emendations to Putnam's argument for natural-kind externalism can be found in his "Other Bodies."

The main pivot of Burge's argument for social anti-individualism is a thought-experiment proceeding in three steps. I have already presented this thought-experiment in section 1.3, but will allow myself a bit of recapitulation here in order to facilitate exposition. In the first step of the thought-experiment, we are to imagine a speaker of English, generally cognizant of the ins and outs of his language, who utters the sentence "I have arthritis in my thigh." We are to ask, which thought is this person giving expression to? The thought-experiment is designed to elicit from us the following knee-jerk reaction: it could not be a thought involving our communal concept *arthritis*, since that concept just means 'rheumatoid ailment of the joints.' However, this knee-jerk reaction is supposed to crumble in the face of deeper reflection on the case. In fact, our initial inclination showcases a peculiar blind spot in our thinking about these matters, which it is crucial for Burge's argument that we isolate and contain. This is the idea that *we can only think with concepts we completely understand*.¹² According to Burge, reflection on the finitude of our cognitive powers suggests that, quite probably, very few of us fully master in this sense any significant range of those concepts which nonetheless appear to be at our cognitive beck and call. If our thinking were constrained to operate with concepts which we completely understand, then little thinking would ever take place. In particular, the idea that we can learn from each other and that, in Frege's phrase, "mankind has a common store of thoughts,"¹³ appears to require that we be able to say, for instance, that a physicist and I share the concept *electron*, even though my grasp of the concept is importantly deficient.¹⁴ If this is true, then the main obstacle to our asserting that it is indeed our concept *arthritis* that the subject of our thought-experiment is thinking with would seem to simply fall away.

¹² See, in particular, "Individualism and the Mental," p. 131. To forestall confusion, I ask the reader to note that where I speak of 'concepts', "Individualism and the Mental" frequently speaks of 'notions' (defined in "Individualism and the Mental," p. 102). Burge later switches back to the more widespread nomenclature.

¹³ Frege, "On *Sinn* and *Bedeutung*," p. 154.

¹⁴ See, for instance, "Individualism and the Mental," pp. 123, 125-126, 149.

In Step 2 of the thought-experiment we are to imagine an individual who is identical to the subject of Step 1 in all respects but one: this individual is a member of a speech community which classifies rheumatoid ailments differently than does English. In this counterfactual language community, the word ‘arthritis’ means simply ‘rheumatosis’. Thus, when this individual utters the syntactic string ‘I have arthritis in my thigh,’ he is, unlike his counterpart in our community, saying something which may or may not be true. In Step 3 of the thought-experiment we are to ask how, more specifically, we should understand what this person is saying, what thought he is expressing with this phrase. Burge’s conclusion is the following: however we are to understand him, we are not to understand him as saying or thinking anything that would involve our concept *arthritis*. The reason is simple: he does not have this concept.¹⁵ Burge writes:

In the counterfactual situation, the patient lacks some – probably *all* – of the attitudes commonly [i.e., actually] attributed with content clauses containing ‘arthritis’ in oblique occurrence. He lacks the occurrent thoughts or beliefs that he has arthritis in the thigh, that he has had arthritis for years, that stiffening joints and various sorts of aches are symptoms of arthritis, that his father had arthritis, and so on. (Burge, “Individualism and the Mental,” p. 106)

Understanding and assessing the validity of this conclusion is the immediate aim of this chapter. If the conclusion is warranted, it provides what Burge needs in order to establish the truth of anti-individualism. The thought-experiment is designed so that only the differences in the social environment could explain why the two individuals entertain different thought contents. This would be sufficient to establish the claim that there is a “necessary or deep individuating relation” between the kinds of thoughts an individual can token and his social environment, thus establishing Burge’s social anti-individualism.

¹⁵ In short, since ‘arthritis’ in our language and ‘arthritis’ in the counterfactual language do not even denote the same range of objects, they cannot express the same concept. See Burge, “Individualism and the Mental,” p. 106.

My claim here is that this conclusion is unwarranted, because it implicitly relies on the two theses discussed above, the concept-cognition thesis (according to which the capacity to think a certain thought requires that one already possess the relevant concepts) and the concept-communication thesis (according to which the capacity of two thinkers to communicate a certain thought requires that they both already possess the relevant concepts). However, some care is needed at this point. As I said above, there might be an innocuous reading of the conclusion which does not support itself on these theses. But on this reading the conclusion would not have the philosophical consequences it is ordinarily taken to have. As I will explain in section 3, Burge's conclusion requires that his argument be invested with a certain kind of modal force – that the counterfactual subject not just *is not*, but *could not be* thinking these thoughts, given how things stand with his conceptual repertoire. This is the reading that cannot be sustained. But without this reading implicitly assumed, Burge's conclusion is either unjustified or trivial, and thus without consequence for how we should otherwise describe the cognitive life of the counterfactual subject. We will see this clearly in section 6 of this chapter, when we turn to describing the communicative possibilities that arise in an encounter between a speaker of English and the counterfactual subject.

2.3 Two vital specifications

Two exegetical details need to be sorted out before we can proceed. First, and to repeat, Burge argues that we should not attribute thought-contents to the counterfactual subject containing 'arthritis' in oblique occurrence. However, Burge's way of stating this conclusion leaves it somewhat unclear exactly what force we are to think of it as having. He prefaces the conclusion by saying "It is reasonable to suppose that" and the conclusion itself says merely that the counterfactual subject "lacks some – probably all" of the attitudes in question. Until we know what these qualifications are

supposed to qualify, we cannot know what kind of consequences are supposed to follow from the thought-experiment.

In order to ensure that the conclusion may have any of the exciting philosophical consequences it is commonly taken to have, I will assume that these qualifications primarily bespeak epistemic uncertainties in determining exactly which concepts the subject may have at his disposal. Thus, Burge writes in support of his conclusion that “It is hard to see how the patient [in the counterfactual community] could have picked up the notion of arthritis” (Burge, “Individualism and the Mental,” p. 106). This statement hints at what I will call the *acquisition-problem of concepts*, the details of which will be further explored in sections 4 and 5 of this chapter.¹⁶

Suppose, however, that we could set these epistemic uncertainties aside and allow ourselves simply to stipulate that the patient does not have the concept *arthritis* (it is a thought-experiment, after all; I reckon such stipulations are open to us): on this assumption, I take it that Burge’s conclusion is supposed to follow with a certain kind of modal force – it is not just that the speaker *does not think* these thoughts, but rather that he *could not think* them, as things currently stand with his conceptual repertoire.¹⁷

This claim may appear controversial, so it is worthwhile stating with some care why this must be so. Why should the anti-individualist commit himself to saying anything about what each

¹⁶ For further invocations of this problem, see Burge, “Other Bodies,” pp. 93, 97; “Two Thought Experiments Reviewed,” pp. 188-189.

¹⁷ This should forestall any objection to the effect that my argument is merely seizing on a strategic mistake in Burge’s early works, a mistake which he has since admitted to making, at least in part, namely an overemphasis on linguistic matters (cf. “Postscript to ‘Individualism and the Mental,’” pp. 157-162, “Postscript to ‘Belief *de re*,’” pp. 65-66, “Phenomenality and Reference: Reply to Loar,” p. 449n3, “Reply to Christopher Gauker,” pp. 246, 250). A person’s membership in this linguistic community or that would be a hugely important epistemic factor in determining which concepts he has at his disposal, but is not the only such factor. Thus, it is conceivable that someone may possess a concept even though he could not have acquired it through the normal social routes. Accordingly, just considering a person’s public language does not allow us to determine *with certainty* whether he is capable of thinking a certain thought. These are welcome admissions, but they do not touch on the question at hand. For these are precisely the kind of epistemic uncertainties that I propose to set aside in the following. The real issue, I take it, is not how we can determine whether our subject has a certain concept, but rather what would follow from determining that he did not. It is on this latter question that Burge and I decisively part ways.

subject is *capable* of thinking, and not just stop at observing that they are not thinking the same thought on the occasion? Burge's own explicit pronouncements on the matter give us no clear answer as to which course he prefers. On the one hand, certain claims in "Individualism and the Mental" appear to favor the weaker reading, for instance:

I also do not claim that the fact that our subject lacks the relevant beliefs in the third step follows from the facts that I have described. The point is that it is plausible, and certainly possible, that he would lack those beliefs. (Burge, "Individualism and the Mental," p. 117)

On the other hand, a slew of later statements would seem clearly to favor the stronger reading. From the recent "Postscript to 'Individualism and the Mental,'" for instance, we cull the following claims:

The arguments center on the point that in the original situation an individual has one set of thoughts, and in the counterfactual situation the individual *cannot have those same thoughts*. (Burge, "Postscript to 'Individualism and the Mental,'" p. 156, my emphasis)

The fundamental reasoning [in "Individualism and the Mental"] concerns *conditions under which one can be in certain sorts of mental states, or have certain concepts*. The intuitions on which the thought experiments rely center on *conditions under which it is possible or impossible to have certain thoughts or perceptions*. (Burge, "Postscript to 'Individualism and the Mental,'" p. 162, my emphases)

The third stage [of the thought experiment] indicates that in such a situation it is *not possible for the individual to have thoughts about arthritis as such*. (Burge, "Postscript to 'Individualism and the Mental,'" p. 162, my emphasis)

Clearer still, here is an excerpt from a recent reply to an argument from Brian Loar:

The main case [for anti-individualism] is a set of thought-experiments that show that a given person can, under certain circumstances, have a given thought or attitude; but if certain environmental conditions were different or lacking, *a counterpart person could not, as a matter of metaphysical necessity, have*

that same thought or attitude. The point can be seen in terms of concept possession: Given certain background conditions, the individual on earth can have a concept *aluminum* or *arthritis* [...], and the relevant individual on Twin Earth cannot. (Burge, “Phenomenality and Reference: Reply to Loar, p. 438, first emphasis mine)

Even as Burge may seem to be in two minds on this matter, there is a simple reason why it is the stronger, modally inflected claims that the debate should centre on. The reason is not, strictly speaking, that it would be insufficient for anti-individualism to look simply at what each person *is* thinking, rather than what they are *capable* of thinking. Rather, the reason is that Burge’s argument is structured in such a way that the only observation which would allow us to establish with any degree of confidence that the counterfactual subject *is not* thinking a thought about arthritis as such is precisely that he *could not* be thinking any such thoughts. And the reason for that would be that he lacks the concept *arthritis*.

An informal reconstruction of the dialectic confirms this. Let us assume Burge opts for the weaker statement of the conclusion: the counterfactual subject is not thinking the same thought as the actual subject. We ask him for a reason to believe this claim. Burge responds: “It is hard to see how [he] could have picked up the notion of arthritis” (Burge, “Individualism and the Mental,” p. 106). But why is *this* a reason to believe that he is not thinking about “arthritis as such”? The reason must be that without the concept, he could not think the thought. The qualifications expressed in the weaker statements must be understood to pertain to what we are able to know about the subject’s conceptual repertoire, not to the relation between his conceptual repertoire and the range of thoughts he is capable of thinking. What the description offered in the thought-experiment makes “plausible” and “possible,” though not certain, is that the counterfactual subject does not have the

concept *arthritis*. What must be certain, if Burge is to have an argument at all, is that if he does not have the concept, he cannot have the thought.¹⁸

Exegetical matter #2: I assume that Burge's conclusion regarding the subject's ability to think thoughts "commonly attributed with content clauses containing 'arthritis' in oblique occurrence" has absolutely no bearing on the subject's ability to frame *de re* thoughts about the matter at hand, e.g., about the condition – arthritis – which his father suffered from for many years.¹⁹ This observation helps determine more precisely what is at stake in the argument. We can very well – indeed, we must – portray the counterfactual subject as thinking *about* his father's condition, which happens to be arthritis. What we cannot do, according to Burge, is portray him as thinking thoughts such as *the condition my father suffered from was arthritis* (here 'arthritis' occurs obliquely). This is so because, per hypothesis, the subject does not have the concept *arthritis*. But why, we may ask, should this fact have any bearing on whether it is appropriate to ascribe to him

¹⁸ On this hangs the fate of anti-individualism as a metaphysical claim about "conditions under which it is possible or impossible to have certain thoughts or perceptions" (Burge, "Postscript to 'Individualism and the Mental,'" p. 162). If the assumption that the subject does not possess the concept *arthritis* merely made probable ("plausible," "possible") the claim that he is not thinking arthritis-thoughts on a given occasion, no such conclusion would follow. All this suggests that we would do well to distinguish two stages in Burge's argument. One affirms a robust but ultimately contingent connection between the conceptual resources of a communal language and the conceptual resources of an individual member of that language community. The other affirms a necessary connection between an individual's conceptual repertoire and the range of his cognitive capability. Insufficient appreciation of this distinction is, I believe, to blame for much confusion regarding exactly what anti-individualism entails and so what would suffice to disprove it.

¹⁹ This claim might also appear controversial, in the following way: someone might argue that the line I take here would be appropriate if what we were considering were a natural kind concept. 'Arthritis,' however, is supposed to denote a social kind. Social kinds do not have individuation conditions at the level of purely physical description. There was no arthritis in the world (no contracts, no three-part fugues, etc.) before there was a social practice to classify it as such. Thus, my claim that the actual and the counterfactual subject can share *de re* belief about arthritis is false, for the *res* of the counterfactual subject's *de re* thought is not arthritis. Even as I have a bit of sympathy with the philosophical motivations behind this view, I think it leads us astray when applied to cases like these. Simple examples should serve for refutation: one might think back to the first time one heard a Bach fugue and had the thought "that is marvelous." Surely, this was a *de re* thought *about* the Bach fugue. It is true, of course, that the thought may not have contained any reference to its *res* as an instance of a particular kind, a fugue. But it would be absurd to make that a required feature of *de re* thought, since that would amount to making possession of a concept a prerequisite for being able to think *de re* about anything falling under that concept. This view can surely not be sustained, for the simple reason that it leaves us with no plausible account of how anyone can come to acquire the concept in the first place. In any case, we can set aside this matter here, since Burge himself has never denied the two subjects can share *de re* thoughts about arthritis. In fact, he affirms just that, for instance in "Two Thought Experiments Reviewed," pp. 184, 189. Moreover, long-standing features of Burge's own best thinking concerning the priority of *de re* belief over *de dicto* belief support the argument I just gave. See, in particular, "Belief *de re*," p. 59, "Postscript to 'Belief *de re*,'" p. 77.

thoughts containing ‘arthritis’ in oblique occurrence? Burge’s answer can be discerned in the following programmatic statement:

Clearly, oblique occurrences in mentalistic discourse have something to do with characterizing a person’s epistemic perspective – how things seem to him, or in an informal sense, how they are represented to him. (Burge, “Individualism and the Mental,” pp. 103-104)

The lesson is this: in attributing that-clauses, we are doing two things at once. We are indicating the thing – the *res* – thought about, *and* we are indicating something about *the manner in which* that thing is thought about.²⁰ Were we to attribute to the counterfactual subject a thought containing ‘arthritis’ in oblique occurrence, we would be picking out the right object. But we would be suggesting something false about the subject’s cognitive perspective on that object. The actual and the counterfactual subject can both be thinking about an instance of the same condition. But the actual subject is thinking about this condition, as it were, *sub specie arthritis*; the counterfactual subject is not – by stipulation, he has no such concept at his cognitive disposal. And since their constituent concepts differ, so must the thoughts, even as they are thoughts about the same thing. In what follows, I will be testing the relevance and plausibility of these observations.

2.4 Further observations on concept attribution

We can now move into the critical, argumentative section of this chapter. Our primary focus will be on questioning Burge’s warrant for drawing different conclusions in Steps 1 and 3 of the thought-experiment. In Step 3, recall, Burge argues that we should not attribute that-clauses containing ‘arthritis’ in oblique occurrence to the counterfactual subject, on the grounds that in so doing, we

²⁰ Cf. Burge, “Belief *de re*,” p. 46. In this way, such attributions connect essentially with action explanation, as Burge observes in “Other Bodies,” pp. 84, 92.

would fail to capture the subject's cognitive perspective on the matter at hand. But if this is so, we must ask why should we feel confident (in Step 1) that such attributions can capture the cognitive perspective of the actual thinker, when it has been admitted that this subject fails to fully master the concept in question (moreover, that he fails to do so in fairly systematic ways).

Burge will likely answer this question by relating a story about the role of our social environment – in particular, our linguistic community – in fitting us with concepts. We have stipulated that the actual subject is a member of a speech community – English – which has the concept *arthritis*, while the counterfactual subject is not. This, along with some minimal suppositions about each subject's basic cognitive capabilities, is what is supposed to entitle us to say that the one is thinking with the concept *arthritis*, while the other could not be. If we are left unpersuaded by this story, we are invited to give some thought to the following question: how could the counterfactual subject have acquired that concept? We may assume that we acquire our social-kind concepts by being inducted into a linguistic community. *Arthritis* is not among the concepts that our counterfactual subject could have acquired in this manner. This is what I referred to above as the acquisition-problem of concepts. Its force will be discussed in section 5 of this chapter.

Some of us might still demur: if accurate characterization of a speaker's cognitive perspective is a central concern of ours in propositional attitude attribution, then how can we rest satisfied with attributing to a speaker a concept he does not fully master? This seems paradoxical from the start. Should we not instead try to find some other concept, perhaps even a concept unique to the speaker if needs be, that better characterizes how the speaker approaches worldly states of affairs in his thinking?²¹

²¹ This question is examined in detail in ch. 4.

This, in outline form, is what Burge calls the *reinterpretation strategy*.²² In order to sustain his own favored *literalist strategy* – the idea that we should, *ceteris paribus*, understand a term as it figures in the attributed that-clause in accordance with its literal meaning in the communal language – Burge’s argument now shifts gear. The first thing he does is to claim that in pushing the re-interpretation strategy, we would in effect be begging the question against his argument. We would be assuming that

what a person thinks his words mean, how he takes them, fully determines what attitudes he can express in using them: the contents of his mental states and events are strictly limited to notions, however idiosyncratic, that he understands, a person cannot think with notions he incompletely understands. But supplemented with this assumption, the argument begs the question at issue. (Burge, “Individualism and the Mental,” p. 131)

Thus, in negating one of his major premises, we are not arguing against him, but simply begging the question against his position. Further discussion of this contention will be postponed until chs. 3 and 4.²³

Burge’s second objection to the reinterpretation strategy starts from an observation concerning common practice in concept attribution: basically, we do this all the time, so what could the problem be? That is, we regularly attribute thoughts using concepts in oblique position even where it is understood via his performance that the attributee does not fully master these concepts. The burden of proof rests squarely on the philosopher who wishes to argue that we should not take

²² There are several, substantially different ways of implementing a reinterpretation strategy, and a variety of philosophical justifications that can be offered for it. Burge deals with a sprinkling of each in “Individualism and the Mental,” pp. 121-132. I analyze Burge’s arguments in detail in ch. 3.

²³ We would do well to note a few things in passing, though. First, it is far from clear that we are, in fact, negating the premise in question: we are asking whether the attributed concept captures the subject’s cognitive perspective to the requisite degree of accuracy, and whether some other notion might not do a better job. There is nothing here amounting to a stipulation that a person can only think with concepts he completely understands. Second, even if we were negating his premise, I would still be baffled by the allegation of question-begging. We would be pointing out that the key premise in his argument is not, in fact, true. Thus, his argument may be formally valid, but still unsound. If this is begging the question, then much of the history of philosophy ought to be rewritten.

this practice at face value; i.e., as involving the attribution of the literally interpreted contents of the ascribed concepts.

This objection raises a series of interesting questions concerning, *inter alia*, whether this actually is an accurate description of standard practice in content attribution, and, if it is, why the philosopher should feel constrained by this fact, i.e., why we should think that ordinary practice is in the right in going about concept attributions in this manner.²⁴

In what follows, I will try a different tack: allow, for the sake of argument, that it is standard practice in concept attribution not to demand that the subject fully understands the concept, and that this practice is basically sound. Thus, our second critical question will not directly concern the warrant for drawing the conclusion of Step 1 of the thought-experiment (the claim that we can think with concepts we do not completely understand), but Burge's warrant for recommending *different* conclusions at Steps 1 and 3.

For a striking fact not commented on by Burge is that it also seems to be standard practice to attribute to foreigners that-clauses containing *our* concepts in oblique occurrence even when it is acknowledged that the foreigner may not have concepts that perfectly match ours. Indeed, the two cases seem to be exactly analogous in this respect. It should be obvious that an English speaker would not hesitate in attributing to the speaker of the counterfactual language the thought *My father suffered from arthritis for years*, even when it is understood that the range of objects denoted by his word 'arthritis' does not precisely match that denoted by our word 'arthritis,' and so that, by Burge's

²⁴ Signal contributions on these questions include Donald Davidson, "Knowing One's own Mind," p. 28n17, Akeel Bilgrami, *Belief & Meaning*, p. 78, and Gabriel Segal, *A Slim Book about Narrow Content*, pp. 121-126. I will have a few things to say about this in chs. 3 and 4.

standards, these words express different concepts. Yet Burge must maintain that common practice is wrong in doing this. What allows him to say this, and what is it that is gotten wrong?²⁵

At this point, it will not do for Burge to simply repeat what he has been saying all along, namely that in the intra-linguistic case the two subjects engaged in communication share the relevant concept, whereas in the inter-linguistic case, they do not. This will not do because we are at this point precisely asking why this fact – if it is a fact – should affect how we describe the pattern of communication flowing between the subjects in each case.

2.5 The acquisition problem

In order to mount our argument, it will be helpful to briefly consider a contrast with Hilary Putnam's famous Twin-Earth thought-experiment. In "The Meaning of 'Meaning,'" as we have seen, Putnam asks us to consider two distant planets, Earth and Twin Earth. Considered as isolated systems, these are exact replicas of each other in all but one respect: there is no water on Twin Earth. Instead, what flows from its faucets, what rains from its skies and so on is an altogether different chemical compound, XYZ. Putnam asks us to consider how we are to understand an

²⁵ Recall Burge's claim that the burden of proof rests on the philosopher who wishes to argue against the soundness of our common practice of thought-attributions. I should specify that what I am invoking here is not just the sort of hermeneutic nightmare that we experience when trying to render, say, Greek philosophical terms such as 'aition' or 'nous' into English (although such examples are also fascinating and important). Here it is typically understood that no English rendering can fully capture the meaning of the Greek term, yet we are constrained by circumstance to render them in *some* English term or other, and so opt for 'cause' and 'mind.' First, such Greek terms lie at the other end of the spectrum from vast majority of the cases we are considering here. A better example might be, say, a World Bank representative negotiating conditions for a loan to an African country, trying to get its politicians to agree that *Public servants should not take bribes*. Does the fact that they may not share our understanding of what a bribe is – that our concept might not be theirs – make it impossible for us to find an agreement about this? Certainly not, which is not to say that we should not be cognizant of the fact that by 'bribe,' they might be prone to mean something different than we are. But – and this is my second point – why should we assume that this is different from what goes on when we are trying to persuade an American politician of the same, when he is unsure of exactly what counts as a bribe? (Does dinner at a 3-star restaurant constitute a bribe?) Why is the latter case a matter of a shared concept, but one person's incomplete understanding of that concept, while in the former there are two different concepts in play? (Recall here that I am not targeting Burge's conclusion at step 3 outright, but rather asking what warrants us in drawing different conclusions at steps 1 and 3.)

Earthling and a Twin-Earthling when each utters, for instance, the phonetic string ‘water is good to drink.’ His point is that they are not talking about the same thing. The Earthling is talking about H₂O, while the Twin-Earthling is talking about XYZ. Now consider what would happen if we placed the two of them in a communicative scenario: it is overwhelmingly likely that each would report the other as assenting to the sentence (in the language of each) ‘water is good to drink.’ Moreover, it is not unlikely that they would never find evidential grounds for doubting their attributions. Nonetheless, according to Putnam, they would both be wrong: they are not thinking (and assenting to) the same thought any more than they would be if each were saying ‘I am six feet tall.’²⁶

A common way of approaching Putnam’s claim that they could not be thinking the same thought (note the modal inflection) is to focus on what I previously referred to as the acquisition-problem of concepts. The Twin-Earthling does not have the concept *water* (‘How could he? There’s no such stuff on Twin Earth ...’) and so could not form thoughts about water (and vice versa for the Earthling with respect to XYZ). Burge’s thought-experiment is modeled on Putnam’s, and it seems clear that the force of the acquisition-problem is supposed to be similar: the counterfactual speaker has no thoughts correctly ascribable using ‘arthritis’ in oblique occurrence, since he does not have that concept (‘How could he? There’s no such concept in his language ...’). In both cases, we are offered a distinctive kind of conclusion: the subjects we are considering are unable to think a certain kind of thought – *water*-thoughts in Putnam’s thought-experiment, *arthritis*-thoughts in Burge’s. Moreover, Putnam and Burge offer apparently similar arguments in favor of these

²⁶ I use this example for illustrative purposes, and do not mean to suggest that natural kind terms function as indexicals, as Putnam conjectured in “The Meaning of ‘Meaning’” (especially, pp. 229-235). For the essential corrective to this line of thought, see Burge, “Other Bodies,” pp. 87-91. In passing, we note a small complication: Putnam originally held that the thought-experiment has no bearing on the subjects’ thoughts, only on the meaning of their words. Simply put, the two subjects are thinking the same thought, but their words still have different meaning (which is why “meanings just ain’t in the head,” etc.). Partly due to pressure from Burge, Putnam has since admitted that this was a mistake, which is why I choose to pass over this wrinkle in this essay. See, for instance, Putnam’s Introduction to Andrew Pessin and Sanford Golberg, eds., *The Twin Earth Chronicles*. For more detail on the significance of this shift, see ch. 1.

conclusions: the subjects cannot think these thoughts because they do not have the required concepts; viz. *water*, *arthritis*. There is a difference, of course: in Putnam's case, it is a reference to the speaker's natural environment which underwrites the conclusion that the subject does not have the concept in question. In Burge's argument, this job is done by reference to the speaker's social environment; specifically, to the conceptual resources provided to him by his membership in a linguistic community. But this is just the difference between Putnam's natural kind externalism and Burge's more encompassing and ambitious social externalism (anti-individualism proper). The force of the argument taking us to the respective conclusions is supposed to be the same.

A little reflection reveals quite clearly, however, that the force of the two arguments is very different. The difference turns on the subjects' ability to form *de re* thoughts about the relevant *res* in the two cases. In fact, it should be clear that it is only on the surface that Putnam's argument turns on suppositions regarding the concept-dependency of thought at all. For the real reason why the Twin-Earthers cannot form thoughts about water is not as such that they lack the concept *water*, but rather that there is no water on Twin Earth to serve as the *res* of their *de re* thoughts. The acquisition-problem has the particularly stultifying effect that it has in Putnam's example, because short of embracing some form of concept nativism, we cannot see how the subject could have acquired the concept *water* when neither he nor any of his compatriots have had any causal interaction with the substance.²⁷

I rehearse these points simply to get the contrast with Burge's argument properly into view. For while the difference between the two arguments should be apparent, the way that this difference

²⁷ Burge notes ("Other Bodies," pp. 96-98) that it is "logically possible" that Twin Earth scientists might develop the concept *water* without ever coming into causal contact with the substance, if, for instance, they have the concepts *hydrogen* and *oxygen*. This might seem to form another parallel with Burge's thought-experiment, in which (as we noted in section 3 above) we might come to allow at least in theory that an individual can be in possession of some social kind concept even though he cannot have acquired that concept in the usual way. But this is no objection to my argument, which is that the Burgean subject's causal interactions with instances of arthritis makes it not just "logically possible" (as in the Twin Earth case) but *overwhelmingly likely* that he is capable of thinking thoughts ascribable using 'arthritis' in oblique occurrence, and this quite irrespective of whether he "has the concept."

affects the force and plausibility of each argument has yet to be sufficiently appreciated. Burge's thought-experiment is designed precisely so as to have the two subjects share physical environment. The point of doing this is to rule out appeals to that physical environment as the difference-maker when it comes to explaining the apparent fact that the two subjects entertain different thought-contents. This leaves only differences in the two subjects' social environment – specifically, differences in the conceptual resources of their respective languages – to account for the difference, thus apparently demonstrating the truth of Burge's particular form of social externalism.

The explanatory order proposed by Burge will thus look something like this:

Conceptual resources of communal language
 → conceptual resources of individual
 → range of thoughts available to the individual

Accordingly, Burge's invocation of the acquisition-problem of concepts takes the following form: 'How could our subject be thinking thoughts attributable with 'arthritis' in oblique occurrence? He does not have the concept. How could he have the concept? There's no such concept in his language.'

This version of the acquisition-problem, I claim, cannot be made both plausible and interesting. It could be made to do the philosophical work required of it only if Burge were entitled to the following assumptions: a person cannot form *p*-thoughts unless he already possesses the concept *p*; two people cannot communicate *p*-thoughts unless they both possess the concept *p*. But as I will show in the next section, these assumptions are false, and without them, Burge's proposed conclusion to Step 3 of the thought-experiment – the conclusion that the counterfactual subject is not thinking arthritis-thoughts – does not come through.

Briefly, then, the reason the acquisition-problem has none of the force in Burge's argument that it has in Putnam's is precisely that in the former, the two subjects share a physical environment.

On the supposition that the two speakers otherwise are of roughly equal cognitive ability,²⁸ this means that they are each perfectly capable of forming and communicating *de re* beliefs about the phenomenon at hand, as I argued in section 3 of this chapter. This is what eludes us in Putnam's thought-experiment. The impossibility of the Earthling and the Twin-Earthling triangulating on the relevant worldly object, and not as such any difference in their respective conceptual repertoires, is what accounts for the impossibility that the two could share such thoughts as 'water is good to drink.'

2.6 Interlingual communication: a counter-thought-experiment

What force, then, can Burge's conclusion have – the conclusion that a certain range of thought contents must remain unavailable to the counterfactual subject, because a certain range of concepts remain thus unavailable to him? In my opinion, very little. Burge's thesis entails that miscommunication between the actual subject and the counterfactual subject must remain endemic until such time as one somehow acquires the concepts of the other. This, to my view, is a gross mischaracterization of the situation. To see this, we need only consider a small sample of communicative interaction between the two.²⁹

It turns out to be quite hard to imagine a conversation between a speaker of English and a speaker of English*, in which the only difference from English is a slight variation in the extension of a single word. Luckily, we have a remedy at our hands. My native language Norwegian

²⁸ Burge explicitly supposes this and more.

²⁹ Someone might object to this strategy, perhaps on the grounds that if we are concerned with determining the range of thoughts a subject is capable of thinking, then placing him in a communicative context with another is, as it were, to tamper with the evidence. This is a misguided objection. If we are sincerely concerned with such matters as the transfer of knowledge and, more generally, with the social dimension of cognition, then communicative contexts are precisely where we should start our inquiry.

individuates rheumatoid ailments pretty much as Burge's counterfactual language community does. That is, Norwegian has a single word 'gikt' which covers all rheumatoid ailments. Accordingly, 'gikt' naturally translates as 'rheumatosis,' not 'arthritis'. (Norwegian has no etymological cousin of 'arthritis' in its lexicon. If one wants to talk specifically about an occurrence of a rheumatoid ailment in the joints, one would use a compound term: 'leddgikt' (= 'joint-rheumatosis'). I shall have more to say about this later.)

Consider now the following fragment of a conversation between a Norwegian speaker **N** and an English speaker **E**. For ease of exposition, I introduce myself as a bilingual translator between the two.

(1) **N**: 'Jeg har gikt i låret.'

I think everybody will agree that in conveying to **E** what **N** just said, I would be wrong to say (1a) "He says 'I have arthritis in my thigh,'" but right to say (1b) "He says 'I have rheumatosis in my thigh.'" What is less obvious, however, is why this is so. We should be on guard against prejudging the situation so as to assume that the reason (1a) is not the correct rendering is that **N** does not have the concept *arthritis*. This is a tempting line to take at this point, but another explanation also offers itself: (1b) is a superior rendering to (1a), because (1a) portrays the speaker as believing a blatant falsehood, when in fact he does not.

The conversation goes on:

(2) **N**: 'Faren min led av gikt i årevis.'

Since 'gikt' could denote any occurrence of rheumatosis, whether in joint or muscle, it should be clear that considerations of truth preservation alone could not force a choice between the following two renderings:

(2a) **Me:** ‘My father suffered from arthritis for years.’

(2b) **Me:** ‘My father suffered from rheumatosis for years.’

So far, then, I, acting as translator, should take counsel elsewhere. Burge’s advice to look to the structure of the conceptual repertoire available to **N** in Norwegian suggests itself with some force.

Suppose, however, that I am an old friend of the family, and I happen to know that the form of rheumatosis from which **N**’s father suffered from for years was in fact arthritis. Here is where Burge must think that I am liable to go wrong. As a translator, I will certainly feel the temptation to opt for (2a), because of the surplus of information contained in it. According to Burge, however, were I to do this, I would crucially misrepresent **N**’s thought-contents, insofar as we have stipulated that **N** does not have the concept *arthritis*, and so could not be thinking any *arthritis*-thoughts. This, I think, is a claim we can only make limited sense of at this point. For quite irrespective of whether he has this concept, it cannot be doubted that **N** is capable of thinking about arthritis. Even if we supposed that Norwegian did not, in fact, have a widespread colloquial usage for ‘leddgikt’ (‘joint-rheumatosis’), it would still have the concept *rheumatosis*, the concept *joint*, and the other standard resources of any natural language. At this point, our motivation for saying that **N** could not think thoughts ascribable with ‘arthritis’ in oblique occurrence – and could not do so until such time as he acquired the concept by learning English or an equivalent language – should start to wane quite dramatically. As we have seen, **N** is as capable as **E** of forming *de re* thoughts about arthritis. He does not need the concept *arthritis* to take note of the fact that rheumatosis can occur in the joints, and that these occurrences may have etiological peculiarities not shared by occurrences of rheumatosis outside the joints, and so on. It is simply no longer clear what our motivation could be

for maintaining that there is some range of thought-contents that **N** could not token, because his language community does not afford him with the concept *arthritis*.³⁰

To see this, let us suppose I choose to follow Burge's advice and render (2) as (2a) "My father suffered from rheumatosis for years." Now **E** asks, with reference back to (1):

- (3) **E**: 'Was your father's rheumatosis lodged in the thigh as well?'
N: 'Nei, i leddene.' ('No, in the joints.')

E: 'OK, I see.'

At this point, it should be patently obvious that **N** and **E** are sharing a thought, viz. that **N**'s father suffered from arthritis for years. But notice that step 3 of the conversation would have been entirely superfluous had I chosen to translate **N**'s (2) as (2a) "My father suffered from arthritis for years" rather than (2b) "My father suffered from rheumatosis for years." Yet, according to Burge I would have been wrong to opt for (2a) over (2b), because in doing so I would have conveyed something false about the *way in which* **N** thinks about his father's arthritis. Nonetheless, mere seconds later we are right to describe **E** and **N** as sharing the thought *N's father suffered from arthritis for years*. What happens in the interim such that this is now the right way to describe **N**'s thought-contents? One thing seems certain: nothing happens that could readily be described as **N** coming to acquire a new concept, *arthritis*. As the example makes clear, all the relevant change in passing from (2) to (3)

³⁰ Could Burge say that the fact that Norwegian *does* have a colloquial usage for 'leddgikt' shows that we could not run the thought-experiment with Norwegian playing the role of the counterfactual counterpart of English? (Cf. "Individualism and the Mental," pp. 105, 106, 112.) To disarm this objection, we would only need to roll back the clock to a time when Norwegian did not yet have this coinage (presumably there was a time when it did not). Still, nothing would prevent the Norwegians from developing that coinage, and indeed, in the normal run of things, one would certainly expect that they would. They know about rheumatosis, and they know that it occurs in the joints. But the point is this: there is no sense in which the Norwegians are adding anything to their cognitive repertoire by adding 'joint-rheumatosis' to their conceptual repertoire.

happens on **E**'s side of the transaction: **E** obtains new information about **N**'s father, information he could already have had, had I translated (2) as (2a) rather than (2b).³¹

And so we seem to have a counterexample to Burge's argument. Like the counterfactual subject in Burge's thought-experiment, **N** does not have the concept *arthritis*. Yet in opposition to the conclusion Burge offers about the counterfactual subject, **N** proves himself quite capable not only of thinking thoughts *about* arthritis (*de re*), but also of thinking thoughts "commonly attributed with content clauses containing 'arthritis' in oblique occurrence" (*de dicto*). No attempt to capture the substance of the communication that flows between **N** and **E** can plausibly deny this. The crucial point here is that **N** comes to exercise this capability without in the meantime acquiring the concept *arthritis*. This shows the concept-cognition thesis and the concept-communication thesis to be false: we cannot argue directly from observations concerning the range of a person's conceptual repertoire to conclusions about the range of thoughts he is able to think and communicate. Yet shorn of support from these theses, Burge's argument for anti-individualism collapses.

2.7 Conventional concepts do not explain communicative possibility

Faced with our example, Burge seems to have two options. He could

- (i) maintain that it would have been wrong to translate (2) using the concept *arthritis*, but join us in agreeing that it would be right to describe **E**'s and **N**'s agreement at (3) using that concept
- (ii) maintain that at no point during the conversation would it be right to describe **N**'s thought-contents by employing the concept *arthritis*

³¹ This rules out the following objection: what my argument shows is simply that **N** is *capable of acquiring* the concept *arthritis*, and so in this wider sense capable of thinking *arthritis*-thoughts. Presumably, Burge would never deny that. I agree, but this is not what is at stake in my argument, since **N** precisely proves himself capable of thinking these thoughts *without* acquiring the concept first.

Recall that we are trying to understand and assess the thesis that a thinker's conceptual repertoire constrains the range of thought-contents which can be correctly ascribed to that thinker. Now, both options agree that it would be wrong to render **N**'s (2) as "My father suffered from arthritis for years." The motivation for saying this is, in each case, the supposition that **N** does not have the concept *arthritis*, and so should not be ascribed thought-contents involving that concept.³² Option (i) then chooses to describe the situation that obtains after (3) as one in which it would be correct to ascribe to **N** thought-contents involving the concept *arthritis*. The problem with option (i) is easy to see: as we just noted, any difference between the situation that obtains after (2) and the situation that obtains after (3) falls squarely on **E**'s side of the transaction, and so cannot readily be described as **N** coming to acquire a new concept. Thus, the diagnosis that option (i) offers of the situation after (3) implicitly amounts to abandoning the motivation for its own diagnosis of the situation after (2).

It would appear that option (ii) would be the most natural way for Burge to continue, especially insofar as (ii) tries to hold on to precisely that motivation that option (i) implicitly abandons. At a first glance it might seem as though option (ii) spells philosophical disaster, insofar as it leaves us bereft of the means to say anything useful about the cognitive agreement evinced by **N** and **E** after (3). As it turns out, however, the situation may not be all that bleak, since we could still rely on the concept *rheumatosis* – presumably shared by **N** and **E** – to convey their agreement.

³² Thus, there is also a third option, according to which it was correct to render **N**'s thought-contents using 'arthritis' all along. It will not concern us here, for the simple reason that it involves abandoning the premise that we started out with, i.e., the assumption that **N** does not, in fact, have the concept *arthritis*, and that this fact rightly constrains the range of thought-contents we can ascribe to him. However, there might seem to be a way out for Burge here. Could he not point to the fact that the relation between the conceptual resources of the communal language and the conceptual resources of the individual is, after all, a contingent one, and so that what I have shown is that **N** must have had the concept *arthritis* after all, even though he could not have acquired it in the usual way? In my view, this would allow Burge to maintain the connection between the conceptual resources of the individual and the range of his cognitive capability only at the expense of depriving it of any philosophical significance (cf. my remarks in section 1): commenting on the range of a person's conceptual repertoire is now seen as just another way of describing the patterns we observe in his thinking, and can in no way contribute to an explanation of these. Moreover, it must be noted that this strategy would deprive the social of any distinctive role to play in Burge's anti-individualism, over and above the minimal role that it plays in Putnam's thought-experiment, through the division of linguistic labor. (Cf. Putnam, "The Meaning of 'Meaning,'" pp. 227-229.)

Closer inspection reveals, however, that this affords only limited consolation. For it is clear that the information packed into the communicative agreement between **N** and **E** after (3), i.e., the beliefs that **N** and **E** have each invested in that situation, substantially outstrip anything we can convey by using the concept *rheumatosis*. Thus, I, in my capacity of translator, am constrained by Burge's counsel to substantially *underdescribe* the cognitive situation between **N** and **E**. Moreover, I am constrained to do this precisely in order to convey what is, according to Burge, essential information regarding the way in which each thinker cognitively approaches the fact believed between them. We can now see that this counsel lands us in an absurd philosophical predicament, in at least two ways: first, as we have just seen, it fails to fully capture the informational state between **N** and **E**; second, it fails, even by Burge's lights, to capture **E**'s cognitive approach to this information, since he presumably *is* thinking with the concept *arthritis*.³³

Burge is still free to maintain his previous conviction, of course: **N** simply does not have the concept *arthritis* ('How could he?', etc.), and so at no point would we be right to ascribe to him thoughts containing 'arthritis' in oblique occurrence. But at this point I think we can see that this conviction is supported by nothing but a philosophical dogma. First, whether or not the Norwegian language provides **N** with the precise concept *arthritis*, it is clear that **N** is, given his other conceptual resources, able to think thoughts about arthritis. The acquisition-problem simply has no force here: it is, as it were, his father's condition which allows him to think these thoughts.

A second problem with maintaining Burge's program runs deeper, and so is harder to get into proper focus. It is also, I think, correspondingly important. One of the major motivations for maintaining the basic thesis concerning shared concepts is the illumination it appears to cast over a

³³ Could Burge at this point claim that we can say that there is a fact believed and agreed to between **E** and **N**, a fact which **E** approaches with the concept *arthritis* and **N** approaches with the concept *rheumatosis*? there would be no need to portray their capacity to agree in thought in terms of both employing the same concept. This provides no comfort for Burge either, since it is, in fact, just a different way of stating my general point. Taking this route amounts to depriving concept-talk of any essential purchase on how we describe the cognitive and communicative possibilities of people.

distinctive and important facet of human cognition, namely its intersubjective dimension. That is, appealing to shared concepts has the promise of shedding light on the conditions of possibility of what is, as Burge convincingly argues, quite essential to maintaining anything like full-blooded human cognition: learning from others in communication.

These ideas are worth stating with some care. The picture of linguistic communication that we are invited to consider is one that goes far beyond its use in merely conveying to each other what is already on our minds. Rather, we are striving to articulate a sense in which partaking in linguistic communication can greatly expand our cognitive horizons, can help us shape our practical and intellectual commitments, and can help us deepen our understanding of what it means to have these commitments, what responsibilities we thereby take on. It is easy to see how reflection on, on the one hand, the role of concepts in cognition and communication, and, on the other, the social constitution of those very concepts, could constitute a plausible starting point in this endeavor.³⁴

I think Burge's reflections along just these lines are immensely rewarding, and constitute a genuine advance in our understanding of "the role of social institutions in shaping the individual and the content of his thought" (Burge, "Individualism and the Mental," p. 100). What my argument suggests is that, at least after a certain point, the appeal to such socially constituted concepts becomes not a help but a hindrance to that task. The reason should be evident: the kinds of concepts around which Burge structures his argument for social externalism can only represent time-slice reifications of what is essentially a fluid and dynamic process of social interaction. Such reification deprives Burge's conclusion of any real plausibility, the conclusion that there is a philosophically consequential sense in which the actual and the counterfactual subject are barred

³⁴ Plausible, but certainly not obligatory. We note, for instance, how talk of shared concepts is conspicuously absent from Burge's more recent writings on the topic of communication and the epistemology of testimony. See his landmark article "Content Preservation" as well as "Comprehension and Interpretation."

from thinking the same thoughts because they are socially fitted with different concepts. Whatever there is to such a claim, it will likely dissipate as soon as they start talking to each other.

What drives this fluid and dynamic process of social interaction by which our concepts are formed is, in the first instance, communication. Thus, if our theory of concepts restricts the range of thoughts we can rightly ascribe to a person entering into and coming out of communicative contexts, we have a problem. What my argument shows is that Burge's particular way of locating the social dimension of cognition – in terms of communally constituted concepts underwriting communication – has precisely this consequence, and so undermines its own purpose.³⁵

By way of conclusion, I hazard the following diagnosis: we may, if we are still so inclined, continue to speak in terms of concepts in discussing these matters.³⁶ But the limited scope of such language must be made explicit – no real illumination is in the offing. We can comment on the range of thoughts a person *is* thinking by commenting on the range of his conceptual repertoire. But in so doing, we are really just providing redescriptions of what we already know. We are in no sense providing explanations. Likewise, we can comment on the quality of communication between two people by saying something about the range of their conceptual overlap. But if we do this, we cannot forget the fact that when things go well, it is *in* communication that people sort out their concepts (which is, for practical purposes, inseparable from reaching the right kind of substantive agreement about the matter at hand). Therefore, sharing concepts could never be rightly counted as a *presupposition* of communication. More generally, whenever two people of approximately equal cognitive powers share a physical environment (as they do in Burge's thought experiment, but not in Putnam's), there is never a philosophical warrant for saying that one can think a range of thoughts

³⁵ Here I draw inspiration from Donald Davidson's analysis, in "A Nice Derangement of Epitaphs," of the role of conventions in explaining communication. See also Bjørn Ramberg's excellent commentary in his *Donald Davidson's Philosophy of Language*, ch. 8.

³⁶ Indeed, for ease of exposition, I have allowed myself to do so in much of the preceding.

that the other cannot; the one can have a set of concepts that the other cannot. I think this follows directly from an appreciation of the role that the objects in our shared environment – the *res* of our shared *de re* thoughts – play in anchoring our cognition and communication.³⁷

Burge pursues a fundamentally correct idea: being in communion with others expands our cognitive horizons, allowing us to grasp thoughts and envisage possibilities that would otherwise have been quite inaccessible to us. But his pursuit of this idea is frustrated by the misguided assumption underlying it: the idea that only a philosophically substantial appeal to shared concepts can show how such communication is possible, and consequently, that the way in which we describe what is going on in communication must be constrained by our assumptions about which concepts the communicants share. I favor something like the following view: if there is any philosophical connection to be drawn between concepts and cognition, it must be that the expansion of our conceptual repertoires and the expansion of our cognitive horizons occur together, and that both occur, typically, in communication.³⁸

Addendum

Toward the end of the final section, I made a substantial appeal to ideas that lie at the heart of Burge's project, ideas that concern the social dimension of mindedness and which describe that expansion of our cognitive horizons which is made possible by our being participants in a certain kind of communicative practice. These are ideas that I have tried to articulate and preserve in my

³⁷ This, to my view, is the often overlooked lesson to be drawn from Davidson's "On the Very Idea of a Conceptual Scheme." See, especially, the conclusion at p. 198: "In giving up the dualism of scheme and world, we do not give up the world, but re-establish unmediated touch with the familiar objects whose antics make our sentences and opinions true or false."

³⁸ I mean this to be evocative of Gadamer's position in *Truth and Method*, especially his doctrine of understanding as a matter of *Horizontverschmelzung*, a 'fusion of horizons.' See Gadamer, *Truth and Method*, pp. 300-307, especially p. 306.

own argument. Thus, although I have stated my argument in my own terms, it is worthwhile to note that a partial motivation for such an argument can be developed independently from ideas that are central to Burge's own project. Indeed, part of what my argument can be taken to show is precisely how uneasily the conclusion of "Individualism and the Mental" sits alongside several longstanding tenets of Burge's program. One such central concern of Burge's is the cognitive fundamentality of *de re* belief.³⁹ The irreducible fundamentality of *de re* belief is crucial, for instance, to any plausible account of language acquisition.⁴⁰ This is a point which I have exploited in my argument against Burge's position in "Individualism and the Mental." My argument, recall, issued in the claim that whenever two people of more or less equal cognitive ability share a physical environment, there can be no philosophical grounds for saying that the one is capable of having a set of thoughts that the other cannot (and similarly for concepts, to the extent that this is a relevant concern). I see this claim as a testament to the cognitive fundamentality of *de re* attitudes. In spite of their initially diverging conceptual repertoires, **N** and **E** are capable of anchoring their communication in shared objects. This allows them to quickly come to share thoughts about arthritis, in all senses of 'sharing thoughts' which we should find worth caring about.⁴¹

In this addendum I will concentrate on another central theme in Burge's project, a theme which develops out of a reading of Frege. It first surfaces in the aforementioned "Belief *de re*," is

³⁹ See, for instance, Burge, "Belief *de re*," p. 44.

⁴⁰ See Burge, "Belief *de re*," p. 51, and "Postscript to 'Belief *de re*,'" p. 77. In the latter discussion he points out that his argument is essentially continuous with that made by Quine in *Word & Object*.

⁴¹ Moreover, my point resonates well with Burge's claim in the recent Postscript to "Individualism and the Mental" that "[a]t the time, I regarded the physical environment as more fundamental than the social environment in determining the natures of mental states. It is more fundamental psychologically, ontogenetically, and phylogenetically. I focused first on the social environment because I thought that its role was less close to the surface, less easily recognized" (Burge, "Postscript to 'Individualism and the Mental,'" p. 153). (Incidentally, Burge gives us no reason to think that he has subsequently changed his mind on this matter.) Chapter 2 could be partly read as making the claim that this insight is not well preserved by the actual argument of "Individualism and the Mental." A similar point surfaces also in Burge's article "Wherein is Language Social?": "Shared idiolectal meanings and shared concepts derive from a shared empirical world and shared cognitive goals and procedures in coming to know about that world" (Burge, "Wherein is Language Social?" p. 290).

developed in several of Burge's writings on Frege,⁴² and comes to full fruition in "Intellectual Norms and the Foundations of Mind" (1986).

Burge's argument in this article starts by warning against a common misunderstanding of Frege: that his notion of *Sinn* is intended to capture anything like conventional linguistic meaning.⁴³ In fact, says Burge, the notion of *Sinn* serves several distinct theoretical functions in Frege's work:

- (i) senses account for 'cognitive value.' "Senses are 'modes of presentation': ways things are presented to a thinker [...]. Not all modes of presentation are senses. But where modes of presentation are senses, they are associated with linguistic expressions."
- (ii) senses fix "the *Bedeutung*, the denotation or fundamental semantic value, of semantically relevant expressions."
- (iii) senses serve "as the denotation of expressions in oblique contexts" (Burge, "Frege on Sense and Linguistic Meaning," pp. 242-243).⁴⁴

These distinctions show that it would be overhasty to simply gloss Fregean senses in terms of conventional linguistic meaning. Nonetheless, as we move to dismiss this interpretation, we should take careful note of the underlying motivations which appear at first glance to speak in its favor. Focusing on the conventional meaning of terms seems like a promising way of satisfying the publicity constraint for senses – the fact that different thinkers can share a mode of presentation and

⁴² For instance "Sinning against Frege" and, especially, "Frege on Sense and Linguistic Meaning."

⁴³ Burge's most recent statement of this argument can be found in the Introduction to his *Truth, Thought, and Reason: Essays on Frege*, pp. 7-8, 35-40, 47, 54. See also "Sinning against Frege," *passim*, "Frege on Sense and Linguistic Meaning" *passim*. Burge namechecks Michael Dummett (*Frege: Philosophy of Language, Origins of Analytical Philosophy*) as a prominent Frege-scholar who ties Fregean *Sinn* to conventional meaning.

⁴⁴ This is a slight refinement of a list that occurs already in "Belief *de re*," p. 59.

thus be capable of thinking the same thoughts – without going Frege’s controversial Platonic route.⁴⁵ Burge admits that there certainly are unattractive aspects to Frege’s ontological Platonism. Nonetheless, he thinks that these unattractive aspects should not lead us to abandon the position just like that:

Frege’s rationalist point of view is unquestionably strange from a contemporary perspective. The taproot of the strangeness is, I think, the conception of fully determinate senses or thoughts that are completely independent for their representational properties of human practices or understanding. This conception has as a corollary the conception of self-evident thoughts that no one understands. Few have found these conceptions palatable. I believe that the point of view has some strengths that survive this unpalatability.⁴⁶ (Burge, “Frege on Extensions of Concepts, from 1884 to 1903,” p. 298)

While Burge is sensitive to the discomfort caused by maintaining a Platonist theory of concepts and thoughts, he is also quite adamant that any attempt to relieve oneself of this discomfort by tying senses to conventional linguistic meaning amounts to throwing out the baby with the bathwater. While tying Fregean senses to a specifiable communal practice appears at first glance like a fruitful compromise, promising both concreteness and supra-individual explanatory power, it will soon transpire that its explanatory power is just too local. We will be faced with the problem of explaining continuity of sense across communal boundaries and over temporal spans. The latter, in particular, is a central concern for Burge, as it was for Frege. Tying senses to conventional linguistic meaning deprives us of the means to account for scientific progress. Our ability to say, for instance, that we now have a fuller grasp of what a gene is than did Mendel, hinges our ability to say that Mendel was

⁴⁵ In particular, many commentators have expressed worries that Frege’s ontology leaves us with an impossible job in accounting for how individual thinkers can ‘grasp’ thoughts, if thoughts are to be seen as Platonic entities. Burge addresses such worries in the Introduction to his *Truth, Thought, and Reason: Essays on Frege*, pp. 29-31. See also “Frege on Knowing the Third Realm.”

⁴⁶ Further partial endorsements can be found in *Truth, Thought, and Reason: Essays on Frege*, Introduction, pp. 5, 27.

working with our concept *gene*, even as he had a demonstrably incomplete grasp of it.⁴⁷ (Although judging from the way that current scientific and philosophical discussions on the foundations of genetics go, one could certainly doubt whether there is such a thing as ‘our concept *gene*.’)⁴⁸ In light of such considerations, we clearly see the problem of tying Fregean sense too closely to communal practices: insofar as our conventions governing the expression ‘gene’ are different than those extant at Mendel’s time, our word ‘gene’ and Mendel’s word ‘gene’ would have to express different concepts, thus showing the attribution of our concept *gene* to Mendel to be false.

This line of thought is consolidated and generalized in Burge’s “Intellectual Norms and the Foundations of Mind.” The argument of that article is structured around the following idea: it must be possible to (rationally) doubt at least a certain range of necessary truths. The necessary truths in question are those that define a concept in terms of its empirical applications (e.g., “a contract is a legally binding agreement,” “arthritis is a rheumatoid inflammation of the joints”). Again, allowing this space for doubt is crucial to explaining scientific progress. This space for rational doubt is what we lose if we tie cognitive significance too closely to conventional meaning. For many path-breaking developments in science were occasioned precisely by scientists coming to question such putative necessary truths (e.g., is the atom really indivisible?). If the content of their concepts were

⁴⁷ For this, and a host of other examples from the history of science and mathematics, see Burge’s Introduction to his *Truth, Thought, and Reason: Essays on Frege*, p. 56. See also “Frege on Sense and Linguistic Meaning,” pp. 256-257, “Concepts, Definitions, and Meaning,” pp. 294-295. The Fregean chapter-and-verse on the matter, frequently cited by Burge, is *Foundations of Arithmetic*, p. vii. We note in passing how controversial is Burge’s stipulation that this is the only way, or even a coherent way, to account for scientific progress. For the classic case against this view, consult Thomas S. Kuhn, *The Structure of Scientific Revolutions*. Burge mentions Kuhn’s work briefly in “Intellectual Norms and the Foundations of Mind” (p. 270), only to say (without argument) that “I think that this sort of attribution [i.e., attribution of our concept *gene* to Mendel, our concept *atom* to Dalton, etc.] is defensible in a wide variety of central examples of scientific developments, even in the light of Kuhnian insights into scientific revolutions.”

⁴⁸ Cf. Kim Sterelny and Paul Griffiths, *Sex and Death: An Introduction to the Philosophy of Biology*, chs. 3-4, Paul Griffiths, “Molecular and Developmental Biology,” pp. 263-265. The same question can be obviously be asked of the biological concept *species*. If we ask, did Darwin think with the same concept *species* as we do, we must first ask, what is our concept *species*? Yet it is far from obvious that there is a single concept *species* in play in current scientific usage; in fact, it is quite clear that there is not.

fixed by the conventional wisdom of science at the time, we would have to portray such doubts either as plainly inconsistent or as really involving some other concept. We could not consistently see them as steps toward a more adequate grasp of what is involved in a given concept.⁴⁹

In the face of such considerations, Burge argues that we must decisively distinguish between what he now calls cognitive value, on the one hand, and conventional linguistic meaning, on the other.⁵⁰ In the course of outlining and defending this distinction, Burge says the following:

The dubitability of meaning-giving normative characterizations can be converted into a demonstration that social practices are not the only or ultimate nonindividualistic factor in individuating mental states and events. [...] even where social practices are deeply involved in individuating mental states, they are often not the final arbiter. This is because the sort of agreement that fixes a communal meaning and norms for understanding is itself, in principle, open to challenge.⁵¹ (Burge, “Intellectual Norms and the Foundations of Mind,” pp. 262-263)

Again, the reason why conventional linguistic meaning cannot be where the buck stops in connection with the individuation of mental states is that an individual can doubt the truth of such conventional conceptual explications without thereby necessarily jeopardizing his claim to master the concept in question. Since we can, and often do, portray an individual as *thinking with* a concept, but *actively doubting* its conventional, communally accepted definition, we must systematically distinguish between the cognitive value of a given term – the role it plays in the mental economy of a given individual at a given time – and its conventional meaning.

⁴⁹ Cf. my remark above (section 7) to the effect that communally constituted concepts can only mark out time-slice reifications of meaning. See also “Concepts, Definitions, and Meaning,” p. 298, where Burge motivates the distinction with considerations concerning “dynamic potential.”

⁵⁰ See, e.g., “Intellectual Norms and the Foundations of Mind,” p. 269. Recall that the first function of Fregean sense as outlined above was to capture cognitive value.

⁵¹ I should add that I find the qualifier “in principle” in the final sentence somewhat misleading. One usually says that something is possible ‘in principle’ if it is not really possible ‘in practice,’ or at least not under non-exceptional circumstances. By contrast, it seems to me a quite ordinary phenomenon that the conventional meanings of our words reconstitute themselves over time, often as a result of such challenges. Burge appears to restrict such challenges to the results of heterodox theorizing about things (e.g., “sofas are religious artifacts”). This strikes me as an undue restriction.

Burge makes some striking and powerful observations in connection with this distinction:

There was never any obvious reason why the two notions should coincide. They are responsible to different paradigms, and their explanatory purposes are distinct. Cognitive value is fitted to explicating possible differences in attitudes – cognitive perspective. Such differences are invoked to explain action and epistemic inquiry. [...] Attributions of continuity through changes of conventional or theoretical meaning provide another touchstone for cognitive value. Conventional meaning, by contrast, is fitted to describing reflective agreement on the means of conveying, in short order, accepted usage. (Burge, “Intellectual Norms and the Foundations of Mind,” p. 271)

And further to the point about the individuation of mental states made above:

Intentional mental states and events are individuated in terms of cognitive value. We have no other systematic, cognitively informative way of individuating them. (Burge, “Intellectual Norms and the Foundations of Mind,” p. 274)

Why should these claims surprise us? They are surprising because they are, if not in outright contradiction, then certainly in significant tension, with the central programmatic ideas of “Individualism and the Mental.” The correct and precise individuation of mental states and events is, we recall, the central concern also of that article. But the common perception of the take-home message of “Individualism and the Mental” is precisely that mental states and events are standardly individuated in terms of conventional linguistic meaning. Compared to this, the message of “Intellectual Norms and the Foundations of Mind” initially seems a bit more mixed. First we are told that social practices are “not the only or ultimate” factor in individuation, and that even where they are a factor “they are often not the final arbiter” (Burge, “Intellectual Norms and the Foundations of Mind,” pp. 262, 263). The article’s final word on the matter, however, seems rather more decisive: “Intentional mental states and events are individuated in terms of cognitive value. We have no other systematic, cognitively informative way of individuating them” (ibid., p. 274).

Compare this starker line with what we are supposed to learn from “Individualism and the Mental.” We are told that “obliquely occurring expressions in content clauses are a primary means of identifying a person’s intentional mental states or events.”⁵² So far, this should be unproblematic. But as we have seen, the characteristic move of “Individualism and the Mental” is contained in the claim that the pool from which are drawn the resources for such attribution is precisely the lexicon of our shared, public language, i.e., concepts defined by the linguistic conventions of our community. This suggests – at the very least – a shift in emphasis between the program of “Individualism and the Mental” and the strain in Burge’s thinking which is borne out in “Intellectual Norms.” In fact, my inclination is to say that there is much more than a mere difference of emphasis in play between the two articles, but nothing I shall say here rides on this stronger claim. For even on the weaker shift-of-emphasis construal, the programmatic statements of “Intellectual Norms” undermine that feature of “Individualism and the Mental” which we have been considering, namely its claim that the actual and the counterfactual subject are, in some philosophically consequential sense, thinking different thoughts *because* they belong to different language communities with different conceptual repertoires.⁵³ As we saw above, the thought-experiment is so designed that the only factor which could account for the supposed difference in mental contents between the actual and the counterfactual subject is the difference in the conventional meaning of the word ‘arthritis’ in the two language communities. With the new line concerning mental-state individuation taken in “Intellectual Norms,” this argument is seriously jeopardized. It might still be, of course, that the actual and the counterfactual subject are thinking relevantly different thoughts. But this conclusion cannot simply be supported by the observation that they belong to language communities with differing conventions governing the expression ‘arthritis.’ For what individuates mental contents is

⁵² Burge, “Individualism and the Mental,” pp. 103-104.

⁵³ For a particularly clear statement, see Burge, “Other Bodies,” p. 84: “By varying the *uses of words* in his linguistic community, I found that the content of his propositional attitudes varied.”

not ultimately linguistic conventions but cognitive value. If the actual and the counterfactual subject differ relevantly, it must be because the cognitive values of their respective words ‘arthritis’ differ.

Do they so differ? This is not an easy question to answer, and I shall have more to say about these matters in my next chapter. Burge’s exposition of the notion does not provide nearly enough detail to fix strict identity conditions for cognitive value.⁵⁴ While Burge is explicit that cognitive value is not a purely individualistic notion,⁵⁵ it should be clear that mental-state individuation in terms of cognitive value will center much more on the thought and action and discriminative abilities of individuals than mental-state individuation in terms of conventional linguistic meaning. For one, as we saw, cognitive value is “invoked to explain action and epistemic inquiry” (“Intellectual Norms,” p. 271). While cognitive value may not exhaust itself in these tasks, it is clear that in these respects at least, the cognitive patterns of the actual and the counterfactual subject will resemble each other in a way that they will not if we individuate their thought-contents in terms of conventional linguistic meaning. One might still feel that this *similarity* does not yet amount to *sameness*. Thus, Burge’s conclusion might still hold: the actual and the counterfactual subject are thinking different thoughts, even though this conclusion is no longer supported by an argument from the different conventionally established concepts that their respective languages afford them.

This might be true. But clinging to this truth will have the pernicious consequence of depriving the phrase ‘sameness of thought’ of any real application. The considerations underwriting identity conditions for cognitive value are likely to be so complex that no two people are ever likely

⁵⁴ See “Intellectual Norms and the Foundations of Mind,” pp. 269-274.

⁵⁵ See, for instance, “Intellectual Norms and the Foundations of Mind,” p. 273: “Schematically, our thesis is that in certain senses, ‘use,’ socially conceived, necessarily ‘fixes’ conventional linguistic meaning; but ‘use’ (individually or socially conceived) does not necessarily ‘fix’ cognitive value.” For similar points, emphasizing the Fregean background, see *Truth, Thought, and Reason: Essays on Frege*, Introduction, pp. 57-58, and “Frege on Sense and Linguistic Meaning,” p. 261.

to think the same thought in this stricter sense.⁵⁶ So this cannot be the sense of ‘sharing thoughts’ that underwrites our philosophical account of communication. Thus, pending a fuller examination of the example, I stipulate the following: the cognitive value of the actual speaker’s concept ‘arthritis’ and the counterfactual speaker’s term ‘arthritis’ is so relevantly similar that the mental contents expressed by each speaker’s “I have arthritis in my thigh” is, for all practical purposes, the same. Thus, Burge’s conclusion in “Individualism and the Mental” is false.

A final note: the argument of “Individualism and the Mental” does display some sensitivity to the considerations that prompt the distinction between cognitive value and conventional linguistic meaning. Consider the following passage:

None of the foregoing is meant to deny that frequently when a person incompletely understands some attitude content he has some other attitude content that more or less captures his understanding. [...] There are also cases in which it is reasonable to say that, at least in a sense, a person has a notion that is expressed by his dispositions to classify things in a certain way – even if there is no conventional term in the person’s repertoire that neatly corresponds to that ‘way.’ [...] On the other hand, the fact that such attributions are justifiable *per se* yields no reason to deny that the subject (also) has object-level attitudes whose contents involve the relevant incompletely understood notion. (Burge, “Individualism and the Mental,” p. 125)

The concession that a person may have “a notion that is expressed by his dispositions to classify things in a certain way – even if there is no conventional term in the person’s repertoire that neatly corresponds to that ‘way’” – is clearly a nod in the direction of the considerations that prompt the

⁵⁶ In all likelihood, the same will hold for a single thinker at two different times. For some thinkers, notably Jerry Fodor, this admission would spell disaster (cf. *Psychosemantics*, pp. 57-60). But then again, Fodor’s program is entirely shaped by the felt need to secure a presumption that we do not share: that scientific psychology must operate on law-like generalizations. Fodor also argues that the common folk would scoff at the idea that relative similarity of thought contents is enough to secure the everyday notion of communication. I think this is entirely wrong, but cannot pursue this matter here.

distinction between cognitive value and conventional linguistic meaning. Nonetheless, Burge goes on to argue that this concession constitutes no reason to deny that the speaker also has the attitude ascribed with the conventional, incompletely understood content.⁵⁷

But is this a stable line to take? I think not. The case we are considering certainly seems to present us with two conflicting mental state attributions, one based on cognitive value and another based on conventional linguistic meaning. The recommendation from “Intellectual Norms and the Foundations of Mind” is, reasonably in my view, to let the former take precedence. Moreover, this recommendation *does* seem to me to yield a reason to deny that “the subject (also) has object-level attitudes whose contents involve the relevant incompletely understood notion.”

(There is, we note, a fairly well-entrenched philosophical strategy the rationale for which is precisely to allow us to take this pluralistic line. Examples include Hilary Putnam’s distinction between narrow content and wide content and Brian Loar’s distinction between psychological content and social content.⁵⁸ Were Burge to jump on this bandwagon, his two-attitude response to the problem would at least receive a theoretical foundation. The problem, we saw in chapter 1.6, is that Burge has repeatedly and resoundingly declined this invitation. Indeed, no one has provided more incisive arguments against such strategies than he.)⁵⁹

New light is cast on this question in Burge’s recent Postscript to “Belief *de re*.” Recall how much of the argument of “Individualism and the Mental” hinges on (i) observations concerning putative ‘common practice’ in belief attribution, (ii) a claim that we should take common practice as

⁵⁷ I will have more to say about this in ch. 3.

⁵⁸ See Putnam, “The Meaning of ‘Meaning,’” pp. 219-222, Brian Loar, “Social Content and Psychological Content.” See also Colin McGinn, “The Structure of Content.”

⁵⁹ For the case against Putnam, see Burge, “Other Bodies,” sections III-IV. For the case against Loar, see Burge, “Phenomenality and Reference: Reply to Loar,” pp. 436-438. For a general statement against two-content strategies, see Burge, *Foundations of Mind*, Introduction, pp. 11-13. Another question to consider is how, unless he is to rely on some such distinction, the non-committal line suggested in the above quote comports with Burge’s frequently professed realism about mental description. (In passing we also note Akeel Bilgrami’s allegation (in his *Belief & Meaning*) that Burge’s anti-individualism has always operated with a covert ‘bifurcationism’ about content. Neither Bilgrami’s allegation nor the general case against such bifurcationism can be addressed here, but will be taken up again in chs. 4 and 5.)

involving conventional concepts, even where these are incompletely understood by the speaker, and (iii) the claim that in this, common practice is successfully tracking the (actual) mental contents of the speaker. These claims on behalf of common practice are certainly put in stark relief by the following comment, looking back at “Belief *de re*”:

One idea in this discussion seems to me undeveloped, but still worthy of development. It is an idea about the relation between what is normally attributed in the language, even in uses of ordinary language geared to describing propositional attitudes, and what the actual representational contents of an individual’s attitudes are. I distinguished sense₃ [i.e., that which is captured in conventional linguistic meaning] – what is attributed in oblique or non-extensional occurrence in attributions of propositional attitudes – from sense₁ – the way of thinking or cognitive value that a thinker associates with an expression. I suggested that sense₃ is coarser-grained than sense₁. That is, what is attributed will often be less specific or fine-grained than the actual way of thinking engaged in by the individual thinker to whom the attitude is attributed. This is so *even for correct oblique occurrences in propositional attitude attribution whose purpose is to describe propositional attitudes as fully as ordinary communicational conditions will allow*. What is easily and conventionally attributed as a way of thinking may be unexceptionable as far as it goes. Yet it may fall short of capturing the individual’s way of thinking.⁶⁰ (Burge, “Postscript to ‘Belief *de re*,’” p. 80)

With this, we have surely left the argument of “Individualism and the Mental” far behind.⁶¹

Nonetheless, there are critical problems also with this new line. For there is a crucial element of the

⁶⁰ See also “Postscript to ‘Belief *de re*,’” pp. 66, 67. It is interesting to note that Burge’s contention here parallels a view that has some currency in linguistics. For instance, Robyn Carston writes that “underdeterminacy is an essential feature of the relation between linguistic expressions and the propositions (thoughts) they are used to express; generally, for any given proposition/thought, there is no sentence which fully encodes it” (Carston, *Thoughts and Utterances: The Pragmatics of Explicit Communication*, p. 29). It should be clear why this is a view that sits uneasily with the program of “Individualism and the Mental.”

⁶¹ Contrast for instance with the following statement of the argument of “Individualism and the Mental”: “By varying the *uses of words* in his linguistic community, I found that the content of his propositional attitudes varied” (Burge, “Other Bodies,” p. 84). We note that this line also pulls the rug out from under the weaker reading of the conclusion which Burge sometimes offers (cf. “Postscript to ‘Individualism and the Mental,’” p. 176). Burge often claims that attribution of incompletely understood content is not automatic, just that it is possible, as is evidenced by the fact that it frequently occurs in common practice. Presumably, the suggestion is also that its occurrence in common practice shows that it is justified. Both the evidentiary status of common practice and the justification that are supposed to flow from it is seriously jeopardized by the line of argument cited above.

old view that survives here, namely the idea that common practice in content-attribution is somehow constrained to operate with conventional linguistic concepts, literally understood. I see no other reason why Burge should feel compelled to infer, from the mere distinction between cognitive value and conventional meaning, the conclusion that even when it works optimally well, communication may fail to capture the full fineness of grain of our thoughts. In my final conclusion in ch. 5, I will have more to say about the predicament that this view leaves us in.

3

INCOMPLETE CONCEPT MASTERY

3.1 Introductory

In chapter 2, I argued against Burge's conclusion that the actual and the counterfactual subject must, in some modally robust sense, be thinking different thoughts. More generally, we found no reason to suppose, as Burge's thesis seems to entail, that the range of socially constituted concepts available to us places philosophically interesting constraints on the range of thought-kinds that we are capable of thinking. Thus, we found no justification for Burge's claim that the counterfactual subject is incapable of thinking *arthritis*-thoughts on account of the fact that there is no such concept in his language. The example I provided shows one case in which the counterfactual subject comes to share in a reflective agreement with the actual subject concerning arthritis, and is therefore presumably capable of thinking the thought embodied in this agreement. He comes to share in this agreement, moreover, without anything happening that could plausibly be described as his coming to acquire a new concept.¹

This was intended as an argument against the conclusion which is supposed to establish the general thesis of anti-individualism. Yet the viability of this conclusion has received strikingly little

¹ Of course, there is always room for arguing that the mere fact that he is capable of sharing in a reflective agreement concerning arthritis shows that he must possess the *arthritis*-concept after all. But this would trivialize the concept-cognition thesis and deprive the subject's social setting of any distinctive role in "shaping the individual and the content of his thought" (Burge, "Individualism and the Mental," p. 100).

attention in the critical literature. Instead, discussion has typically centered on a move which occurs earlier in the argument, namely the claim that we can rightly portray a subject as thinking with a certain concept (e.g., *arthritis*) even though he possesses a demonstrably incomplete grasp of that concept (e.g., he thinks that arthritis may occur outside the joints).

In focusing their attention on this move, would-be critics are following Burge's own lead. Burge presents the *doctrine of incomplete concept mastery* (as I will refer to it) as the theoretical focal-point of anti-individualism, the assumption without which the thought-experiment which purportedly establishes anti-individualism could not even get off the ground.² Most of the argumentation on offer in "Individualism and the Mental" is directed at undermining considerations that would seem to speak against the doctrine of incomplete concept mastery.

I have chosen to approach the matter from the opposite direction: *first* examine the validity of the ultimate conclusion at step 3 of the thought-experiment (chapter 2), *then*, in this chapter, examine the philosophical move that licenses the crucial intermediary conclusion at step 1 of the thought-experiment. The reason is simple: the argument of chapter 2 shows that even if we were to grant to Burge the doctrine of incomplete concept mastery, the major conclusion of the thought-experiment still does not follow. That is, unlike Burge himself, and unlike so many of his critics, I do not think that acknowledging the thesis of incomplete concept mastery by itself establishes anti-individualism.

This insight dramatically alters the dialectical position we find ourselves in with respect to Burge's argument. In order to resist Burge's conclusion, we no longer suffer the pressing need of a robust *reductio* argument against the very idea of incomplete concept mastery. As it is, I can take a more relaxed stance on the matter: my aim in this chapter is not so much to *defeat* all talk of incomplete concept mastery as to *deflate* it. In brief: I will aim to show that while attributions of

² See, in particular, Burge, "Individualism and the Mental," p. 111, 112.

thoughts containing incompletely understood, socially constituted concepts may in some sense be correct (i.e., they may provide true descriptions of the subject's thought contents), rival attributions involving different, perhaps idiosyncratic, concepts can do (at least) as well. I will show, then, that none of the arguments that Burge offers for the doctrine of incomplete concept mastery decisively favors his view over the competitors. In fact, when all the qualifications that Burge enters on his theory are duly noted, it will be seen that the anti-individualist thought attributions and its competitors are actually just notational variants of attributions of the same attitude.

This 'deflationary' approach may seem overly relaxed and conciliatory, insofar as it allows that both the anti-individualist thought-attribution and the variant notations that I will lay out here can both, in a sense, turn out to be correct. Thus, it falls to the argument of chapter 4 to show why we nonetheless have strong reasons to favor my view. The reasons have to do with the role that thought attributions play in portraying the subject as a rational agent engaged in intentional action. These are roles that our attributions cannot play if they follow the pattern that Burge argues that they do. In brief, then, my argument is that anti-individualism may be able to provide true descriptions of the agent's thought contents, but descriptions which are nonetheless inappropriate to the tasks of capturing the rational force of the agent's reasoning and explaining his actions.

Pace Burge, my argument will not pivot on the claim that the very idea of incomplete concept mastery names a metaphysical impossibility. I can allow, for instance, that this is how we should describe someone who is genuinely confused, someone who is committed to all of p , $p \rightarrow q$, and $\text{not-}q$.³ The important point to note is that the subjects of Burge's thought-experiments are *not* genuinely confused in this sense. Do they, then, "completely master" their concepts? It will transpire that the whole matter of "conceptual mastery" in Burge's sense is something of a red herring. What

³ But, of course, there are notorious problems with saying precisely in what sense such a person engaging in thought at all on that occasion. More on this, and how this differs from the subject putatively suffering from incomplete concept mastery, in ch. 4.8-4.9.

matters is not whether a subject can say, as if in advance of reflection on the matter, what his concepts apply to and what applying a concept on some occasion commits him to. The point is that his thinking about just such questions – his (critical) reasoning – exhibits, when seen in the right light, a certain coherent and cogent pattern, a pattern which it is our task as interpreters and fellow human beings to capture and consider. What I will show in chapter 4 is that this is a “light” that Burge’s incompletely understood concepts are incapable of casting.⁴

But these are matters for chapter 4. The object of this chapter is to obtain a conclusion which will, I hope, nicely complement the conclusion we obtained in chapter 2. In chapter 2, recall, we proceeded in the following way: we assume, with Burge, that “I have arthritis in my thigh” as uttered by the actual speaker and the same words uttered by the counterfactual speaker are expressions of relevantly different thoughts.⁵ What follows from this concession about each thinker’s *capability* to think the other’s thought? Our answer: “Nothing.” By contrast, this chapter aims to withdraw the concession given in chapter 2: these are not, so far as Burge has given us any reason to think, different thought contents at all.

Unfortunately, Burge provides little by way of a direct exploration of the idea of incomplete concept mastery. This may be because he thinks that the practice of attributing incompletely understood concepts is deeply ingrained in our ordinary manner of understanding each other, and so that we *know* what the phenomenon of incompletely understood concepts is already. As Burge claims that it is philosophical sophisticates, not the ordinary folk, who typically object to this practice, most of

⁴ Is it nonetheless a consequence of my view that appropriately attributed thought contents must be restricted to those that only involve concepts the subject completely masters, in Burge’s sense? Not at all. The concept we should attribute to the subject is determined by the *actual* pattern of his reasoning. Certainly we hope that this is a pattern that the subject himself will be able to retrieve or recognize, at least on reflection. But lack of such reflective activity in no way entails that he is not thinking with the concept in question.

⁵ That is to say, they are different expressions of different thoughts not merely in virtue of the indexical expressions contained in each.

what he has to say about it is defensive, designed to deflect the various misgivings that theoretical reflection might shore up about the idea of incomplete concept mastery. In lieu of a direct statement, we are forced to go the circuitous route of considering the set of hypothetical objections which Burge entertains and then rejects at “Individualism and the Mental” IIIc-d.⁶ These are a motley crew, as are the arguments that Burge levies against them. I am afraid that this form of paragraph-by-paragraph commentary will make for grueling reading, but it will be our best way to get a handle on the topic of this chapter.

3.2 Incomplete concept mastery: the basics

As Burge himself points out, the idea of incomplete concept mastery is the *conditio sine qua non* of the arthritis thought-experiment.⁷ As we have seen, the thought experiment presents us with a subject, dependably rational and mostly conversant with the ins and outs of his language, who utters the sentence “I have arthritis in my thigh.” Our question is, how are we to understand what this person is saying?; which thought is this person giving expression to? The thought-experiment is designed so as to elicit from us the following knee-jerk reaction: whatever he is thinking, it cannot be a thought involving our (public) concept *arthritis*, for that concept simply means “rheumatoid ailment of the joints.” The intuition underlying this knee-jerk reaction appears simple enough: to be counted as being in possession of a concept – as capable of thinking with that concept – one must in some sense know what that concept is. Call this the *knowledge-requirement* of concept possession.

Assume, then, that we could define the concept *arthritis* as *rheumatoid ailment of the joints only*. Now if we were to portray our subject as thinking with our concept *arthritis*, then we should be able

⁶ Burge, “Individualism and the Mental,” pp. 116-130.

⁷ Burge, “Individualism and the Mental,” p. 111, 112.

to substitute the expression ‘rheumatoid ailment of the joints only’ for the expression ‘arthritis.’⁸ The resultant attitude-attribution – ‘I have a rheumatoid ailment of the joints only in my thigh’ – sees us well on our way toward portraying our subject as a logical idiot, contradicting the assumption of rationality that we started out with.⁹ Thus, unless our subject truly is a logical idiot (in which case our attempt at uncovering the conceptual basis of his thinking would be fruitless anyway) we must assume that something is amiss. We should strive to ascribe to him a concept that is in consonance with the underlying drift of his thinking. Perhaps he has simply misunderstood the meaning of the word ‘arthritis.’ Perhaps, for instance, he is really thinking with the concept *rheumatosis*, and mistakenly assumes that this is the concept most people express with the word ‘arthritis.’ If this is true, then we should modify our interpretation accordingly.

Such reflections provide a rationale for what Burge calls the reinterpretation-strategy. According to this line of thought, the cognitive error that a literal interpretation would attribute to the subject is such as to force a reinterpretation of his words, portraying him as thinking with a different concept than most of us do when we use the word ‘arthritis.’

In the face of these considerations, Burge maintains that nothing need stand in the way of a literal (or conventionalistic) interpretation of the subject’s words.¹⁰ Burge meets our knee-jerk reaction by attempting to undermine the intuition that underlies it, namely that in order to be correctly said to be thinking with a certain concept, a subject need possess full understanding of it (i.e., a strong insistence on the knowledge-requirement of concept possession). Burge disputes this:

⁸ As should be clear, it is the knowledge-requirement which would license this substitution in what would otherwise be an opaque context.

⁹ Cf. Akeel Bilgrami, “A Trilemma for Redeployment,” pp. 25-26. Kirk Ludwig, “The Myth of Social Content,” pp. 5-13, argues that the independence of syntax and semantics forces this conclusion.

¹⁰ It is important here to be precise about the scope of Burge’s claim. His claim has never been that we are always bound to, or always justified in, giving a literal interpretation of a speaker’s words, as some commentators seem to assume (cf. Burge, “Social Anti-Individualism, Objective Reference,” pp. 308-309). The claim is rather that incomplete mastery of a concept does not automatically force reinterpretation of the speaker’s words, and that “literal interpretation is *ceteris paribus* preferred” (Burge, “Individualism and the Mental,” pp. 116-117). Thus, bringing out supposed counterexamples in which a speaker’s incomplete understanding does seem to force reinterpretation is of limited use.

thinking with a concept does, he argues, require a certain minimal level of understanding of the concept, but by no means a full understanding. Call this the *entry-level comprehension requirement* of concept-possession.¹¹ When the strong version of the knowledge-requirement falls away, so does any general justification for intersubstitutability, and we no longer run the risk of portraying our subject as a logical idiot by attributing to him the thought “I have arthritis in my thigh.”¹² The only cognitive shortcoming we would thereby attribute to him is that of failing to know that arthritis may not occur outside the joints. This is a perfectly explicable error.

3.3 Objection from *de re* attitudes

Burge organizes the hypothetical objections he considers under two headings; methodological and philosophical. I will follow him in examining the methodological proposals first. Burge lists four such proposals.

¹¹ Thus, Burge is willing to exempt, for instance, foreigners and children from the literalist bias, as they may fail to satisfy the entry-level comprehension requirement. See Burge, “Individualism and the Mental,” p. 119.

¹² Bilgrami raises the worry of what is left of one’s externalism if one ends up disallowing intersubstitutability in this sense. (See Bilgrami, “A Trilemma for Redeployment,” pp. 25-26.) The question is an interesting one, but not one that I can pursue here. We can note, however, that as Bilgrami levels the charge, it rests upon a simple misreading of Burge. Bilgrami writes: “But it *does* look as if externalist doctrines which have it that a subject’s concepts are determined by their extension or reference (at least as Burge, Kripke, and Putnam understand reference) will make Burt come out as being inconsistent. By these theories, the ‘water’-concept [...] in Burt’s judgment is to be specified in terms of its extension, which is the class of all instances of the substance H₂O. This allows us to *plug in* ‘substance which is H₂O’ for ‘water’ in reporting the content of his state of mind, the content of his belief. So he must be said to believe that a substance which is H₂O is not H₂O” (ibid.). It should be clear that this argument completely overlooks the fact that this is expressly not Burge’s view, and that Burge indeed criticizes Kripke and Putnam precisely for this reason. (Recall the argument concerning narrow and wide content in ch. 1. Bilgrami is in effect imputing to Burge the view that if mental content is not narrow, then it must be wide. But as we have seen, this is precisely *not* Burge’s view.) Meanwhile, Bilgrami is entirely justified in asking what remains of one’s externalism if one does not allow intersubstitutability in this sense. Burge’s answer is, of course, that what remains is an anti-individualist form of externalism: it is true that reference is mediated by concepts, but the concepts in question aren’t *individualistic*, but are defined by linguistic conventions and belong in some kind of social space. This is why Burge can claim both that mental contents are individuated at the level of concepts and that his theory is externalistic. I have no doubt that this is a coherent view of mental content. The object of this dissertation is to argue that it is not practically sustainable in view of the various tasks we need the theory of content to aid us in the execution of, like for instance the explanation of action.

- i) Interpret the attitudes in question as “*de re* attitudes *of* entities not denoted by the misconstrued expression. For instance, the subject’s belief that he has arthritis in his thigh might be interpreted as a belief *of* the non-arthritic rheumatoid ailment that is in the thigh” (Burge, “Individualism and the Mental,” p. 121).
- ii) Claim that where incomplete understanding occurs, “the subject’s attitude or content is indefinite” (p. 122).
- iii) “Attribute a notion that just captures the misconception, thus replacing contents that are apparently false on account of the misconception, by true contents” (p. 122); (e.g., replace the attribution of *arthritis* with an attribution of *tharthritis*, the concept of a rheumatoid ailment that strikes also outside the joints).
- iv) Treat the speaker’s error as a “purely metalinguistic” error: “Thus the patient’s apparent belief that he has arthritis in the thigh would be reinterpreted as a belief that ‘arthritis’ applied to something (or some disease) in his thigh” (p. 122).

For all of these proposals, Burge stipulates the following burden of argument: “in order to overturn the thought experiment, these methods must not only establish that the subject held the particular attitudes that they advocate attributing; they must also justify a *denial* of the ordinary attributions literally interpreted” (pp. 122-123).¹³

Burge is quite brief with the two first proposals. The problem with the proposal invoking *de re* attitudes about objects not denoted by ‘arthritis’ is that it “ignores the oblique occurrence of

¹³ For the record, I have strong doubts that this burden of proof is fairly allotted, or, for that matter, that it is consistent with Burge’s own best thinking. Since Burge must presume that the attitudes in question are *different* attitudes, then establishing that the subject has this other attitude will come partly as a result of undermining our reasons to think that he has the attitude that Burge advocates that we attribute. Perhaps it is conceivable that he also has the latter attitude, but nothing Burge will have said should incline us to think that he did. See chapter 2, Addendum, and section 8 of this chapter for further remarks on this.

‘arthritis’ in the original ascription,” an occurrence which “bear[s] on a characterization of the subject’s viewpoint” (Burge, “Individualism and the Mental,” p. 121-122). The patient, Burge concedes, probably has a *de re* attitude of the sort invoked in the critical proposal. Everyday communication seems rife with similar examples: I ask you “why don’t we have our coffee on the veranda?” and without flinching you follow me out to the porch; I tell you how much I enjoyed your lute rendition of that Kapsberger piece, and you nod and say “thank you” even though you were playing the theorbo.

Nonetheless, our subject thinks that the disease in his thigh is arthritis, and this is not something we are able to convey with this kind of reinterpetive strategy. In this sense, “the appeal to *de re* attitudes [...] is not adequate to the task of reinterpreting these ascriptions so as to explain away the difference between the actual and counterfactual situations. It simply overlooks what needs explication” (ibid., p. 122).

Burge’s point is that this ‘objection’ is not really an objection at all. *Of course* the subject is thinking of an object not denoted by the term ‘arthritis’. So far from being in dispute, this fact is precisely what lends vividness to the example under consideration. The sense in which this emphasis on *de re* reference “simply overlooks what needs explication” is that specifying full-blown representational content requires that we note not just the object that the person is thinking about, but also the particular conceptual angle he takes on that object.¹⁴ Thus, while the subject is indeed thinking of an object not denoted by the expression ‘arthritis,’ he does appear to be thinking of that object by deploying the concept *arthrititis*. And that is our problem.

So far, Burge’s response is adequate, although perhaps only in light of the inanity of the hypothetical objection he is responding to. For the response gives us nothing more to work on in our assessment of the idea of incomplete concept mastery. Burge may be right to point out that

¹⁴ See section 1.5 for more.

attributions of expressions like ‘arthritis’ in oblique occurrence can “bear on a characterization of the subject’s viewpoint,” and that this fact is simply overlooked on the *de re* objection we are currently considering. But this observation adds no support to Burge’s claim that it is indeed the term ‘arthritis’ which best captures the subject’s viewpoint in this particular case. That is, when it comes to the assessment of the incomplete concept mastery thesis, we are still where we were: the fact that the subject uses the word ‘arthritis’ to express his thought might be construed as *prima facie* evidence in favor of the attribution of the concept *arthritis*, and so in favor of the idea of incomplete concept mastery. On the other hand, the fact that the subject so obviously misapplies this word might be construed as *prima facie* evidence against this attribution. Burge is right to point out that the *de re* objection is easily refuted, but would be wrong to suggest that his case is strengthened by the refutation.

3.4 Objection from indefiniteness of content

According to objection ii) (which Burge attributes to Descartes) we should think of the subject as expressing no definite attitude at all in cases where such conceptual confusion is involved.¹⁵ Burge concedes that there may be cases where the subject is *so* confused that it would be appropriate to describe him as having nothing particular in mind. He argues, however, that this is not one of those cases.¹⁶ We know this, says Burge, because in our case, “[t]he subject and his fellows will typically know and agree on precisely *how to confirm or infirm his beliefs*” (Burge, “Individualism and the Mental,” p. 122).

¹⁵ Cf. Burge, “Individualism and the Mental,” p. 122.

¹⁶ I will be exploiting a similar observation in ch. 4.7.

I have less interest in the fate of this objection, in some part because I am unsure how we should approach this talk of attitudinal definiteness and indefiniteness. Burge invokes Descartes' name, but gives us little to work with otherwise. There is, however, a fuller discussion of Descartes later in the paper.¹⁷ Here Burge draws a contrast with an argument due to Bertrand Russell, according to which the mind is infallibly 'acquainted' with its contents.¹⁸ Descartes, on the other hand, does hold that one may fail to grasp the contents of one's thoughts, thereby, according to Burge, implicitly nodding in the direction of the doctrine of incomplete concept mastery. However, Descartes spoils this insight by assuming that such problems can be overcome essentially by focusing one's mental energies better. Where the proper concentration fails us, Descartes seems to hold that "no definite content can be attributed" (Burge, "Individualism and the Mental," p. 133).

I am inclined to agree with Burge that the task of achieving clarity of thought is not in general something we can entrust to introspection alone, and I have no interest in defending Descartes on this score. Meanwhile, this tells us nothing about how we should approach the question at hand. I find the implied distinction between confused thoughts adequately grasped (Descartes) and clear thoughts inadequately grasped (Burge) quite unhelpful. Both elements of this distinction posit a certain kind of substantial epistemic relation between ourselves and our thoughts – grasping or understanding – which might turn out to be part of the problem rather than part of the solution.¹⁹

While we have found little cause for concern in the argument from indefiniteness of content, Burge's take on Descartes is worth noting for another reason. Burge repeatedly insinuates that any resistance to the notion of incomplete concept mastery must betray covert allegiance to some sort of

¹⁷ Burge, "Individualism and the Mental," pp. 133-134.

¹⁸ Burge cites Russell's *Mysticism and Logic*, p. 221.

¹⁹ Looking back at the matter, Burge might agree with this point, since so much of the criticism of anti-individualism has come to focus on the question of self-knowledge. I examine the problem of self-knowledge in relation to anti-individualism in chapter 4.

Cartesian model of introspection (or one of its empiricist successors). Burge does not offer an argument for this insinuation, except insofar as an argument might be thought to be embedded in the distinction we just discussed. To that extent, the argument is clearly a bad one. No argument I offer in this dissertation will rely, explicitly or implicitly, on the notion that people can only be credited with thinking with concepts they completely understand.²⁰

3.5.0 Objection from variant concepts

So much for objections i) and ii). Objections iii) and iv) will require more attention. Objection iii), recall, urges us to “attribute a notion that just captures the misconception, thus replacing contents that are apparently false on account of the misconception, by true contents” (Burge, “Individualism and the Mental,” p. 122).²¹ According to objection iv), on the other hand, we should treat the speaker’s error as “purely metalinguistic”: “Thus the patient’s apparent belief that he has arthritis in the thigh would be reinterpreted as a belief that ‘arthritis’ applied to something (or some disease) in his thigh” (Burge, “Individualism and the Mental,” p. 122). As Burge notes, proposals iii) and iv) are

²⁰ We also note in passing that this Cartesian claim concerning the ‘indefiniteness’ of thoughts involving incompletely understood concepts should not be confused with another idea that Burge occasionally levies arguments against, namely the Quinean-Davidsonian claim that the contents of a subject’s attitudes may be *indeterminate*. The reason is clear: the indeterminacy thesis applies to all specifications of thought-contents, not just those involving incompletely understood concepts (if indeed Quine and Davidson should find any use for such talk, which is highly doubtful). Further, the argument Burge gives against the claim concerning indefiniteness – that the subject and his fellows agree on how to confirm or infirm the belief in question – simply does not touch on the claim concerning indeterminacy. For the indeterminacy thesis is precisely the claim that there may be multiple non-equivalent specifications of the contents of a body of beliefs, which yet have the same truth-conditions. (Which showcases another difference: the indefiniteness thesis is applied to single, individuated attitudes, whereas the indeterminacy thesis applies to bodies of attitudes.) For more on the indeterminacy thesis, see Quine (*Word & Object*, §16, “Ontological Relativity”), Davidson (“Radical Interpretation,” pp. 151-154, “The Inscrutability of Reference”). Illuminating commentary can be found in Bjørn Ramberg, *Donald Davidson’s Philosophy of Language*, ch. 7.

²¹ I.e., this objection urges us to replace the attribution of *arthritis* with an attribution, say, of *tharthritis*, the concept of a rheumatoid ailment that strikes also outside the joints.

often invoked in tandem, “attempting to account for an ordinary content attribution in terms of a reinterpreted object-level content together with a metalinguistic error” (ibid., p. 122).²²

We will start by looking more closely at Burge’s arguments against the object-level reinterpretation strategy. Burge offers a series of critical angles on the matter. I will deal with these *seriatim* in the following. Again, bear in mind the following: the objections which Burge considers all attempt, in one way or another, to show that the subject *must* be thinking with some concept other than *arthritis*, thus rendering false any attribution invoking that concept. This is not quite my aim in this chapter. My aim here is simply to show that attributions involving variant concepts can *also* provide true descriptions of the subject’s thought contents, and that – so far as Burge has given us any reason to think – these may be true descriptions of the same thought contents as those that feature in the anti-individualist attributions. (Again, it falls to the argument of chapter 4 to show why we should nonetheless prefer the attributions involving variant concepts.) Thus my arguments in this section will not necessarily constitute attempts to defend these hypothetical objections against Burge’s arguments.

3.5.1 No-retroactive-reinterpretation argument

The object-level reinterpretation-strategy, says Burge, has little to commend it as a general strategy for dealing with these kinds of cases. For instance, learning of a subject’s incomplete understanding of a concept does not typically lead us to retroactively reinterpret previous attributions involving the concept which were “intuitively unaffected by the error” (Burge, “Individualism and the Mental,” p.

²² There is a complex dialectic between objection iii) (with or without supplementation by objection iv)) and objection i). It is plausible to see objection iii) as simply trying to supply what was missing in objection i), namely some mention of the conceptual angle the subject takes on the object he is thinking of. Meanwhile, objection iv), as we shall see, seems patently incapable of delivering anything on its own. Its purpose seems only to be to provide plausibility to the move contained in objection iii).

123). For instance, learning that Lester does not know that an oral agreement could also constitute a binding contract will not typically incline us toward reinterpreting an attribution to Lester of the belief that he is bound by a one-year contract with his landlord. In fact, although Burge does not note this, the same holds of contemporaneous cases. I may be your legal advisor, and I happen to know of Lester's incomplete mastery of the concept *contract*. But if what we have before us is a draft of a written contract, I may simply tell you that Lester is happy with the proposal and is willing to sign. In this particular case, I do not feel compelled to specify to you that Lester really is operating with a slightly different notion of what a contract is than you have.

Our question, recall, is whether we should portray Lester as thinking with a different concept than the rest of us do when he considers contractual matters, on account of the fact that he does not know that an oral agreement can also be legally binding. Burge argues that, as a matter of fact, we don't. True, where oral agreements are involved, we would do well to keep Lester's incomplete understanding of the concept "in mind." But where oral agreements are not involved, we simply portray Lester as *agreeing to the contract*, and feel no compulsion to specify that Lester thinks 'contract' applies to a slightly narrower set of objects than the rest of us do. If we really thought Lester was operating with a different concept than we do, we should substitute Lester's personal concept for our concept *contract* in our specifications of his beliefs across the board. Since we do not do this, we are, Burge argues, implicitly portraying him as thinking with the concept *contract*, but suffering from an incomplete understanding of that concept, which ignorance is relevant in some contexts, but not in others.

This is an interesting observation, but it does not clearly favor Burge's view in the way that he supposes. Burge's argument is based on the observation that actual attributive practice does not retroactively reinterpret except where this is relevant to understanding the agent's words and actions. But it would be amply sufficient to vindicate the reinterpreter's position if Burge were to concede,

as he certainly seems to concede here, that we *do* reinterpret in cases that *are* affected by the error. If we are less vigilant about cases where the “error” is simply not relevant, this can no be chalked down to just that – it is not relevant.²³ We simply have no need to keep track of indefinitely fine distinctions among possible background conceptualizations where the differences in question do not affect our ability to reach reflexive agreement about the matter at hand.²⁴ The point about relevance suggests another problem in Burge’s argument here. In cases where I am aware that you are thinking of the matter “in a different way” than I am, yet I refrain from remarking on this because it is irrelevant to our present concerns, we should precisely *not* take the occurrence of the problematic word in my attitude report as “bearing on a characterization of [your] viewpoint.” This is an exception licensed by the irrelevance. This can be seen in interlinguistic contexts, where we breezily (and appropriately) render, say, an English speaker’s words “It’s cold out; you should wear a hat” in Norwegian as “det er kaldt ute; du burde ta på deg lue,” even though I am quite aware that the English word ‘hat’ and the Norwegian word ‘lue’ do not pick out the same extension.

Burge’s reading of the situation seems to confront us with the following choice: either (i) attribute to Lester an eccentric concept throughout, but specify that this concept deviates from the standard one only where difference is relevant to our practical concerns; or (ii) attribute to Lester the public concept *contract* throughout, but specify that his beliefs about the concept – his *conception*, as

²³ We can say this without endorsing the full force of, say, Dan Sperber and Deirdre Wilson’s programmatic claim that our communicative practices are geared toward the maximization of relevance. See Sperber and Wilson, *Relevance: Communication and Cognition*; “Relevance Theory.”

²⁴ It might be useful to compare this with a line taken by Burge in a later essay: “what is attributed will often be less specific or fine-grained than the actual way of thinking engaged in by the individual thinker to whom the attitude is attributed. This is so *even for correct oblique occurrences in propositional attitude attribution whose purpose is to describe propositional attitudes as fully as ordinary communicational conditions will allow*. What is easily and conventionally attributed as a way of thinking may be unexceptionable as far as it goes. Yet it may fall short of capturing the individual’s way of thinking” (Burge, “Postscript to ‘Belief *de re*,’” p. 80). I think this dramatically overstates the case for a discrepancy between what people are thinking and what we might ordinarily report them as thinking (for more, see my remarks toward the end of ch. 2, Addendum). The point I want to make here, though, is that sometimes it is simply not within our interest to report the full detail of someone’s thought contents – because, as Burge observes, the detail may not be relevant – but that this does not mean that, on these occasions, we take that subject to be thinking the thought conventionally associated with the words we use in the attribution.

we might say – are eccentric only where this is relevant to our practical concerns. But so far, Burge has given us no reason to think that these two strategies will even produce specifications of different thought contents. Moreover, he also gives us no reason to prefer alternative (ii), if it were settled that these were indeed different. So far as I can tell, the only sketch of a systematic argument that Burge offers for this stems from the supposedly privileged role of public concepts in shaping and conveying reflective agreement among persons. But this supposed privilege is precisely part of what is in dispute (as per my argument in chapter 2). Hence the no-retroactive-reinterpretation objection is in no way decisive.²⁵

3.5.2 Shared-belief argument

A further objection is that we need to portray Lester and his lawyer as at least capable of sharing beliefs. For instance, we want to make sense of the idea that the lawyer can successfully convey information to Lester concerning the various contractual relations in which he finds himself (cf. Burge, “Individualism and the Mental,” p. 123). We will be frustrated in this endeavor if we decide on portraying Lester as thinking in terms of one concept, his lawyer in terms of another.

My response to this follows closely on what I said in response to the previous objection, as well as on my argument in chapter 2. Burge appears to think that it is only if we portray two people as thinking with the same concept that we can portray them as thinking about the same state of

²⁵ There is also a hint here of a more serious problem, which I shall explore in more detail in a later section. It would seem that Burge takes the following datum to count as evidence that common practice favors alternative (ii) above: we do not coin and define a new term to use in our attributions of thoughts to the subject. But, as we shall see, this datum is irrelevant to the question at hand. What matters is not which *word* we use in our attributions, but which *concept* we attribute with that word. For obvious reasons, the anti-individualist cannot without begging the question simply assume that we are attributing the concept that is conventionally associated with that word (i.e., *arthritis*). And, in fact, what Burge says on closer analysis actually supports our view, not his: in attributions that are affected by the error, we might still deploy the problematic word in our attributions, but “keep in mind” the fact that the subject thinks the word applies to something it does not. But that is just what it is to attribute a diverging concept – a diverging “way of thinking” – to him.

affairs, and thus as capable of reaching reflective agreement on the matter at hand.²⁶ Burge's argument acknowledges, indeed exploits, the notion that they may have substantially different beliefs about – conceptions of – the relevant concept. Nonetheless, he holds that it is the (presumed) fact that these are substantially different cognitive takes on *the same concept* which allows them to communicate about the matter at hand.

This contention is not well supported. As my argument in chapter 2 shows, two people can perfectly well reach reflective agreement about a subject matter even though we must, even by Burge's lights, conclude that they do not share the relevant concept. The choice between portraying them as sharing a concept but disagreeing about what is involved in that concept, and as thinking with different concepts but largely agreeing about the matter at hand is not, so far as we can tell at this point, one on which we should expect philosophical guidance. Burge is wrong in the first place to assume that it is only by (already) sharing concepts that the two can agree or disagree about the matter at hand.

3.5.3 *Ad-hoc* argument

Burge suggests that one reason ordinary practice does not follow the strategy of object-level reinterpretation is that “finding a reinterpretation in accord with the method is entirely untrivial” and, moreover, that it would be hard to non-arbitrarily delimit its use once found. He writes:

²⁶ It should be noted that Burge does not always put it quite so strongly. When discussing the matter explicitly, he often operates with more guarded claims. Thus, for instance, “Words interpreted in conventionally established ways [...] are critical in maximizing interpersonal comparability” (“Individualism and the Mental,” p. 149); “I believe that the stronger description in terms of shared concepts [...] is part of the best explanation of the transmission of knowledge” (“Postscript to ‘Individualism and the Mental,’” p. 164). There are legitimate questions to be asked, however, about whether Burge's metaphysical arguments are can be sustained on this weaker, more pragmatically oriented approach.

Consider the arthritic patient. Suppose we are to reinterpret the attribution of his erroneous belief that he has arthritis in the thigh. We make up a term ‘tharthrititis’ that covers arthritis and whatever he has in his thigh. The appropriate restrictions on the application of this term and of the patient’s supposed notion are unclear. Is just any problem in the thigh that the patient wants to call ‘arthritis’ to count as tharthrititis? Are other ailments covered? What would decide? The problem is that there are no recognized standards governing the application of the new term. In such cases, the method is patently *ad hoc*. (Burge, “Individualism and the Mental,” p. 123)

Moreover, there is also another sense in which this strategy is *ad hoc*, namely in that it “proliferates terminology without evident theoretical reward” (ibid.). We do not, Burge thinks, improve our understanding of the patient by saying that he has some true belief involving the concept *tharthrititis*, rather than a false belief involving the concept *arthritis*:

It is simpler and equally informative to construe him as thinking that arthritis may occur outside the joints. When we are making other attributions that do not directly display the error, we must simply bear the deviant belief in mind, so as not to assume that all of the patient’s inferences involving the notion would be normal. (Burge, “Individualism and the Mental,” pp. 123-124)

There are a few things we must note in connection with these claims. First, Burge’s stipulation that the reinterpretation strategy involves “making up” a new *term* is false, or at best, inessential. Moreover, it suggests a profound confusion. For obvious reasons, the advocate of the reinterpretation strategy has no need of making up a new term. He can use the word ‘arthritis’ in his attributions as surely as the anti-individualist. What matters is what *concept* we think of ourselves as attributing with that word on the occasion. Burge here comes extremely close to begging the question in favor of anti-individualism by assuming just the kind of convention-bound word-to-concept (or, if you will, syntax-to-semantics) matching that is in dispute. That is to say, Burge cannot assume that the fact that we go on using a certain word constitutes evidence that we go on

attributing the concept that is conventionally associated with that word without precisely begging the question in favor of conventionalism.

Nonetheless, the problem that Burge points to here does seem to arise at the level of delimiting the employment of the new concept, even if no new word is involved. In brief: if we introduce a new concept to capture our subject's thinking, then we will face the problem of trying to non-arbitrarily delimit the application of this new concept. If we cannot find a way of delimiting the concept (which we won't) then the introduction of this new concept is *ad hoc*, and so, presumably, unjustified. This argument is reminiscent of Wittgenstein's famous 'rule-following considerations,' at least in Saul Kripke's popular but controversial exposition.²⁷ On a widespread understanding of Wittgenstein's argument, we are to conclude that a private language would be impossible, because the subject would have no objective checks on whether he applies his private terminology correctly or incorrectly. As such, a private concept could be applied without censure to anything, and so would fail to embody any meaning at all.

Whatever the merits of this line of reasoning in its original context, it wields no force in the present context. When Burge asks "Is just any problem in the thigh that the patient wants to call 'arthritis' to count as *tharthritis*? [...] What would decide?", he is invoking a problem that simply does not arise. Burge's argument would have some force if what were in question is whether *we* should follow the subject in adopting the concept *tharthritis* in favor of our old concept *arthritis*. The problem would be that where our old concept has "recognized standards governing application," the new concept does not, and so if we adopted this new concept we could not know what we were thinking.

But this is not what is at stake here. We want to understand what and how our subject is thinking, his cognitive perspective on the matter at hand. The concern that fuels the objection is that

²⁷ Ludwig Wittgenstein, *Philosophical Investigations*, §§147-202; Saul Kripke, *Wittgenstein on Rules and Private Language*.

in portraying him as thinking “I have arthritis in my thigh” we may fail precisely in this task. Could it be that our subject is thinking with some other concept, perhaps the concept of a rheumatoid disease that could occur in both joint and muscle? Burge appears to be worried that if we were to try to understand him as thinking with this other concept, then we too would have to take on and think with that concept in our effort to understand him. And this is what we could not do, according to this line of thought. If our only means of saying what the new concept applies to is what our subject fancies that it applies to, then we simply have no concept to share with him, and understanding must fail.

But the error of this way of thinking should be clear by now: Burge thinks that understanding – sharing belief – must proceed *via* shared concepts. We must portray the subject as either thinking with our concept (albeit eccentrically) or thinking with a different concept, in which case we would have to acquire that concept in order to understand him. Since we would have no way of acquiring that other concept, or because there is no such concept to acquire, understanding could not proceed in this manner. *Ergo*, if there is to be understanding at all, it must proceed on the assumption that he is thinking with our concepts, albeit eccentrically.

This is false. Understanding can proceed perfectly well without our sharing concepts. Thus, were we to choose to portray him as thinking with the concept *tharthritis* rather than the concept *arthritis*, we would not be forced to adopt the concept *tharthritis* for ourselves, in such a manner as to be cognitively paralyzed by the question of how this new concept applies. As Burge himself says, we would simply be trying to capture something about his way of thinking about the world. And if this is our project, then having the extension of the variant concept be determined by “whatever the patient wants to call ‘arthritis’” is, in the initial stages at least, exactly what we need to do, for this is the only way to gather up and systematize the available data concerning his manner of thinking

about the world.²⁸ In short, the application of the new concept is not *ad hoc*, but is constrained precisely by our need to make sense of the agent's behavior; constrained, that is to say, by the considerations that led us to abandon the use of the original concept in the first place. In other words, "the recognized standards governing the application of the new term" is the quality of the rationalization it permits.

Burge is right, of course, to point out that finding out what concept it is that our subject is thinking with is entirely non-trivial. But this is as it should be, and to argue that the non-triviality suggests arbitrariness is quite inappropriate. The task of finding out what concepts our subject is thinking with is exactly as non-trivial as finding out what he is thinking, for these are indissolubly part of the same task. There is nothing arbitrary about letting ourselves be guided in this task by the subject's overt actions and verbal testimony, for there are no other sources of evidence that we could turn to for guidance.

Meanwhile, Burge presents the following consideration in favor of his view: "[i]t is simpler and equally informative to construe him as thinking that arthritis may occur outside joints" (Burge, "Individualism and the Mental," p. 123). We note, first, that if both alternatives really are "equally informative," then presumably no information is lost on either choice. This amounts to an implicit acknowledgment that information flow between thinkers does not require sharing concepts. Once we have abandoned the idea that thinking about – making reference to – arthritis requires that one possess the concept *arthritis*, then we understand that we can track his thinking equally well by construing the subject as thinking with a different concept.

²⁸ Bear in mind that this in no way amounts to stipulating that the subject needs to possess complete mastery over his concept. For there is no requirement that the *subject* need be able to say how he applies the concept, or be able to give a perspicuous definition of it, or anything of the sort. Rather, we are letting his usage, his actual pattern of thinking, determine what the concept is. We would normally hope that the subject possess some recognition of the patterns that his thinking exhibits. But it is his actual patterns of thought that determine what his thinking is about, not what he might think his actual pattern of thought is.

But further, once we have disabused ourselves of the idea that the attribution of a different concept must be syntactically marked with the deployment of a different word in the attribution, then we see that Burge's proposal actually favors our view, not his. For if we must "bear the deviant belief in mind" in characterizing the subject's pattern of reasoning, "so as not to assume that all of the patient's inferences involving the notion would be normal" (Burge, "Individualism and the Mental," pp. 123-124), then it would seem that we are already attributing to him a different concept, whether or not we deploy the old word in the attribution. Burge will protest that this is all wrong: we do attribute the conventional concept, but make due note of the fact that the subject's understanding of the word departs from the norm. But this view does not stand up to scrutiny: minimally, it involves (i) negating the premise that we started out with, namely that a concept is what captures the subject's cognitive perspective – his way of thinking –, and (ii) transferring all the explanatory tasks we have associated with the notion of cognitive perspective *from* the concept *to* the subject's conception – his deviant understanding of that concept. The anti-individualist is free to talk this way if he wishes, but must recognize that it comes at the price of depriving concept-talk of any significant explanatory tasks.

Even if we assume that this way of speaking passes muster, Burge has yet to provide an argument that decisively favors his view. We seem to be faced with a choice between, on the one hand, construing the subject as thinking with public concept C, with qualifications for incomplete understanding x, y, z, and, on the other, construing him as thinking with (individualistic) concept C', which differs from the public concept C in respects x, y, z. In lieu of some argument as to why we should take public concepts to have a privileged role in communication between people, neither theoretical nor practical concerns favor the former over the latter. Indeed, so far as Burge has given us any reason to assume, these are simply competing descriptions of the same concept.

3.5.4 Not-true-to-the-evidence argument

Another objection to the reinterpretation method is that it fails to properly account for the subject's behavioral patterns. This objection is particularly important to my line of inquiry here, since I am arguing precisely that my line connects better with action-theoretical concerns than does Burge's. What, then, is the evidence that the reinterpretation line fails to account for? Burge writes:

When caught in the sorts of errors we have been discussing, the subject does not normally respond by saying that his views had been misunderstood. [...] This sort of response would be disingenuous. Whatever other beliefs he had, the subject thought that he had arthritis in the thigh. In such cases, the subject will ordinarily give no evidence of having maintained a true object-level belief. In examples like ours, he typically admits his mistake, changes his views, and leaves it at that. Thus the subject's own behavioral dispositions and inferences often fail to support the method. (Burge, "Individualism and the Mental," p. 124)

There are two lines of thought running through this argument. The most salient line is an empirical claim about how people are disposed to respond when they learn of their mistakes: they correct themselves. The other is a normative claim to the effect that such self-correction is also the appropriate way to respond: it would be "disingenuous" to go on insisting that *we* had misunderstood him. This normative claim is presumably meant to connect with a theme that Burge sounds elsewhere in his article, namely that of a certain form of obligation that speakers stand in to the conventions governing their language.²⁹

²⁹ Cf. Burge, "Individualism and the Mental," p. 147: "Crudely put, wherever the subject has attained a certain competence in large relevant parts of his language and has (implicitly) assumed a certain general commitment or responsibility to the communal conventions governing the language's symbols, the expressions the subject uses take on a certain inertia in determining attributions of mental content to him. In particular, the expressions the subject uses sometimes provide the content of his mental states or events even though he only partially understands, or even misunderstands, some of them. Global coherence and responsibility seem sometimes to override localized incompetence." See also, *ibid.*, p. 148: "I think it profitable to see the language of content attribution as constituting a

I will deal with each theme in turn. Burge observes that “[w]hatever other beliefs he had, the subject thought that he had arthritis in the thigh.” This is, of course, true. Burge would be wrong, however, to assume that we cannot account for this fact without ascribing to him the concept *arthritis*.³⁰ Burge further writes that our subject “will ordinarily give no evidence of having maintained a true object-level belief.” This claim strikes me as presumptuous, if not outright false. Of course the subject will maintain a true object-level belief, and will maintain it through the self-correction. For he will maintain *some* belief about the ailment in his thigh; for instance, that he *is* afflicted with an ailment of some sort, and, perhaps, that the ailment bears striking etiological similarities with arthritis.³¹ This points to an important difference between ordinary empirical mistakes and the kind of empirical-*cum*-conceptual mistakes that Burge is working from. If I thought that the *Rite of Spring* was written before *The Firebird* and you informed me of my mistake, I would simply abandon my belief and “leave it at that.” There is nothing that I was right about in my original belief. By contrast, if I thought that the pains in my thigh were symptomatic of arthritis, and you informed me that arthritis cannot, by definition, occur outside the joints, then I would certainly

complex *standard* by reference to which the subject’s mental states and events are estimated, or an abstract grid on which they are plotted.” I will have more to say about this in ch. 5.

³⁰ It might seem that my argument here is somewhat misguided: in the ordinary case, the attribution of the concept is not so much part of an explanation of the subject’s thought as simply part of the thought content we attribute. This is, of course, true. But which concept we attribute, and thus which thought we attribute (assuming, with Burge, that thoughts are individuated at the level of concepts) are precisely what is in question here. The fact that we use the word ‘arthritis’ in the attribution is, I claim, no evidence that it is the concept *arthritis* that we are ascribing. All I mean to highlight above is that one might easily get oneself into a position of thinking that since we are attributing to him a thought that is (at least in part) *about* arthritis, then it had better be a content that involves the concept *arthritis*. But this is false, as per my argument in ch. 2. We do not need to ascribe to our subject the concept *arthritis* in order to ascribe to him thoughts about arthritis. Burge might, of course, try again and argue that in order to attribute to our subject any thoughts about “arthritis as such” (cf. Burge, “Postscript to ‘Individualism and the Mental,’” p. 162), then we had better attribute to him the concept *arthritis*. This may be so, but in that case what is in question is precisely whether our subject is thinking thoughts about arthritis “as such.” The fact that he thinks that arthritis may occur in the thigh is precisely what would support our contention that he is not. (This in response to a query from John McDowell.)

³¹ It might seem that Burge can concede this point and still maintain that this is not the belief that he expresses with the words “I have arthritis in my thigh,” and so the existence of this other belief does not touch on the proper interpretation of those words. This will not do. For Burge cannot simply assume that the belief our subject expresses with the words “I have arthritis in my thigh” is the belief yielded by a literal interpretation of those words. This must remain an open question at this point, something that needs to be argued, not merely assumed.

admit that I was wrong in thinking it might have been arthritis. But I will still claim that I was right about something, namely that I have an ailment in my thigh, etc.

The normative question enters here. Will I, *pace* Burge, claim to have been misunderstood? Would my response be disingenuous if I did? The answer to both questions relies, I think, on what *your* response would be. If you were to claim, “no, you are wrong, there is nothing wrong with your thigh, because arthritis cannot occur outside the joints,” then I think I can say, quite legitimately, that you have misunderstood me. You would have misunderstood me, insofar as you would have thought that the only, or at any rate, the primary communicative purpose I had was to inform you that I thought the ailment in my thigh was arthritis. If you, acting as my doctor, refused to take a further look at my thigh to see what was wrong, I would tell you, and rightly so, that you had misunderstood me. On the other hand, if your response was, “well, it can’t be arthritis, but let’s have a look anyway,” then it would be disingenuous, if not to say bizarre, of me to claim that you had misunderstood me because I was thinking with the concept of a disease that could occur in joints and muscles. This shows, I think, that the subject’s response to the correction will vary with the response of his communication partner. In particular, it will vary with whether or not the communication partner correctly comprehends his communicative purpose.

Burge’s quick response from a disposition to self-correction ignores the possibility that I might self-correct primarily in order to more effectively communicate to you what I still maintain is a true object-level belief. Whether this is so will largely depend on your response; more specifically, whether you understand me as having communicated a true object-level belief at all. Thus, Burge’s claim is false: in many of these cases, our subject will maintain a true object-level belief throughout the self-correction. Whether he *explicitly* maintains such a belief will depend on what his interlocutor initially took him to be saying. In short, the evidence on offer does not systematically favor Burge’s view over its competitors.

The reader might protest that I have yet to deal properly with the argument from normativity here. The fact, for instance, that we hold the actual speaker as being in error, but not the counterfactual speaker, shows, according to the protest, that we hold them to be thinking with different concepts and thereby to be thinking different thoughts. The normativity is baked into the concept itself, and the fact that the actual speaker is willing to correct himself once notified of his mistake is confirmation that it is indeed the concept *arthritis* that he is thinking with.

A fuller treatment of this topic will have to wait until ch. 5. All I have aimed to accomplish here is to put in relief Burge's assumption that the subject's willingness to submit to authority and change his views is evidence that he was thinking with the concept *arthritis* rather than some other concept. As I shall emphasize later, what is uncontroversial here is that our subject comes to learn something new – and recognizes himself as coming to learn something new – about the world, viz., that arthritis can only occur in the joints. To try to capture this bit of learning in terms of the subject coming to improve his mastery of a concept that he already possessed and employed is certainly not obligatory and in many ways quite misleading. By contrast, it seems more natural to say that with the new knowledge comes a new “way of thinking” about arthritis. And so, if a concept is a way of thinking, as Burge himself assumes, then we should describe him as coming to acquire a new concept in and through this process of learning. Thus, Burge is right to observe there are norms in play, and that the subject recognizes the force of these norms and adjusts himself accordingly. But little is illuminated and much obscured by attempting to argue that the norm was somehow inherent in the concept that he was already thinking with, albeit poorly.

3.5.5 His-beliefs-aren't-true-anyway argument

A strong motivation for the reinterpretation strategy is the felt need to attribute true beliefs to subjects wherever possible.³² The basic idea is that attributions of true beliefs provide a better grip on the subject's thinking, and, through that, a better explanatory grip on the subject's actions. Crudely put, we understand someone only if we understand how their words and thoughts could be true; their actions, expressions of a rational point of view on the world.³³

One of Burge's most important objections to reinterpreting the patient's attitudes in terms of true object-level beliefs comes from considering the holistic relations between the problematic attitudes that we are currently debating whether to attribute to him and the comparatively unproblematic (because true) attitudes which we might already have attributed to him, *involving the same terms*. Assume that the subject knows that he has arthritis in his knees and ankles, and that he has long feared the disease will spread. Feeling a new pain one morning, he thinks his fears have come true – his arthritis has spread to his thigh.

Is his belief true? Are his fears realized? It is easy to see how the fate of the reinterpretation strategy might hang in the balance here. Suppose that we do opt for a reinterpretation of our subject's thought such that it is not about arthritis, but about tharthritus, the rheumatoid ailment that may also occur outside the joints. If so, then the current argument suggests that we should also reinterpret his previous thoughts and statements regarding his rheumatism in knees and ankles accordingly.³⁴ Now we would be presenting the subject as thinking a true object-level thought. But

³² Cf. Davidson's appeals to the principle of charity, according to which our hermeneutic endeavors should be guided by an effort to "maximize agreement" (and thereby, presumably, truth) (cf. Davidson, "Radical Interpretation," p. 136). See also Akeel Bilgrami, *Belief & Meaning*, pp. 51-56.

³³ This view is not without its detractors. Thus, for instance, Stephen Stich: "What knowledge [i.e., truth] adds to belief is psychologically irrelevant" (Stich, "Autonomous Psychology and the Belief-Desire Thesis," p. 574). It is important to note, however, that such arguments are often (as they are in Stich's case) coupled with a general skepticism toward so-called 'folk-psychological' forms of explanation (cf. Stich, *From Folk Psychology to Cognitive Science: The Case against Belief*). The anti-individualist certainly does not share this kind of skepticism.

³⁴ Recall the point about retroactive reinterpretation explored in section 3.5.1.

the problem is that our subject is *not* thinking a true thought, because the disease he is suffering from cannot spread to the thigh. Burge writes:

When told that arthritis cannot occur in the thigh, the patient does not decide that his fears were realized, but perhaps that he should not have had those fears. He does not think: Well, my tharthrititis *has* lodged in the thigh; but judging from the fact that what the doctor called ‘arthritis’ cannot occur in the thigh, tharthrititis may not be a single kind of disease; and I suppose I need not worry about the effects of its occurring in various locations, since evidently the tharthrititis in my thigh is physiologically unrelated to the tharthrititis in my joints. There will rarely if ever be an empirical basis for such a description of the subject’s inference. The patient’s behavior (including his reports, or thinkings-out-loud) in this sort of case will normally not indicate any such pattern of inferences at all. But this is the description that the object-level reinterpretation method appears to recommend. (Burge, “Individualism and the Mental,” p. 124)

This strikes me as a much stronger objection than any of the foregoing, in large part because it succeeds in touching on the central motivation behind the reinterpretation strategy.³⁵ This argument reminds us that in trying to preserve truth and rationality, we cannot consider each belief attribution in isolation. Rather, we must place any belief in a larger network of the subject’s other beliefs.³⁶ Our belief-attributions, then, are constrained not only by truth (as the current objection emphasizes), but by consistency. Taking a consistent line with the subject’s beliefs entails taking his belief that his arthritis has spread to his thigh as false, or else taking him to have been thinking of something else (e.g., tharthrititis) all the while. The subject’s own behavioral dispositions suggest that the former is the correct view to take. After consulting the doctor, the subject takes himself to have previously held a false belief involving the concept *arthritis*. Thus, Burge’s literalism comports better with the evidence than the reinterpretation strategy does.

³⁵ Which is *not* the idea that we can only think with concepts we completely master. The whole idea of ‘concept mastery’ in Burge’s sense is quite beside the point.

³⁶ This, incidentally, is a point that I will be exploiting, *against* Burge, in ch. 4.

Before going on to evaluate this objection, we must note that it should not be understood as a strike against the idea that true beliefs have a privilege, however circumscribed, in attitude attribution. In fact, the objection implicitly confirms this view. The view was never that every belief we attribute to the subject must be true. Holistic considerations weigh strongly here. The truth-norm pertains in the first instance to bodies of beliefs. There is no particular problem in attributing a single false belief, so long as that belief finds a place in a network of other, mostly true beliefs. Burge's objection to the reinterpretation strategy here relies on this fact. We can confidently ascribe a false belief because there is a network of other true beliefs that we can confidently ascribe to him – pertinently, that he believes he suffers from arthritis in wrists and ankles – and among which this false belief finds its place.

This observation suggests the relevance of a larger theme which goes largely unexplored in Burge's writings, namely the importance of background beliefs in determining what someone is thinking on occasion. Burge conveniently deploys this idea here, but it is not easy to see how such factors can enter into the determination of mental contents on the anti-individualist view.³⁷ In particular, it is difficult to see how such factors can be weighted against other factors such as causal relations or conventional linguistic norms on a given occasion. My argument in ch. 4 is to a large extent designed to show that giving due weight to the subject's background beliefs is inherently problematic for the anti-individualist, and that his theory suffers as a result.

Let us turn to assessing the substance of Burge's objection. The objection turns on a certain need for constancy and consistency in interpretation. It turns on the need to give a univocal interpretation of the term 'arthritis' as it occurs in the subject's initial formulation of his fear and the

³⁷ All in all, Burge's stance on holistic tendencies in philosophy of mind and language is a complex one. His resistance to the idea is by no means as fervent as that of say, Jerry Fodor (cf. *Psychosemantics; Concepts: Where Cognitive Science went Wrong*). Nonetheless, it seems that anti-individualism depends on delimiting the scope of holism. Local holism is good; global holism is not. Thus, for instance: "Holism has been enormously overplayed in philosophical accounts of meaning and reference" (Burge, "Psychology and the Environment: Reply to Chomsky," p. 452).

term ‘arthritis’ as it figures in his subsequent relief that it cannot be arthritis after all. Otherwise, we cannot make sense of his relief, because his relief must precisely be a relief at discovering that his previously held belief (*I have arthritis in my thigh*) was false.

However, if this is the argument, it contains a simple oversight. The attributions occur at different times. Importantly, they flank the time at which the subject came to acquire an important new piece of knowledge, namely that arthritis cannot occur outside the joints. Burge advocates that we see this event simply in terms of an uptake of new factual information, and that we should construe the subject as thinking with a single concept throughout. (And, moreover, that this concept had better be *arthritis*.) But there are no decisive theoretical reasons why we should choose to portray this as a matter of thinking in terms of a single concept throughout, but achieving a gradually better grasp of that concept, rather than present him as thinking of the ailment in his thigh using a different concept once he learns that arthritis cannot occur outside the joints. Concepts, after all, are “ways of thinking about things,” and what our subject acquires when he learns that arthritis cannot occur outside the joints can easily be described precisely as a new “way of thinking” about that disease.

We have already seen one theoretical consideration that *would* decisively favor the former over the latter, namely the idea that we could only portray communication between our patient and his doctor by assuming that they share concepts. My argument in chapter 2 was designed to dismantle this idea. Now we are catching a glimpse of a second kind of consideration that would favor Burge’s view, namely the idea that we could only present the continuity of the subject’s own thinking by having him think with a single concept throughout, of which he achieves a better grasp over time. I shall have more to say about this idea below.

In any case, it is probably true that the subject will express relief upon learning that it is not arthritis which he is currently feeling in his thigh. On the other hand, he is still likely to express

continued concern about the pain in his thigh. He should by no means feel compelled to give up his claim to have a true belief about *that*, just because his initial layman's diagnosis turned out to be incorrect. So far as Burge has given us any reason to think, the relief as well as the continued concern could be equally well expressed using one concept as using the other.³⁸

By way of wrapping up, we should return to Burge's claim that the literal interpretation comports better than its rivals with the evidence on offer. Burge has now offered us two perspectives on this claim: one was a disposition on the subject's part to self-correction; the other the assumption that the subject would express relief upon learning that arthritis cannot occur outside the joints, thereby implicitly confirming the literal interpretation of that belief as a false belief involving the concept *arthritis*, not as a belief involving some other, idiosyncratic concept, say, *tharthritis*. The supposition that either of these could constitute decisive evidence in favor of one interpretation over the other is testimony to Burge's operative assumption that there is a philosophically significant distinction to be drawn between presenting someone as thinking with an incompletely understood public concept and presenting him as thinking with a slightly different idiosyncratic concept (or between presenting someone as thinking with a single concept over time but changing his beliefs about that concept and presenting him as changing his concept with the belief), and so, that these must be attributions of different beliefs. Burge has given us no reason to think that this is the case. So far as we are concerned, then, these could be construed as different descriptions of the same belief. Which description is appropriate relative to central concerns such as assessment of the subject's rationality and explanations of his actions is still an open question. In chapter 4, I will present an argument that

³⁸ One objection might still seem untouched by these arguments, namely that the reinterpretation strategy would have us render as true beliefs which are in fact false. I will address this in ch. 5.

militates against construing the subject as thinking with the incompletely understood public concepts.³⁹

Meanwhile, one might find it tempting to think that there must still be some fact of the matter here, and that the considerations I have levied here merely trace out an epistemic problem of some kind. As Burge sometimes invokes the subject's inferential dispositions to settle the matter, we might think, perhaps as some part of a general assumption about first-person authority, that the subject himself can know in a way that bystanders cannot whether he is changing his belief with respect to a concept or changing his concept with the belief. I think this is wrong. Our ordinary first-personal way of thinking simply does not feature any explicit thematization of concepts. It is even doubtful that more sophisticated kinds of thinking – for instance, what Burge explores under the heading of 'critical reasoning'⁴⁰ – does so. Critical reasoning typically involves thinking about one's justification for thinking p rather than q , or what believing that p entails and rules out. It does not turn on thematizing the conceptual basis of one's thinking. Nor is there any reason why it should. Thematization of the conceptual basis of thought belongs essentially in the sphere of interpersonal comparison. Speaking in terms of diverging concepts is a way of coming to grips with the fact that different rational beings may come to think differently about the same thing, even in assessing the same body of evidence.⁴¹ And even here, concept-talk will only be a conveniently abbreviated way of talking about the different persons' beliefs.

³⁹ My claim here – that so far as Burge has given us any reason to think, these are not attributions of different beliefs, but simply different descriptions of a single belief – may seem to fly in the face of a Fregean orthodoxy that Burge's anti-individualism supports itself on, according to which thoughts are individuated at the level of sense (i.e., concepts). The short answer is that the impression would be false, at least on a proper construal of 'sense,' viz. a construal that connects it decisively to the subject's cognitive perspective. When I enter the qualification "so far as Burge has given us any reason to think," I mean to point out, precisely, that Burge has so far not given us any reason to think that these are in fact different concepts, rather than merely different descriptions of the same "way of thinking."

⁴⁰ I will explore self-knowledge and the idea of critical reasoning further in chapter 4.

⁴¹ As in, "ah, I get it, by 'contract' you mean only written agreements."

Should we assume that our subject is in a better position than the rest of us to assess whether he is now thinking with a different concept or with the same concept only better understood? I doubt that this question points us in the direction of a significant distinction at all. One thing is clear: the subject comes to learn more about how things stand in the world. We have different ways of describing this development, but as yet no argument as to why we should feel compelled to choose between them. No insight is gained by saying that the continuity-through-change in the subject's thinking is anchored in the continued deployment of a (public) concept. It is more intuitive, and philosophically more perspicuous, to say that the continuity of a subject's thinking is anchored directly in the worldly objects that his thoughts are about. One motivation for maintaining the opposite view is the assumption that we could only account for the possibility of communication between subject's with different beliefs if we assumed that they were thinking with the same concepts. I have dealt with this idea in ch. 2. Another motivation for this view is that it promises to preserve an important sense of normativity: it tells us something about why the subject was *right* to change his beliefs in the way that he did. I will have things to say about this motivation in ch. 5.

3.6 Metalinguistic reinterpretation

Burge now turns to what he calls 'metalinguistic arguments for reinterpretation.' Initially, there would seem to be little here that connects with the kinds of concerns we have been exploring so far. Linguistic and metalinguistic matters certainly do matter to our inquiry. But Burge presents the metalinguistic strategy as an attempt to account for our subject's error "as purely a metalinguistic mistake" (Burge, "Individualism and the Mental," p. 125). I fully agree that this is a non-starter.

Attempting to account for our subject's error purely in metalinguistic terms would require arguing that whenever someone employs a word in violation of that word's conventional meaning we should construe the subject as really thinking about the meaning of the word, *as opposed to* thinking about the worldly phenomenon the word ordinarily refers to. I agree with Burge that this is a grotesquely implausible assumption (although I am curious to know who Burge might think would offer such an argument).

However implausible such a purely metalinguistic strategy would be, it should not stop us from exploring the explanatory power of 'non-pure' metalinguistic strategies. For as Burge concedes, metalinguistic notions are ordinarily indissolubly intertwined with ordinary object-level belief attributions. Pertinently, if we interpret our subject's utterance "I have arthritis in my thigh" as expressing the proposition *I have arthritis in my thigh*, then we will also in the same motion stand prepared to attribute to him the metalinguistic belief that the word 'arthritis' applies also to ailments outside the joints. However, Burge claims, "[a]ccepting these metalinguistic attributions, of course, does nothing *per se* toward making plausible a denial that the subjects in our examples have the counterpart object-level attitudes" (Burge, "Individualism and the Mental," p. 125). This I agree with fully. It would be quite wildly implausible to claim that *all* our thoughts are about the words we use to express those thoughts, *as opposed to* being about the usual worldly objects that we take ourselves to be talking about. Thus, it would be *ad hoc* to claim that it is only the beliefs involving putatively incompletely understood concepts which are purely about words, rather than about the ordinary worldly objects.

Burge concedes that it is hard to draw a fine distinction between the relative relevance of metalinguistic beliefs and object-level beliefs.⁴² In particular, he notes that it may be hard to "find

⁴² While Burge does not mention it at this point, this is pretty much what one would expect from anyone who sympathizes with Quine's rejection of the analytic-synthetic distinction. See Quine, "Two Dogmas of Empiricism."

evidential grounds for attributing an object-level attitude *as opposed* to its metalinguistic counterpart,” because “when a person holds one attitude, he often holds the other” (Burge, “Individualism and the Mental,” p. 126). Nonetheless, he claims that there are systematic philosophical reasons weighing in for a bias in favor of an object-level construal of the patient’s thought. In particular, this holds in the case of ascriptions of occurrent beliefs, as opposed to standing attitudes. Burge comments: “if the reasoning is not concerned with linguistic issues in any informal or antecedently plausible sense, attribution of an object-level thought content is supported by the relevant evidence, and metalinguistic attribution is not” (ibid., p. 126). Further,

there appears to be a general presumption that a person is reasoning at the object-level, other things being equal. The basis for this presumption is that metalinguistic reasoning requires a certain sort of self-consciousness about one’s words and social institutions. This sort of sophistication emerged rather late in human history. (Cf. any history of linguistics.) Semantical notions were a product of this sophistication. (ibid., p. 126).

I note with approval Burge’s claim that we should, *ceteris paribus*, think of people as reasoning at the object level. This is the basis of my argument in chapter 2. Nonetheless, while the conclusion is right, the argument that Burge offers for it is not obviously relevant. Even granting that metalinguistic thinking requires a level of mental sophistication, why should we not attribute it, once the sophistication is in place? Note here that I am not saying that metalinguistic considerations are in any way more important than object-level considerations. My claim is rather that it is fruitless to deny that the two operate as a pair in attitude attribution, even when what is at stake is occurrent

Burge declares his allegiance to this rejection in “Individualism and the Mental,” p. 117, “Two Thought Experiments Reviewed,” p. 190, “Intellectual Norms and the Foundations of Mind,” pp. 256-257, “Wherein is Language Social?” p. 290#14. Whether Burge’s reliance on a notion such as ‘conceptual error’ is *ipso facto* a reliance on a notion of analytic truth is the subject of an interesting exchange between Åsa Wikforss and Sarah Sawyer. See Wikforss, “Social Externalism and Conceptual Errors,” Sawyer, “Conceptual Errors and Social Externalism,” Wikforss, “Externalism and Incomplete Understanding.”

beliefs. If we ascribe to someone the occurrent belief that his arthritis has spread to his thigh on the basis of his uttering the words “I have arthritis in my thigh,” then we had also better stand prepared to ascribe to him the belief that the word ‘arthritis’ could apply also to ailments in the thigh. Of course, it need not be properly a part of the content of his occurrent thought about the pain in his thigh that the word ‘arthritis’ applies to it. Nonetheless, if our subject is rational, this will be a relevant part of a true description of his thought contents. I fail to see why the relative cognitive sophistication required to sustain such thinking should be a concern here. If we ascribe to our subject any occurrent thought about arthritis (where ‘arthritis’ occurs obliquely), then we are, *ipso facto*, ascribing to him the relevant degree of mental sophistication. Whether he is self-consciously exercising that sophistication on the occasion is beside the point.

At this point, we should ask why this is a concern to Burge at all. Everybody agrees that the pure metalinguistic strategy is a non-starter. But why should Burge feel the need to go on to attempt to drive a wedge between attributions of object-level attitudes and attributions of metalinguistic attitudes? I suspect the following: conceding that the speaker’s metalinguistic attitudes are relevant at all will tend, in these problematic cases, to erode our confidence in the literal construal of his words. Once we start emphasizing the fact that the subject quite plainly does not know the meaning of the word ‘arthritis,’ it will tend to become less and less clear why we should portray him as thinking with the concept which is conventionally associated with that word.

This touches on a difficulty in understanding and assessing Burge’s general argument. Many commentators assume that anti-individualism entails a direct line from the *words* a speaker uses on occasion (on their conventional interpretation) to our attribution of mental content: whenever a speaker uses words with a certain meaning in his language, we are to ascribe to him a mental state in

accordance with the meaning of those words.⁴³ Against, this Burge can rightly protest, first, that his claim was never that strong, and secondly, that this reading gets the focus of his argument all wrong. His argument is fundamentally not about words, but about concepts. If this distinction is sustainable, then Burge can simply shrug his shoulders at any metalinguistic considerations we might introduce. Whatever metalinguistic beliefs a subject might have (i.e., whatever beliefs about the meanings of words), these may fail to be relevant to determining which concept that person is thinking with. There is no such tight connection between the conceptual basis of our thinking and the words we use to express ourselves.⁴⁴

But this line of thought seems to run up against a fairly salient aspect of our manner of communicating and understanding each other. For we frequently do use metalinguistic inquiries as a way of getting to understand precisely what someone is thinking. If my interlocutor says something that puzzles me, then my asking him “what do you mean by *x*?” is typically a first step towards a better grasp of what he is thinking. And if grasping what someone is thinking is, as Burge suggests, at least partly a matter of uncovering the conceptual basis of their thought, then such metalinguistic inquiries would be a reliable way of uncovering the concepts that person is thinking with. In short (and this is the valuable point that the metalinguistic strategy could bring to the table, if managed

⁴³ Cf. Christopher Gauker, “Social Externalism and Linguistic Communication,” p. 1, Donald Davidson, “Knowing One’s Own Mind,” p. 28. But certainly Burge himself must shoulder much of the blame for the misapprehension. Consider claims like the following: “wherever the subject has attained a certain competence in large relevant parts of his language and has (implicitly) assumed a certain general commitment or responsibility to the communal conventions governing the language’s symbols, the expressions the subject uses take on a certain inertia in determining attributions of mental content to him” (Burge, “Individualism and the Mental,” p. 147; “By varying the *uses of words* in his linguistic community, I found that the content of his propositional attitudes varied” (Burge, “Other Bodies,” p. 84).

⁴⁴ This line of thought is taken to its extreme in Burge’s essay “Intellectual Norms and the Foundations of Mind.” Here he distinguishes between the ‘cognitive value’ of a term and its ‘conventional meaning,’ concluding that there never was “any obvious reason why the two notions should coincide” (Burge, “Intellectual Norms and the Foundations of Mind” (p. 271). I raise the question of whether this line is compatible with the argument of “Individualism and the Mental” in chapter 2, Addendum. My assessment of the situation is that this is indeed the line that we must ultimately take if we are to be conventionalists about linguistic meaning. What this shows, however, is why we should not be conventionalists about linguistic meaning, not that we should dissociate semantic content and psychological content.

correctly), further knowledge of which metalinguistic beliefs a speaker holds is typically relevant to ascertaining which belief he is giving expression to on occasion.

In connection with assessing the metalinguistic strategy, Burge briefly considers a related issue, which will concern us at greater length in ch. 4. The question is whether “what a person says and how he behaves [...] infallibly determine what his attitude contents are” (Burge, “Individualism and the Mental,” p. 127). Burge willingly acknowledges that there are cases in which action and utterance do not determine attitude. However, this point “is often invoked in a sloppy and irresponsible manner. It is incumbent on someone making it (and applying it to cases like ours) to indicate considerations that override the linguistic and behavioral evidence” (Burge, “Individualism and the Mental,” p. 127).

This question is awkwardly put, and it is not immediately easy to see what is at stake in it. Asking whether what one says and does *determines* what one thinks seems like grabbing the wrong end of the stick. A more natural way of asking the question is in terms of whether what one says and does infallibly *reflects* what one thinks. Assuming that this is the question Burge has in mind, the answer is clearly ‘no.’ To be sure, we expect a high degree of consonance between our utterances, our actions, and our thoughts. Nonetheless, this basic requirement of rational consistency is not so strict as to rule out the possibility of the occasional slip. Simple examples are easy to come by: I decide to push the green button, but due to carelessness or haste I end up pushing the red button instead. My action did not properly reflect what I was thinking.

But this rather pedestrian observation is not the real issue here. Burge considers this objection in the following form: assume that this is what the subject is saying; should we now take what he is saying to be the proper expression of what he is thinking? Put in those terms, Burge is right that we must present strong evidence in order to overrule the assumption that what he is

saying is what he is thinking. But, of course, the question at stake between Burge and us is not whether what he is saying is what he is thinking, but rather whether what the anti-individualist is taking him to be saying (and thinking) really is what he is saying (and thinking). Here Burge is guilty of sneaking in illicit assumptions on behalf of the literalist strategy. For the various so-called ‘reinterpretation strategies’ that Burge is considering simply do not cede the interpretation of the speaker’s *words* to the literalist strategy, so that philosophically pregnant questions could arise about the relationship between the subject’s words and his thoughts (for which the reinterpretivist would opt for a non-literal construal).⁴⁵ On the contrary, if we find that we cannot ascribe to the subject a thought in accordance with the literal interpretation of his words, then we are finding that the literal interpretation of his words is not the right one. We are not arguing that what he is thinking differs from what he is saying.

In other words, the reinterpretation strategy fully shares Burge’s concern that we should expect a certain kind of continuity between what a person says and how he thinks and behaves. So far I have been focusing mostly on the relationship between words and thoughts. Burge’s argument seems to lean programmatically in favor of taking the words at face value and ascribing attitude content accordingly. The various reinterpretation strategies we have been considering all seem to advocate that the ordinary bias in favor of literal interpretation be overruled in certain local cases. Here, some assumption of cognitive cogency – for instance that which is encapsulated in the principle of charity – is invoked to bolster the claim that we should not understand the speaker’s

⁴⁵ It might be useful here to draw a comparison with debates concerning the distinction between semantics and pragmatics in philosophy of language, or between semantic content and something like ‘speech act content’ (cf. Cappelen and Lepore, *Insensitive Semantics*). I take it that Burge would agree with me that this distinction is of no help here. Of course, we could speak of the conventional meaning of the words uttered in isolation from any question of what the speaker might have meant by them, and so construct a technical notion of semantic content on that basis. But insofar as this is the first step precisely toward distinguishing such semantic content from speech act content, this move could no longer serve Burge’s purposes, since he is looking for a single notion of content that could work for both. The question that concerns us, then, is whether there is any good argument in favor of taking the (conventionally determined) semantic content of an utterance to also yield the speech act content in cases where the subject is plainly ignorant of (has false beliefs about) the conventions that determine this content.

words in the manner recommended by the literalist strategy. Burge maintains that the burden of evidence lies on the reinterpretation strategy.

What kind of evidence is it that the reinterpretation strategy can claim on its behalf? This question will be subject to much more detailed discussion in chapter 4 of this dissertation, but it behooves me to say something about it here. One of the main purposes of the attribution of attitudes is to provide the raw materials for the explanation of action. As Burge himself emphasizes, in explaining behavior, we seek to portray the subject's actions as flowing rationally from his way of viewing the world.⁴⁶ It is this way of viewing the world – what I have called the subject's cognitive perspective – that we seek to capture with the invocation of concepts. As we saw in chapter 2, it is Burge's contention that even as the actual and the counterfactual subject are thinking of the same worldly item, they are doing so armed with different concepts, and so are viewing this state of affairs from different cognitive perspectives. This difference in cognitive perspective yields a difference in their thought contents, and so, one might think, should matter to the way we explain their respective actions.

But here lies the puzzle. Recall that we are discussing Burge's question of whether “what a person says and how he behaves [...] infallibly determine what his attitude contents are.” Burge concedes, as he must, that they do not *infallibly* determine attitude contents – allowance must be made for foreigners, children, and even the occasional slip-up by otherwise competent speakers. But exceptions aside, it is a signal virtue of anti-individualism that it preserves the tight, systematic

⁴⁶ A familiar Davidsonian story lurks in the background here. Many balk at what they see as an overly instrumentalistic ring to Davidson's take on attitude attribution – i.e., that the ‘whole point’ of attitude attribution is action explanation (cf. Moran, “Interpretation Theory and the First Person”). Burge does not seem to be among these. In “Other Bodies,” he writes that “it seems unexceptionable to claim that the obliquely occurring expressions in propositional attitude attributions are critical for characterizing a given person's mental state. Such occurrences are the stuff of which explanations of his actions and assessments of his rationality are made” (p. 84); “We thus describe his perspective on his environment and utilize such descriptions in predicting, explaining, and assessing the rationality and the correctness of his mental processes and overt acts. These enterprises of explanation and assessment provide much of the point of attributing propositional attitudes” (p. 92).

connections between what a person says, what he does, and what attitudes we should ascribe to him. Of course, it is my claim that these are precisely the connections that anti-individualism cannot preserve; in fact, this is precisely where anti-individualism does significantly worse than the reinterpretivist strategy. First, as we just saw, Burge cannot without begging the question equate “what the speaker said” with what is yielded by an interpretation in accordance with the conventional semantic values of the terms that he used. Second, it seems that if we were to follow Burge’s conventionalism in the cases that we are considering, what we would lose would be precisely the ability to use those attributions – our understanding of what the subject said – to explain his behavior. Indeed, it seems most peculiar that Burge should attempt to claim this as a strength of his theory, when it so clearly is one of its weaknesses. Consider, for instance, the following:

[I]o a fair degree,⁴⁷ mentalistic attributions rests not on the subject’s having mastered the contents of the attribution, and *not on his having behavioral dispositions peculiarly relevant to those contents*, but on his having responsibility to communal conventions governing, and conceptions associated with, symbols that he is disposed to use. (Burge, “Individualism and the Mental,” p. 148, my emphasis)

It should be easy to see how this is problematic for anti-individualism. Actions are explained, in the basic instance, if they can be seen to flow rationally from the agent’s beliefs. The patient tells his doctor “I think I have arthritis.” The doctor attributes to him the belief *I have arthritis*, and proceeds to ask, “Where?” The patient gestures to his wrists. There is a clear sense in which the patient’s action is rationalized by the foregoing attitude-attribution, and thereby also serves as corroboration of its correctness. But imagine instead that the patient were to gesture toward his thigh. There is a corresponding sense in which this action is *not* rationalized by the foregoing attitude-attribution. The subject does not, as Burge puts it, have “behavioral dispositions peculiarly relevant to” the thought

⁴⁷ And so, presumably, in the cases we are considering here.

contents we are attributing to him. Yet this should not, according to Burge, shake our belief that we are attributing to him the correct psychological attitudes. It follows that Burge cannot consistently claim as a virtue of anti-individualism that it preserves a tight, systematic connection between what a person says, what he thinks, and how he behaves.

3.7 Philosophical arguments for reinterpretation

Having dispensed with these four proposals for methods of reinterpretation, Burge turns to considering three philosophically motivated arguments for the same.⁴⁸ These are:

- i) An argument that “the content clauses we ascribed must be reinterpreted so as to make reference to words because they clearly concern linguistic matters – or are about language” (Burge, “Individualism and the Mental,” p. 129).
- ii) An argument from charity, according to which “we should not attribute to rational people beliefs like the belief that one may have arthritis in the thigh” (ibid., p. 129)
- iii) An argument according to which “we should not attribute contents involving incompletely understood notions because *the individual must mean something different by the misunderstood notion than what we non-deviant speakers mean by it*” (Burge, “Individualism and the Mental,” p. 130).⁴⁹

We shall consider Burge’s counter-arguments to each of these in turn.

⁴⁸ Burge speaks of these as relying “purely on intuitive or *a priori* considerations” (Burge, “Individualism and the Mental,” p. 128-129).

⁴⁹ Burge declares that he finds argument iii) to be “perhaps the most interesting” (Burge, “Individualism and the Mental,” p. 130).

Argument i) is clearly a close cousin, if not an identical twin, of the metalinguistic reinterpretation strategy which we considered in the previous section, and many of the same considerations apply. Burge holds that “[e]ven if this argument were sound, it would not affect the thought experiment decisively” (Burge, “Individualism and the Mental,” p. 129). He agrees that some of the relevant attitudes may be linguistic in the sense that their truth is tested by recourse to dictionary entries rather than ordinary empirical states of affairs. Nonetheless, “it is simply a mistake to think that these facts entail, or even suggest, that the relevant contents are metalinguistic in form” (ibid., p. 129). I have already spoken my piece on the metalinguistic construal of content (section 3.6), and will not pursue the matter further here.

Burge is brief also with argument ii), the argument from charity. He points out that the argument fails to “touch most of the attitudes that may launch the thought experiment; for many are straightforwardly true, or false on ordinary empirical grounds” (ibid., p. 129). Furthermore, “there is nothing irrational or stupid about the linguistic or conceptual errors we attribute to our subjects. The errors are perfectly understandable as results of linguistic misinformation” (ibid., p. 129). In fact, arguing from considerations of charity in this context begs the question against anti-individualism: “A belief that arthritis may occur in the thigh appears to be inexplicable or uncharitably attributed only if it is assumed that the subject must fully understand the notions in his attitude contents” (ibid., p. 129). That is to say, it is only if we insist that possessing a concept requires knowing the definition of that concept⁵⁰ that a literal interpretation of “I have arthritis in my thigh” would constitute an attribution of logical idiocy. But as Burge argues, in order to make this objection stick, we would first have to show that this requirement is appropriate. If we can do that, then we will already have defeated anti-individualism, and we will have no need of this further argument.⁵¹

⁵⁰ I.e., what I above called the knowledge requirement for concept possession.

⁵¹ See section 3.2 for more.

In response, we note first that it is not in dispute that our subject thinks that arthritis may occur outside the joints. So far, however, this is a purely *de re* specification of his thought content. What *is* in dispute is how to characterize the subject's particular conceptual angle – his cognitive perspective – on that thought content. Again Burge appears to be assuming that in order for us to ascribe to him any thought about arthritis at all (even a false thought), we need to ascribe to him the concept *arthritis*. On this view, making reference to – thinking about – arthritis is an accomplishment that requires a philosophical explanation. The only such explanation we would seem to have is that the subject possesses and deploys the concept *arthritis* on the occasion. On the other hand, it does not matter whether he completely understands the concept. This first of these claims is false, as my argument in chapter 2 shows.⁵² What is at stake here is the second claim, according to which we can think with concepts that we fail to grasp in systematic ways.

Second, Burge writes that raising charity-related concerns in this context is misguided, because there is nothing “stupid or irrational” about the mistakes we attribute to the subject. Indeed, such mistakes can be easily understood as “results of linguistic misinformation.” However, this is clearly not enough to get Burge what he needs. An appeal to linguistic misinformation takes us no further than to the kind of metalinguistic error which Burge himself has explicitly dismissed as not quite relevant. We are still no nearer to understanding in what sense our subject might still be said to be thinking with the concept that he incompletely understands. Differently put, allowing incomplete concept mastery to be explained or otherwise illuminated by linguistic misinformation would amount to bringing words and concepts much closer together than I think Burge's program can be comfortable with.⁵³

⁵² Moreover, and as I also suggest in ch. 2, even if the view were true, there would be legitimate questions to be asked about its explanatory power. In other words, one senses that if the view is true, it is trivially true.

⁵³ Certainly, Burge's later statements regarding an overemphasis on linguistic matters in his earlier arguments for anti-individualism seem to support this point (cf. “Postscript to ‘Individualism and the Mental,’” pp. 157-162, “Postscript to

Third, Burge is right to point out that we would not typically raise charity-related concerns in cases where the subject might not fully understand the concept we ascribe to him, but in which the attributed belief is true or merely a matter of an empirical mistake.⁵⁴ But what does this show? I assume that we are supposed to feel a concern about consistency here: it would be inconsistent of us to object on grounds of charity in the arthritis-case, if we are not also prepared to do so in these other cases. But this concern with consistency is quite misplaced. In fact, it can be readily seen that Burge is misrepresenting the motivation behind the argument. Our concern is not *per se* with the attribution of contents involving concepts that the subject cannot recount the definition of. As I pointed out in chapter 2, I think there are many cases in which we unhesitatingly and quite justifiably attribute contents involving concepts which we make no assumption that the subject even possesses, let alone that the subject fully masters.⁵⁵ Rather, what we are saying is that in contexts where the cognitive divergence matters,⁵⁶ then it would *prima facie* appear to be in conflict with charity to ascribe an attitude which involves a blatant misunderstanding of that concept. We suspect this in no small part because we have yet to find an enlightening statement of what it might mean to say that someone is thinking with a concept that he does not fully understand. To say that we are not attributing a self-contradiction and thereby violating charity, because we need not presuppose that

'Belief *de re*,' pp. 65-66, "Phenomenality and Reference: Reply to Loar," p. 449n3, "Reply to Christopher Gauker," pp. 246, 250). Nonetheless, even in recent years Burge has occasionally seemed prone to conflate the two, e.g.: "a primary impetus for my discovering the thought experiments was *recognizing how many words or concepts I went around using which I found, on pressing myself, that I did not fully understand*" (Burge, "Postscript to 'Individualism and the Mental,'" p. 175, my emphases). Kent Bach catches Burge in the making of a similar conflation in "Individualism and the Mental," and perceptively observes: "Burge describes his well-known 'arthritis' example as involving 'incomplete understanding of concepts' or 'conceptual errors,' even though it is more aptly described as involving incomplete understanding of words or linguistic errors. Indeed, sometimes he seems to equate the two, as when he says that 'the error is linguistic or conceptual' [...] though surely these are distinct. Burge offers no positive account of what it is to think with a concept that one incompletely understands. As for me, it is clear how one can use a word one incompletely understands, but I have no idea what it is to think with a concept that one incompletely understands, for I have no idea what it is to understand a concept over and above possessing it" (Bach, "Burge's New Thought Experiment: Back to the Drawing Room," p. 88n3).

⁵⁴ Which is not to say that there are not many attributions of ordinary empirical mistakes which should also trigger concerns about charity.

⁵⁵ An example that I exploit in chapter 2 is attributions of beliefs in inter-linguistic contexts. See section 2.6 for more.

⁵⁶ Cf. the point about relevance in ch. 3.5.1.

the subject has full mastery of the concept *arthritis* and thereby knows that arthritis cannot occur outside joints, does not help. What matters is this: the subject has all his thoughts about arthritis, some true, some false. If all we can say to shed light on the idea of incomplete concept mastery is point to the fact that the subject is linguistically and / or empirically misinformed, then Burge has simply failed to give us any real philosophical grip on what it is to think with a concept one incompletely masters. What this offers us is simply a version of what I above called a ‘non-pure metalinguistic strategy’⁵⁷: to evade the charge that attributions of thoughts involving incompletely mastered concepts violate charity, he himself reverts to attempting to account for the error “in terms of a reinterpreted object-level content together with a metalinguistic error.”⁵⁸

It is argument iii), according to which the individual must mean something different by his word ‘arthritis’ than the rest of us do and should be understood accordingly, that Burge spends most time on trying to refute. Burge reminds us how he understands the burden of proof here: to be successful, such an argument would not only have to establish that the subject “has notions which are not properly expressed in the way he thinks they are.” It would also have to show that the subject does not – indeed, cannot – have the attitude attributed according to the literal interpretation (Burge, “Individualism and the Mental,” p. 130).

Burge suggests two ways in which such an objection might proceed. On a first interpretation of the scenario, we might emphasize the fact that the speaker thought the word he used applied to a different range of objects than it does. This observation is not disputed by Burge. Indeed, it is clearly the starting point for Burge’s own thought experiment. A second interpretation is more ambitious: according to it, the speaker must “have had *in mind* something that the words do not denote or

⁵⁷ See section 3.6.

⁵⁸ Burge, “Individualism and the Mental,” p. 122.

express” (Burge, “Individualism and the Mental,” p. 130). I take it that the difference between the two interpretations is the following: the mere observation that the speaker is mistaken about the meaning of his words is no problem to Burge. Therefore, the second interpretation attempts to take this observation further: the speaker’s mistaken thinking that the word applies not just to this, but to that as well, shows that he had something else “in mind”; i.e., he was thinking with a different concept than (most of) the rest of us do when we use the word ‘arthritis.’⁵⁹

Burge brushes aside this argument in characteristic fashion. The argument, he says, is a *non sequitur*:

It does not follow from the assumption that the subject thought that a word means something that it does not [...] that the word cannot be used in literally describing his mental contents. It does not follow from the assumption that a person has in mind something that a word does not denote or express that the word cannot occur obliquely (and be interpreted literally) in that-clauses providing the content. (Burge, “Individualism and the Mental,” p. 130)

Burge concedes that there might be cases in which reinterpretation along these lines would be warranted.⁶⁰ However, “the present argument needs to show that deviant speaker meaning always forces such reinterpretation” (*ibid.*, p. 130).⁶¹

Thus, rather than dispute the observations and ideas underlying the objection, Burge now attempts to appropriate them for his own purposes. It follows from the way in which his thought experiment is set up that the speaker thinks the word applies to something it does not. Thus, this

⁵⁹ Thus, there is a clear sense in which the dialectic between methodological objections i) and iii) is reproduced here.

⁶⁰ Examples would include the utterances of children and of foreigners just learning the language, as well as phenomena like slips of the tongue and malapropisms. See Burge, “Individualism and the Mental,” pp. 118-121.

⁶¹ We note in passing Burge’s view of the structure of the argument: we who object to anti-individualism must be operating with a universal claim – it is never right to ascribe thought contents including incompletely understood concepts in oblique position. Burge can then establish anti-individualism with an existential claim: it is sometimes justifiable to ascribe thought contents with incompletely understood concepts in oblique position. This testifies to the fact that Burge thinks what is at stake is the question of incompletely understood concepts as such. I think the situation is much more complex than that, but will not pursue the matter here.

observation cannot be held against it. Moreover, the more demanding idea that the speaker may have had some other concept “in mind” does not entail that he does *not* (also) have beliefs ascribable using the concept which he incompletely understands. And since the burden of argument was such that the objection had not only to establish that he has one kind of thought, but also that he does *not* have the other kind of thought, the objection fails.

Burge again invokes the subject’s tendency to self-correction to bolster his argument:

The subject’s willingness to submit his statement and belief to the arbitration of an authority suggests a willingness to have his words taken in the normal way – regardless of mistaken associations with the word. Typically, the subject will regard recourse to a dictionary, and to the rest of us, as at once a check on his usage and his belief. When the verdict goes against him, he will not usually plead that we have simply misunderstood his views. This sort of behavior suggests that [...] we can say that in a sense our man meant by ‘arthritis’ *arthritis* – where ‘*arthritis*’ occurs, of course, obliquely. We can say this despite the fact that his incomplete understanding leads us, in one of the senses explicated earlier, to say that he meant something different by ‘arthritis’. (Burge, “Individualism and the Mental,” p. 131)

Let me start with the point about self-correction. I have dealt extensively with an argument of this general kind earlier in this chapter, and here it will suffice to merely repeat the main outlines of my conclusion. The fact that we frequently display a disposition toward self-correction in these kinds of cases, and that we will admit to being caught in some kind of error, is not in dispute. What is in dispute is what this fact shows. The subject comes to acquire new knowledge about arthritis, and comes to recognize himself as having previously held a false belief about the disease. To say, in addition to this, that he also thereby passed to a new level of mastery of a *concept* that he already possessed and deployed is non-obligatory and in many ways misleading.

Burge writes that the subject intends (presumably at the time of utterance) that his words be “taken in the normal way,” i.e., literally.⁶² But of course he does – anything else would be exceedingly eccentric.⁶³ However, we do not – and certainly an anti-individualist like Burge should not – hold that it is up to the individual to decide what his words mean, by simply *intending* for them to mean this or that. The relation between intention and utterance meaning is much more complex than that. Speaker’s intention is certainly an important factor, but the problem here is that the speaker’s performance appears precisely to preclude our taking his words in quite the normal way.⁶⁴ We have two ways to inform our subject that he failed in executing his intention: either tell him that arthritis does not occur outside the joints, or tell him that the word ‘arthritis’ does not apply to ailments outside the joints. It does not matter which we choose: if he correctly appropriates the one, he will also appropriate the other.⁶⁵

Burge appears to be presuming that the speaker has an opinion, or can at least be called upon to form an opinion, as to whether he is thinking with this concept or that. Thus, upon consulting with authority, he will consent to our suggestion that he was thinking with the concept *arthritis*, even as he had a confused grasp of that concept. I think this assumption is at best

⁶² See also Burge, “Social Anti-Individualism, Objective Reference,” p. 309.

⁶³ Except, perhaps, in the familiar cases of poetic metaphor, and so on.

⁶⁴ Thus, I am in no way operating with what semanticists sometimes refer to as a distinction between wide and narrow context, i.e., a distinction between interpretations that do and those that do not take speaker’s intention into account (cf. Francois Recanati, *Literal Meaning*, p. 56). Quite the contrary. My point is rather the familiar one that sometimes our actions (including our speech acts) are botched, and we do not succeed in executing our intentions. Burge says that the speaker intended his words to be taken in the normal way. This is probably true. But he also intended to refer to a disease that could occur in the muscles as well as the joints. These two intentions cannot be jointly satisfied with the word ‘arthritis.’ Our question, then, is which of these two component intentions should guide our hermeneutic endeavor? We note that my interpretation seems in better keeping with Burge’s injunction that we should, *ceteris paribus*, treat people as thinking and reasoning at the object level (see section 3.6, “Individualism and the Mental,” p. 126). Here Burge’s argument attempts to sway us precisely by appeal to considerations regarding the subject’s metalinguistic intentions (his intending that his words be taken in the normal way). My interpretation, by contrast, is based firmly on considerations at the object level (his intending to pick out a member of this rather than that range of objects).

⁶⁵ Indeed, why even bother to bring concepts into the reckoning here? I suppose that Burge will argue that we are, as a matter of fact, already dealing with concepts here. I take the subject’s word ‘arthritis’ to refer to arthritis. I am thereby already taking him to be in possession of the concept *arthritis*. Again, this does not follow: I also take the Norwegian speaker’s word ‘gikt’ to refer to arthritis, but do not thereby make any supposition about him having that concept.

misleading. As I have argued, concept-talk has no real purchase on the subject's own thinking about his thinking. He is simply not thinking of himself as thinking with any given concept. He is thinking that the world is a certain way, viz. that he has arthritis in his thigh. It follows, given minimal suppositions, that he thinks arthritis can occur outside the joints. If concepts are "ways of thinking about objects," as we have been assuming, then we should conclude that our subject is manifestly *not* thinking with the concept *arthritis*; he is thinking with the concept of a rheumatoid ailment that can occur outside the joints, and he is thinking that it is arthritis. About the latter, he is wrong, and he accepts this upon consulting with authority. But his correcting his own error does not mean that he will admit that he was thinking with the concept *arthritis*. I imagine that he will simply look quizzically at us if we were to ask him this question.⁶⁶ He was wrong to think that arthritis could occur outside the joints. What could we mean by asking further whether the erroneous thought contained a standard, public concept non-standardly understood, or a non-standard concept?

Moving on, Burge claims that even if he concedes that the subject had something else "in mind," it does not follow that we could not present him as also having attitudes involving the incompletely understood concepts. Burge can concede this – and apparently does concede this – because he thinks that the reinterpretation strategy needs not only establish that he has this other thought, but also that he does *not* have the thought involving *arthritis*. I take it that this "something else" which the subject has "in mind" is something that Burge will concede that we could appropriately describe as a variant concept (e.g., *tharthritis*). Support for this interpretation can be found in the following passage:

⁶⁶ I imagine Burge responding in the following way: "you say that 'he is thinking with the concept of a rheumatoid disease that can occur in joint and muscle, and he is thinking that it is arthritis.' This last clause shows that he is thinking with the concept *arthritis* after all." Suppressed premise: one could only make reference to or think about arthritis using the concept *arthritis*. As I have shown in chapter 2, this is false, and the objection therefore unpersuasive.

There are cases in which it is reasonable to say that, at least in a sense, a person has a notion [i.e., concept] that is expressed by dispositions to classify things in a certain way – even if there is no conventional term in the person’s repertoire that neatly corresponds to that ‘way.’ [...] On the other hand, the fact that such attributions are justifiable *per se* yields no reason to deny that the subject (also) has object-level attitudes whose contents involve the relevant incompletely understood notion. (Burge, “Individualism and the Mental,” p. 125)

When conjoined with the concession that the speaker might have had something else “in mind,” this passage gives us what we need to complete our argument.⁶⁷ Burge has all the while been maintaining that choice of concept is crucial to thought-individuation, because that is the level at which thoughts are individuated: *I have arthritis in my thigh* and *I have tharthrititis in my thigh* are simply not the same attitude-content.⁶⁸ Our subject utters the words “I have arthritis in my thigh,” and we are asking which thought he is giving expression to. Burge is now arguing that establishing that he is expressing the one thought does not rule out that he is (also) expressing the other. This could mean one of two things: either he is expressing two thoughts at the same time, or it does not matter which of the two thoughts we choose to think of him as expressing. The first option is surely not sustainable,⁶⁹ whereas the second option implicitly reneges on Burge’s official line and suggests the following diagnosis: it does not matter which thought we portray him as thinking, for they are attributions of the same thought. Thus, the argument meant to establish anti-individualism is demonstrably unsound: not only are the actual and the counterfactual subject not, in any philosophically relevant

⁶⁷ See also “Individualism and the Mental,” pp. 122-123: “in order to overturn the thought experiment, these methods must not only establish that the subject held the particular attitudes that they advocate attributing; they must also justify a *denial* of the ordinary attributions literally interpreted.”

⁶⁸ Cf. the conclusion offered to establish anti-individualism at the thought-experiment’s step 3.

⁶⁹ Bear in mind, again, that Burge cannot here plausibly locate the two contents on either side of a divide between, say, semantics and pragmatics. Our question is what thought the subject is giving expression to, not what his words mean according to the conventions of his language. Burge’s startling claim is that we should determine which thought the subject expressed by interpreting in accordance with these linguistic conventions even in cases where the subject clearly demonstrates his lack of knowledge of them. Thus, there is a sense in which the debate between Burge and me mimics the debates concerning the relative importance of semantic content and pragmatic content. But Burge and I both agree that there is only one kind of content that is relevant to our present concerns. This is the assumption that seems jeopardized by Burge’s flirtations with the two-content strategy.

sense, incapable of thinking the same thoughts; so far as we can tell at this point, they are actually thinking the same thought on the occasion under consideration.

As I will argue in chapter 4, however, this diagnosis is only partially correct. These attributions are indeed rival forms of attribution – different descriptions – of a single thought content. But it *does* matter which description we choose. *Why* and *how* it matters is something that we can only see when we turn to considering our attributions in the wider context of trying to make sense of the agent as engaged in reasoning and intentional action. Once the wider context is brought to bear, we will see, as Burge rightly points out, that only one of these attributions can succeed in capturing the subject’s cognitive perspective. But we will also see that the attribution that does capture the subject’s cognitive perspective is not the one that involves the public, conventional concept, but rather the attribution that involves the variant concept.⁷⁰

⁷⁰ This is also how I read the message of “Intellectual Norms and the Foundations of Mind” (see ch. 2, Addendum, for a detailed discussion). Burge’s contention in “Individualism and the Mental” that it is “common practice” (and presumably correct) to attribute thoughts involving incompletely understood concepts is put in relief by the passage in “Postscript to ‘Belief *de re*’” (p. 80) which I also discussed in that section, according to which “what is attributed will often be less specific or fine-grained than the actual way of thinking engaged in by the individual thinker to whom the attitude is attributed. This is so *even for correct oblique occurrences in propositional attitude attribution whose purpose is to describe propositional attitudes as fully as ordinary communicational conditions will allow*. What is easily and conventionally attributed as a way of thinking may be unexceptionable as far as it goes. Yet it may fall short of capturing the individual’s way of thinking.” The Burge of “Individualism and the Mental” reads “common practice” correctly as regards the attribution of thoughts involving incompletely understood *words*, but wrongly assumes that common practice here presumes itself to be laying bare the conceptual basis of the attributed thoughts.

4

COGNITIVE PERSPECTIVE AND SELF-KNOWLEDGE

4.1 Introductory

In chapter 3, I mounted an argument against anti-individualism and the doctrine of incomplete concept mastery the conclusion of which may have seemed overly irenic: that an anti-individualist theory of thought attributions and the theory that I favor can both yield true attributions on some occasion. It falls to this chapter to show that we nonetheless have strong reasons to favor my theory. The conclusion of chapter 3 – that there is a sense in which attributions of thoughts containing incompletely understood concepts can be true – will still stand. Nonetheless, we will find that such attributions are inappropriate. They are inappropriate because they will hinder us in the execution of many of the key tasks that we need such attributions for, viz. assessing the character and force of a person’s reasoning and explaining his actions. If we cannot use our attributions in these tasks, then these attributions do not yield *understanding*, however much they may yield true descriptions of thought contents.¹ As I shall say in this chapter, such descriptions will not succeed in capturing the

¹ I have in mind an analogy with Davidson’s famous point concerning the individuation and description of actions (derived ultimately from Anscombe’s *Intention*, §§ 19, 23). Davidson’s idea is that actions are multiply describable, even as not all true descriptions bear in the right fashion on the subject’s intentions, and so have the right kind of explanatory force. In Davidson’s famous example “flipping the switch,” “turning on the light,” “illuminating the room,” and “alerting the prowler to the fact that I am home” may be four true descriptions of my action (Davidson, “Actions, Reasons, and Causes,” pp. 4-5). Even as all of these may be (extensionally) true descriptions, however, it may be that not all of them *explain* the action in the right kind of way (and are thus, one would want to say, *intensionally appropriate*). A description of an action is extensionally true but intensionally inappropriate if it cannot enter into a rationalization of that action. The basic idea is that a description cannot enter into rationalization of my action if it crucially cites factors of

subject's *cognitive perspective*. This is relevant, since it is precisely Burge's contention that attributions of thoughts containing incompletely understood concepts may succeed in capturing the subject's cognitive perspective, and that this is why we should favor them over rival attributions. To show that these attributions will systematically fail in capturing the subject's cognitive perspective is to show that Burge's anti-individualism fails – even by its own lights – to supply a viable theory of thought attribution.

The cognitive perspective metaphor is notoriously elusive, but nonetheless fundamental to any theory of thought attribution. The underlying idea can be spelled out roughly as follows: we must seek to capture not just *what* the subject is thinking about – what worldly object, say – but also *how* she thinks about what she is thinking about. Grasping the subject's cognitive perspective bears on the task of gaining the right, rich sense of understanding of our subject: the sense of understanding we have when we can not only describe that person's doings in the world, but we can also claim to understand *why* she thinks and acts as she does; how thinking and acting like that could flow rationally from a certain view of how things are.

Consider a small example: say I tell you that “Peter thinks the murderer should walk free.” It could be that I am saying something true here, if the person we are talking about is indeed guilty of murder. But this can hardly be how Peter is thinking about it. Peter precisely thinks that this person is not guilty of murder, which is why he thinks he should walk free. Thus, by using the term ‘the murderer’ in my attribution of this thought to Peter, I have failed to convey to you how Peter's position may be a rational position to take. For it *is* the rational position to take if you believe, as Peter does, that this person is innocent. That this is how Peter thinks about it is something I failed to remark on in my report of his belief, and which indeed my use of the term ‘the murderer’ did

which I am incognizant. These reflections, while developed in the context of a theory of action explanation, bear quite directly on the topic of thought attributions, and suggest one way in which the theories of linguistic ascription, psychological attribution, and action explanation are intimately related.

everything to cover up. In this sense, we should say that my attribution – though true – may have failed to capture Peter’s cognitive perspective on the matter. I may have picked out *what* Peter is thinking about – namely that person over there – but I have failed to properly indicate *how* Peter is thinking about what is he thinking about, namely not as a murderer, but as an innocent man, falsely accused. My attribution, then, yields a true description of Peter’s attitude but fails to convey understanding. The attribution fails to render Peter’s way of thinking and acting intelligible.

The requirement that our attributions capture the subject’s cognitive perspective bears in an obvious way on our ability to use those attributions to form a picture of the subject as a rational agent engaged in intentional action. Connoisseurs will recognize this requirement as a variation on Frege’s insistence that we distinguish between *Sinn* and *Bedeutung* (which I introduced in chapter 1) and on the theme of *de dicto* and *de re* modes of thought attribution (which formed a centerpiece of my argument in chapter 2). In what follows, I will provisionally accept Burge’s implicit assumption – ultimately derived from Frege – that the task of capturing the subject’s cognitive perspective can be modeled in terms of the task of finding the right concepts to employ in our attributions. Against this background, I will show that attributions of thoughts containing concepts that the subject incompletely masters will fail to capture the subject’s cognitive perspective in the requisite way.

Much of my argument concerning cognitive perspective will be channeled through a discussion of self-knowledge. The question of the compatibility of self-knowledge and anti-individualism is one of the most hotly debated topics to follow in the wake of Burge’s work. I shall argue that these debates have by and large had the wrong focus. Critics have mainly homed in on a problem that is not unique to Burge’s position at all, but which is generic to any externalist position worthy of the name. This problem, however, is merely apparent, and can be resolved by recourse to a purely deflationary argument, which Burge has done more than any other externalist to explore and defend. However,

there is also another problem of self-knowledge that lurks in this vicinity, which has received much less attention in the critical literature. This problem arises as a consequence of the doctrine of incomplete concept mastery and so is unique to Burge's position. As we shall see, it cannot be resolved by recourse to a purely deflationary argument, because it bears on what I shall call a *substantive* notion of self-knowledge (to be defined further below).

On the few occasions that Burge has addressed this latter, more specific challenge in print, he has been quite dismissive of it, arguing that it rests on theoretical presuppositions that beg the question against anti-individualism. I will show that this dismissive attitude cannot be sustained. What is at stake in substantive self-knowledge is the ability of subjects to engage in critical reasoning over the contents of their thoughts. Burge has done more than any other recent philosopher to showcase the cognitive fundamentality of critical reasoning capabilities and their connection with self-knowledge.² Yet we will find that the doctrine of incomplete concept mastery is inconsistent with seeing critical reasoning capabilities, and the kind of self-knowledge on display in the exercise of such abilities, as placing fundamental constraints on the theory of thought attribution.

This connects the themes of cognitive perspective and self-knowledge in a relatively straightforward way. It is in critical reasoning, in the exercise of substantive self-knowledge, that the subject's cognitive perspective takes shape. In this sense, the failure of anti-individualist thought-attributions to capture the subject's cognitive perspective, on the one hand, and their failure to display the subject in his exercise of substantive self-knowledge, on the other, are merely two manifestations of a single theoretical problem, a problem which stems from the doctrine of incomplete concept mastery.

² Cf. Burge, "Our Entitlement to Self-Knowledge," "Reason and the First Person," "A Century of Deflation and a Moment of Self-Knowledge."

In my view, an exploration of the task of capturing the subject's cognitive perspective and an exploration of the requirement that attributions be consistent with self-knowledge can thus be jointly illuminating. In catchword form, my view concerning cognitive perspective is that we should view it as a *holistic* notion: we can only get a real sense of a subject's cognitive perspective if we consider a larger set of her beliefs and the dynamic interactions between these beliefs, that is, the inferential connections she draws between them. We cannot, I argue, gain a well-founded sense of a subject's cognitive perspective by looking at her beliefs one by one, in isolation from all the rest. Consider a simple and intuitive argument in favor of this view: we said that capturing the subject's cognitive perspective is connected to action explanation. The standard philosophical picture of action explanation is that it proceeds through a reconstruction of a piece of practical reasoning.³ But practical reasoning essentially involves stringing together several beliefs in the right, rational pattern, namely a pattern that shores up a justification or a motivation for the action in question. What we fundamentally want, I argue, is not just to draw up a list of which beliefs some agent has, but to understand and appreciate the role of these beliefs in her larger cognitive economy. In a turn of phrase that will become central to my case, the aim of capturing the subject's cognitive perspective is the aim of *understanding what*, from the agent's point of view, *is a reason for what*; what, in her view, justifies her acting in this way rather than that, in maintaining this attitude rather than that. But such justifications essentially span over multiple beliefs and attitudes.

Similarly for self-knowledge: we should view it as a notion that essentially involves several beliefs and the relations between them. The substantive notion of self-knowledge that I will argue for involves being able to say something about what follows from what one is thinking, to say *in other words* what one is thinking, or what, by thinking this, one is committing oneself to. Such self-knowledge is what we exercise when we reflect on what justifies us in acting this way rather than

³ Cf. Davidson, "Actions, Reasons, and Causes," Anscombe, *Intention*.

that, in maintaining this attitude rather than that. And just as I remarked above, such justifications cannot but range over multiple beliefs and attitudes.

If both of these core claims come across as crashing truisms, I sympathize. Thus, much of my argument will be directed toward showing that Burgean anti-individualism puts both ideas in jeopardy. Burgean anti-individualism, and specifically, its theoretical centerpiece, the doctrine of incomplete concept mastery, makes philosophical claims which presume the plausibility of construing both cognitive perspective and self-knowledge as notions that need range over no more than single beliefs at a time. Once we see what is wrong with anti-individualism as a theory of thought attribution, we shall see what is wrong with this atomistic way of construing these notions.

4.2 Burgean anti-individualism and cognitive perspective introduced

As we have seen, anti-individualism is typically introduced by way of a characteristic kind of thought-experiment, a thought-experiment which pivots on a speaker making a conceptual error of a certain kind. Thus, in the most famous of these thought-experiments, we encounter a person who goes to his doctor and says “I have arthritis in my thigh.” Recall that we are supposed to understand here that the error here is conceptual rather than empirical, because ‘arthritis’ simply means *rheumatoid ailment of the joints*, and so could not, by definition, occur in the thigh. Burge’s point is that this glaring conceptual error nonetheless need not deter us from giving a literal interpretation of the word ‘arthritis’ as it figures in his utterance, and thereby say that it is indeed the concept *arthritis* that our subject is thinking with. This, in a nutshell, is the doctrine of incomplete concept mastery. The fact that our subject is prone to misapply some concept – and not due to carelessness, but because he actively maintains the belief that the concept applies to things it does not – should not deter us from saying that he still possesses that concept and even that he is employing that very concept on

the occasion of the misapplication.⁴ Moreover, Burge holds that such incomplete concept mastery is not merely a marginal phenomenon in our cognitive lives, such that it would be of primarily theoretical interest. Quite the contrary, he holds that it is “common or even normal in the case of a large number of expressions in our vocabulary” (Burge, “Individualism and the Mental,” p. 112).⁵

On Burge’s view, concept and cognitive perspective are closely related notions. Approaching some matter from one cognitive perspective rather than another can be represented as a matter of thinking with this concept rather than that. Thus, to capture the subject’s cognitive perspective is, on this view, to find the right concept to use in the attribution. If it is – as Burge claims – the concept *arthritis* that our subject is thinking with, then that is the concept that captures his cognitive perspective. The central anti-individualist claim, then, is that a certain concept can capture the subject’s cognitive perspective on some matter even in a case where the subject does not fully “master” the concept in question; indeed, even in a case where he is systematically disposed to misapply that concept.

I have said that capturing the subject’s cognitive perspective is connected to our ability to use these attributions to illuminate the subject’s pattern of reasoning and to explain his actions. Burge fully concurs with this assessment:

⁴ One motivation for this view is our apparent need to say that the patient and his doctor may in other circumstances be successful in communicating thoughts about arthritis. Our ability to model such communication may seem to presuppose that they share the concept *arthritis*, even though the patient’s grasp of this concept is clearly a bit shaky. In chapter 2 of this dissertation, I have argued that the appearance of a need to say these things is simply an illusion. Crudely put: we have no need to see the doctor and the patient as sharing the concept *arthritis* in order for us to be able to see them as sharing beliefs about arthritis (the disease). Fundamentally, thinking about arthritis does not presuppose having the concept *arthritis*; nor does communicating – sharing – beliefs about arthritis presuppose that the interlocutors share that concept.

⁵ See also the recent Postscript to “Individualism and the Mental” (p. 175): “A primary impetus for my discovering the thought experiments was my discovering how many words or concepts I went around using which I found, on pressing myself, that I did not fully understand. I came to realize that this was not just a personal weakness. It was part of the human condition, at least in complex societies.”

Clearly, oblique occurrences in mentalistic discourse have something to do with characterizing a person's epistemic perspective - how things seem to him, or in an informal sense, how they are represented to him. ("Individualism and the Mental," pp. 103-104.)

If there were not deep and important connections between propositional attitudes and understanding, one could hardly expect attributions of mental content to facilitate reliable predictions of what a person will say, do, or think. ("Individualism and the Mental," p. 118.)

[T]he obliquely occurring expressions in propositional attitude attributions are critical for characterizing a given person's mental state. Such occurrences are the stuff of which explanations of his actions and assessments of his rationality are made. ("Other Bodies," p. 84)

The point of such attributions is to characterize a subject's mental states and events in such a way as to take into account the *way* he views or thinks about objects in his environment. We thus describe his perspective on his environment and utilize such descriptions in predicting, explaining, and assessing the rationality and the correctness or success of his mental processes and overt acts. These enterprises of explanation and assessment provide much of the point of attributing propositional attitudes. ("Other Bodies," p. 92)

According to Burge, then, concepts which the subject incompletely understands can still convey essential information about "the way he thinks about these things," can still capture how things "seem" or "are represented to him," they can play an essential role in explanations of the subject's actions and in assessments of his rationality. These are all claims that I will try to undermine in this chapter.

I will start by noting a small but important reservation concerning the methodology of these anti-individualist thought experiments. While the reservation is of a general kind, it is especially relevant to our present concerns, and will become important as we go on. Approaching the question of thought attribution from the vantage point provided by such thought experiments will tend to

distort our view of what is at stake. In a nutshell, these thought experiments implicitly encourage an atomistic approach to thought attribution, quite at odds with the holistic approach I sketched out above. Here is what I have in mind: I think everyone will admit that the thought-experiments are schematic (as indeed all thought-experiments must be). They present us with a brief description of some salient features of a social and physical environment, and a set of words passing from a subject's lips. We are to ask: what thought is this person giving expression to? In this sense, the thought-experiments ask us to consider thought-attributions one by one, thereby implicitly assuming that we have some way of determinately saying what some person is thinking on such minimal information. In my view, the way to approach Burge's question "what thought is this person giving expression to?" is neither to affirm nor deny Burge's proposal, but to say, simply, "I don't know; tell me more." The request for more information would be an expression of the kind of view that guides me here, namely that an attribution of a thought does not yield understanding until we know something about the role of that thought in the subject's larger cognitive economy. The mere fact that the subject has uttered words with such and such semantic values conventionally attached to them is, of course, a relevant datum. But the only thing that could entitle us to say that the literalist interpretation is the *right* interpretation is to see how an attribution based on the literal interpretation would fare with respect to the tasks of illuminating the subject's pattern of reasoning and explaining his actions. As I will try to show, in cases of what Burge calls incomplete concept mastery, the literal interpretation will fail with precisely these tasks. But these problems are obscured by the methodology of Burge's thought experiments, which ask us to consider attributions of thoughts one by one.

In this sense, Burge's claim that attributions of thoughts containing incompletely understood concepts can serve to capture the subject's cognitive perspective has a sheen of plausibility only against the background of these atomistic assumptions artificially imposed through the methodology

of the thought-experiments. Since the thought-experiments ask us to consider attributions one by one, accepting the challenge they present to us can lead us to think that the cognitive perspective that we need to capture is somehow inherent in each thought, considered in isolation from all the others.

At first glance, it might seem as if the idea that cognitive perspective is inherent in each thought individually can support itself on a prestigious philosophical pedigree, namely the Fregean distinction between *Sinn* and *Bedeutung* that I mentioned earlier. Frege's famous puzzles are resolved by finding the right concept – the *Sinn* – with which to describe the subject's cognitive perspective on the object that he is thinking about – the *Bedeutung*. This certainly seems like a matter that can be explored on a case by case, thought by thought, basis.

Exactly how to approach this idea is a delicate matter, and Frege-scholarship is not my task here. I will nonetheless cautiously suggest the following: this interpretation, with its implicit atomism, cannot capture the Fregean approach at its strongest or most plausible. Briefly and somewhat metaphorically, we should view the concepts that figure in Frege's theory as nodes in vast inferential networks. A concept is defined by its role or location in this network. To claim that *this* concept as it figures in *this* thought captures the subject's cognitive perspective in a way that *that* concept would not is to presume the background of this inferential network. We would be working, then, precisely on the idealizing presumption that the subject to whom we are attributing the concept in question *knows his way around* the inferential network that defines that concept. But this is exactly the assumption that anti-individualism's doctrine of incomplete concept mastery urges us to let go of. On the anti-individualist view, a person can be said to be thinking with some concept even though he "incompletely masters" it. This concept can capture his cognitive perspective – can characterize his way of thinking – even though he does not know his way around the inferential

network that defines the concept, indeed, even though he has false beliefs about the inferences that defines the concept.⁶

In my view, this lacks plausibility. The claim that anti-individualist attributions can serve to capture the subject's cognitive perspective has a sheen of plausibility only because the thought-experiments artificially constrain us to consider such attributions one by one. This apparent plausibility crumbles when we widen our focus, as I think any serious theory of thought attributions must. My claim is that attributions of thoughts containing incompletely understood concepts will tend to frustrate our ability to capture the rational dynamic interactions between the subject's many beliefs, in other words, our ability to see that subject as engaged in reasoning. But it is only here, in the subject's pattern of reasoning, that we could plausibly seek to anchor our picture of his cognitive perspective, not in any one belief that we are considering attributing to him. Thus, Burge's claim that anti-individualist thought attributions can serve to capture the subject's cognitive perspective cannot be sustained upon critical scrutiny.

4.3 The generic problem of externalism and self-knowledge

So far, I have sketched and motivated my initial claim that we should view cognitive perspective as a holistic notion, anchored in a subject's pattern of reasoning, not in concepts or beliefs one by one. I

⁶ There is much confusion about these matters in the current literature, partly, I believe, as a result of attempts to accommodate Burge's arguments. For instance, even as careful a thinker as Christopher Peacocke, in his much cited book *A Study of Concepts*, will say first, in a programmatic introductory section, that "[t]he view that a thinker's incomplete understanding of a word does not prevent true ascriptions of propositional attitudes using that word in oblique occurrences has been extensively defended by Tyler Burge. I take his arguments as compelling." (p. 29). However, when time comes around for a more detailed analysis, he writes as follows: "A possession condition for a particular concept specifies a role that individuates the concept. The possession condition will mention the role of the concept in certain transitions that the thinker is willing to make." (p. 107). It should be clear that these two claims are in tension, unless one subsequently goes on to say exactly which inferential transitions are the ones that a thinker must be willing to make, if he is to be rightly said to possess the concept. Peacocke could avert this problem by saying "mastery condition" rather "possession condition." But that would leave him instead with the problem of saying in what sense someone who does not fulfill the mastery condition nonetheless can possess and be thinking with the concept in question.

now want to turn our attention to the problem of self-knowledge. At first glance, this may seem an unrelated matter. On the contrary, I think reflecting on self-knowledge will allow us to make significant headway on the matter of cognitive perspective.

Burge's writings on self-knowledge stake out two distinct approaches to the problem. The first approach appears to take the form of a rather traditional epistemological investigation concerning my warrant in framing a certain narrowly specified range of judgments, namely self-ascriptions of occurrent thoughts: e.g., 'I think, with this very thought, that water slakes thirst.' Burge calls these "*cogito*-judgments," and argues that the product of such judgments constitutes "basic self-knowledge."⁷ The reflexive character of these judgments guarantees that we cannot be wrong in making them: they are, in a word, *self-verifying*. One could not misidentify the content of one's occurrent thought in such a case, because the content is "thought and thought about in the same mental act" (Burge, "Individualism and Self-Knowledge," p. 116).

The second approach takes a more circuitous route, and constitutes, in effect, a transcendental argument.⁸ This argument starts from an observation concerning the crucial role of self-knowledge in what Burge calls *critical reasoning*. Crudely put, if we did not know what we were thinking, we could not engage in critical reasoning (or for that matter, in any reasoning at all). The argument then goes on to note that we are, in fact, critical reasoners. From these two premises, we can then draw the conclusion that we must possess the relevant kind of self-knowledge.⁹ If this second argument is successful, it will have secured a philosophical foundation for self-knowledge on which can rest a broader class of judgments than the first approach was capable of supporting.

⁷ Burge, "Our Entitlement to Self-Knowledge," p. 240.

⁸ Thus, one might be tempted to assume that the difference between the two approaches mirrors that between a broadly Cartesian and a broadly Kantian approach to the epistemology of self-knowledge. This interpretation receives some measure of support from Burge's own characterization of the project. Nonetheless, we should also not underestimate the Kantian overtones to the argument from *cogito*-thoughts. I have in mind, of course, Kant's famous claim in the B-deduction that "It must be possible for the 'I think' to accompany all my representations" (*Critique of Pure Reason*, B131).

⁹ Burge, "Our Entitlement to Self-Knowledge," p. 246.

In Burge's presentation, these arguments outline two sources of our entitlement to self-knowledge.¹⁰ We have reason to ask, however, whether two such different arguments can really succeed in triangulating on a single notion of self-knowledge, that is, whether the sense of self-knowledge that figures in each argument is the same. The fact that Burge does not even consider this question suggests that he certainly thinks they do.¹¹ It will be central to my case that they do not. I will argue that the two arguments speak to two radically different notions of self-knowledge, which I will call the *deflationary* and the *substantive* notion, respectively.¹² I will say much more about the difference between these two notions after we have reviewed both of Burge's arguments concerning self-knowledge in more detail. We will find that Burge's two arguments respond to two very different philosophical challenges. The *cogito*-argument is a response to what has come to be called "the incompatibilist challenge," which raises a problem that is not unique to Burge's position at all, but which I shall call the "generic problem of externalism and self-knowledge." By contrast, the argument from critical reasoning has the right shape for an answer to a challenge concerning what I have called substantive self-knowledge. What we shall ask is whether this is an answer that Burge is in fact entitled to, given his other philosophical commitments. I will argue that the doctrine of incomplete concept mastery prevents us precisely from seeing the agent as engaged in critical reasoning in the right kind of way, and so prevents us from portraying him as exercising the substantive kind of self-knowledge.

¹⁰ See Burge, "Our Entitlement to Self-Knowledge," p. 245. "Entitlement" here is a semi-technical notion introduced by Burge. It will be reviewed in more detail below.

¹¹ To my knowledge, the short passage in "Our Entitlement to Self-Knowledge" from which I just quoted is the only place where he discusses the two arguments side by side.

¹² I note that a similar terminology is employed, although in the context of drawing different morals than I will do here, in Andre Gallois, *The World Without, The Mind Within: An Essay on First-Person Authority*, ch. 9.

4.3.1 The challenge of incompatibilism

According to the standard picture of self-knowledge, rational human adults possess an unmediated (non-inferential), non-empirical, and close to infallible knowledge of their own occurrent thoughts. Such knowledge, moreover, is taken to be an essential aspect of rational cognition and agency. Yet, by tying the contents of a subject's thought directly to contingent states of affairs in the external world, the externalist program would quickly arouse the suspicion that its espousal would be incompatible, at least on occasion, with the exercise of self-knowledge, so conceived. In brief, if we can only gain knowledge of these external determinants of thought in a posteriori manner – by relying on sense experience – how could we nonetheless come to gain knowledge of the contents of our thoughts in an a priori manner, as common sense, or at least the traditional philosophical picture of self-knowledge, seems to entail?¹³

While in what follows, I will be mostly concerned with this so-called 'achievement'-problem, we may note that in many presentations, it marks only one half of an apparent dilemma, the other half of which is often called the 'consequence'-problem.¹⁴ Burge's anti-individualism accords a decisive role to the causal impact of a speaker's environment in providing content to the speaker's thoughts and utterances. Thus, *water*-thoughts are *water*-thoughts in virtue of certain (causal) relations

¹³ With almost universal scorn, Gilbert Ryle's *The Concept of Mind* (ch. VI) is usually singled out as the only place where a philosopher of note dares experiment with the view that self-knowledge may be perfectly ordinary empirical knowledge. First-person authority, on this view, would be merely the result of the fact that we have more and better empirical evidence to work with in our own case than with any others. See, e.g., Ryle, *The Concept of Mind*, pp. 155-156, 179-181. Donald Davidson's critical citation in "First Person Authority" (pp. 5-6) is what brought Ryle's idea back into the discussion.

¹⁴ The terminology is due to Martin Davies, "Externalism, Architecturalism, and Epistemic Warrant." Michael McKinsey ("Anti-Individualism and Privileged Access") is usually given the nod as the first-in-print with this objection, although it has been explored in details also by a host of other philosophers (notably, Martin Davies (ibid., "The Problem of Armchair Knowledge"), Paul Boghossian ("Content and Self-Knowledge," "What the Externalist can know a priori") Brian McLaughlin ("McKinsey's Challenge, Warrant Transmission, and Skepticism"), Crispin Wright ("Some Reflections on the Acquisition of Warrant by Inference"), Jessica Brown ("The Incompatibility of Anti-Individualism and Privileged Access"), and Gary Ebbs (*Rule-Following and Realism*, ch. 9). Since then the question of the compatibility of anti-individualism and self-knowledge has generated an enormous literature. A first round of papers is collected in Peter Ludlow and Norah Martin, eds., *Externalism and Self-Knowledge*, while Susana Nuccetelli, ed., *New Essays on Semantic Externalism and Self-Knowledge* does much to bring the debate up to date. A useful monograph on the matter (by a writer who has in turn occupied both sides of the debate), is Jessica Brown's *Anti-Individualism and Knowledge*.

obtaining, in the basic instance, between the thinker and the substance water. Hence, it would seem that a speaker's finding himself in a position to frame thoughts about water presupposes that there is (or has been) water in the speaker's physical environment to provide the relevant causal input. Much as the patient in the linguistic community in which rheumatoid ailments are classified differently than in ours cannot be said to think with our concept *arthritis*,¹⁵ it seems that a subject in a world in which there is no water, and who has no compatriots who have had any causal interaction with water, cannot be said to think with the concept *water*. Following this train of thought, we can see that if a subject is to be in a position of knowledgeably claiming a priori that she is thinking about water, then some apparently contingent empirical matter of fact, namely the existence of water in that subject's environment at some point in time, must obtain. The anti-individualist is then faced with a dilemma:

either

(i) allow that the exercise of self-knowledge (knowing that one is thinking about water) can also provide non-empirical knowledge of contingent states of affairs (knowledge that water does exist or has existed),

or

(ii) give up the idea that we can possess self-knowledge independently of particular bits of empirical knowledge.¹⁶

¹⁵ See section 1.3, as well as chapter 2, for an extended critical discussion.

¹⁶ It is important to note that it is the dependence of self-knowledge on *particular* bits of empirical knowledge that is in question. Few current authors would be willing to claim that self-knowledge should be independent of any and all empirical knowledge whatsoever. It is, I think, one of the most interesting and innovative aspects of Burge's epistemological theory that he has systematically sought to dissociate the question of whether a given epistemic warrant is *a priori* from whether it is a warrant that could be had in the absence of any empirical knowledge at all. On this, see e.g., Burge, "Content Preservation," pp. 459-460.

Thus, by tying our thoughts directly to objects in the world (rather than, say, to mere mental representations of objects in the world), it seems to tie our *knowledge of our thoughts* to our *knowledge of worldly objects*. If we suppose that we could only gain knowledge about these external determinants of thought in an a posteriori manner, by relying on sense experience, we then face the problem of how could we nonetheless know the contents of our thoughts in an a priori manner. If we insist that we can have such a priori knowledge of our own thoughts under externalism, then we are faced with the problem of explaining why the subject does not now possess a new route to a priori knowledge of worldly states of affairs. In short, it seems that if externalism is correct, then the two kinds of knowledge would seem to have to be of the same epistemic kind, no matter which kind that will turn out to be.

A common way of approaching the consequence problem is in terms of the transmission of epistemic warrant: supposing that a subject has a priori knowledge of the fact that she is thinking about water, and knows, having read and understood the arguments of Putnam and Burge, that she would be in a position to think about water only in an environment that contains or has contained water, is she not now in a position to draw the conclusion that she lives in an environment that contains or has contained water?¹⁷ And would not the result of this inference constitute a priori knowledge, insofar as it is deduced from two premises which are likewise assumed to be known a priori?¹⁸

¹⁷ We note in passing also another attempt at defusing this problem. According to this line of thought, what our subject could not know a priori is that ‘water’ denotes a natural kind and so is subject to Putnam’s argument. See, for instance Susana Nuccetelli, “Knowing that one knows what one is talking about” and Burge, “Reply to Martin Davies,” pp. 263-264. This response, however, seems quite inadequate to address the unique challenges raised by Burge’s *social* externalism. According to Burge, the thought experiment could run on “an artifact term, an ordinary natural kind word, a color adjective, a social role term, a term for a historical style, an abstract noun, an action verb, a physical movement verb, or any of various other sorts of words” (Burge, “Individualism and the Mental,” pp. 106-107).

¹⁸ I mention this because it has seemed to several authors that finding a non-*ad hoc* rationale for blocking the transmission of warrant in these cases would be a promising way of defanging the incompatibilist challenge. See, for example, Crispin Wright (“Cogency and Question-Begging: Some Reflections on McKinsey’s Paradox and Putnam’s Proof,” “Some Reflections on the Acquisition of Warrant by Inference”) and Martin Davies (“Externalism,

4.3.2 Resolving the achievement problem

Burge presents the incompatibilist challenge as highlighting the difficulty of “understanding how we can know some of our mental events in a direct, non-empirical manner, when those events depend for their identities on our relation to the environment” (“Individualism and Self-Knowledge,” p. 650). In brief, if externalism is correct, then self-knowledge and knowledge of contingent matters of fact in the external world seem like they would have to be of the same epistemic kind, whichever kind – a priori or a posteriori – we take that to be. Burge’s strategy is to show that this is not so: even assuming the truth of externalism, we can dissociate the two forms of knowledge: “one can know what one’s mental events are and yet not know the relevant general facts about the conditions for individuating those events” (“Individualism and Self-Knowledge,” p. 651). More specifically:

To know that water exists, or that one is touching water, one cannot circumvent empirical procedures. But to *think* that water is a liquid, one need not *know* the complex conditions that must obtain if one is to think that thought. [...] It is enough that they [these conditions] actually be satisfied. (Burge, “Individualism and Self-Knowledge,” p. 654)

Very roughly, then, the strategy is this: while it is true that possessing knowledge of any worldly x requires empirical investigation, one’s knowing that one is thinking about x does not. Self-knowledge, thus understood, requires merely that the relevant conditions obtain as a matter of fact, not that this matter of fact be antecedently known to the thinking agent. The error in thinking otherwise, Burge will go on to say, is to hamper externalism with a peculiarly internalist conception

Architecturalism, and Epistemic Warrant,” “Externalism and A Priori Warrant”). It is important to note that this is *not* Burge’s line. See Burge, “Reply to Martin Davies,” p. 261. See also Jessica Brown, *Anti-individualism and Knowledge*, ch. 7, for an extended review of these arguments.

of knowledge.¹⁹ As we shall explore in more detail shortly, Burge here moves to draw a distinction between two kinds of epistemic warrant: an externalist notion of warrant – entitlement – which a subject can enjoy even in cases where she cannot cite the conditions that enable her to claim this knowledge, and an internalist notion of warrant – justification – which requires that she can also claim knowledge about these enabling conditions.²⁰

4.3.3 Slow-switching

For an argument toward the conclusion that knowledge of one's own occurrent thoughts does not require knowledge of the conditions that make thinking the thought in question possible, Burge examines the apparent problem of *slow-switching*. Slow-switching thought-experiments are closely related to the thought-experiments familiar from “Individualism and the Mental,” with one important difference: in the original experiments we consider numerically distinct individuals in different possible environments, while slow-switching concerns the verdicts we should draw with regard to the contents of the thought and speech of a single individual being moved between different such environments.

Using the basic set-up of Putnam's twin-earth scenario, we may thus imagine not two distinct individuals existing in two different possible worlds, but one individual being moved from

¹⁹ This is not to say that someone who embraces an externalist position in the philosophies of mind and language must, on pain of inconsistency, embrace what has come to be called an externalist theory of epistemology. The point is merely that the incompatibilist challenge appears to be structured around some distinctly internalist assumptions concerning knowledge and justification. These assumptions can certainly be challenged, and it would seem that someone who embraces an externalist position in the philosophies of mind and language is in a particularly good position to issue just such a challenge. Moreover, the counter-challenge could be issued on general grounds, and so would not constitute an *ad hoc* measure to rescue anti-individualism from problems of self-knowledge.

²⁰ The distinction between entitlement and justification is of a quite general nature, applicable to a large range of philosophical problematics. It might indeed be more instructive to first see it applied to, say, the epistemology of visual perception. Thus, in a normal case of visual perception, I may be *entitled* to claim knowledge that there is a red book in front of me, even though I have not secured many of the elements some philosophers might say would enter into a full *justification* of such a knowledge-claim, viz. that I am not fooled by light conditions, tricked by an evil demon, etc.

earth to twin-earth. Against this background, the problem of self-knowledge is easy to frame: given that we assume (as per the original thought experiment) that the two subjects in the original thought-experiment will be entertaining different thoughts, in spite of the fact (per hypothesis) that nothing cognitively available to either of them would reveal this difference, does it not follow that a single individual, transported without her knowledge from one scenario to the other, will now be thinking different thoughts without her being aware of any change occurring? In other words, may she not ignorantly go on thinking that she is thinking about water, whereas in the new environment she is in fact thinking about XYZ? And would not this misidentification of one's own thoughts amount a failure of self-knowledge?²¹

It would not take one very far here to argue that the causal connections that individuate content evince a certain degree of inertia, so that immediately upon switching, the person would still be properly understood as thinking about water even in an environment where such stuff does not exist. After all, having never previously had any causal interaction with XYZ, she would not be properly understood to have any thoughts about it. Immediately upon switching environments she would still, then, be thinking about water, and rightly thinking of herself as thinking of water, but simply be prone to misidentify a certain substance as water. This is an ordinary empirical mistake, and not at all a threat to self-knowledge.

However, little by little the relevant causal connections would be set up between the subject and the counterpart substance in her new environment. And so, after a suitable period of time she would be properly understood by her new compatriots as talking and thinking about XYZ rather

²¹ This, I should add, is what has come to be called the problem of *ignorant slow-switching*. The question of *knowledgeable slow-switching* poses slightly different problems. In such a scenario, the subject would know that such switches occurred from time to time, and would know the ramifications of such switches, but would not know which environment she currently finds herself in. This is also the home turf of Paul Boghossian's so-called 'memory-argument' ("Content and Self-Knowledge"). Recent contributions to the question of knowledgeable slow-switching can be found in Bernard W. Kobes, "Mental Content and Hot Self-Knowledge," and Tyler Burge, "Mental Agency in Authoritative Self-Knowledge: Reply to Kobes." Burge confronts Boghossian's argument in detail in "Memory and Self-Knowledge."

than water.²² A *slow switch* from one set of thoughts to another would then have occurred. We note again, however, that this slow switch could, per hypothesis, occur without the subject being aware of it. As we have stipulated, there need be no phenomenological tip-off to the difference, and so the subject's failure to suspect a difference has no bearing on the thought experiment. Thus, the original question arises again: would this not be a scenario in which our subject persisted in (mistakenly) thinking that she was thinking about water, whereas she is correctly understood by her new peers as speaking about XYZ?²³

4.3.4 *Cogito*-thoughts and basic self-knowledge

This raises a problem concerning the subject's ability to know her own occurrent thoughts. What we have here is a seeming mismatch between the speech community's identification of the thought content ("She is thinking about ...") and the subject's own identification of that state ("I am thinking about ..."). Since the speech community and the subject identify the thought contents differently, there seems to be a conflict of interpretation at hand.²⁴ And in these cases, Burgean anti-

²² Note the echo of Davidson's infamous Swampman (in "Knowing One's Own Mind"). While Swampman, Davidson thinks, cannot be said to have any thoughts right away, given that he has never been in the appropriate causal relations, nothing prevents Swampman from *subsequently* having all these thoughts and more, once the appropriate conditions are in place (see Davidson, "Knowing One's Own Mind," p. 19n3). For a critical discussion of whether Davidson got the conclusion of the Swampman thought-experiment right (even by his own lights), consult Martin Hahn, "When Swampmen get Arthritis," pp. 43-45.

²³ Here is a view of the same problem from a rather higher altitude: while the traditional internalist picture of the mind ties our thoughts to mental representations of the world, an externalist theory has thoughts reflect directly on things in the world. On the internalist view, these mental representations are conceived in such a manner as to be self-intimating: there could not be a difference between two mental representations that would remain cognitively irretrievable even upon the most thorough investigation. On the other hand, since the contents of these mental representations do not necessarily reflect how things stand in the subject's environment, this model gives rise to a now familiar kind of *external-world skepticism*. The problem now is to present externalism in such a way as not to invite a parallel, but apparently even more damaging, problem of *internal-world skepticism*. This is precisely the charge that we seem to invite when we say that the very contents of our thoughts could change in ways that would remain systematically elusive to us.

²⁴ I note in passing that this conflict need not, of course, be apparent to either party. Both parties employ the same syntactic string, agree to occasions on which this string should be employed, and otherwise agree to the wider range of behavioral dispositions appropriate with respect to the occasions in question.

individualism seems programmatically committed to resolving the dispute in favor of the speech community, thereby seemingly depriving the subject of her grasp of *which thoughts she is currently thinking*.

Not so, says Burge. The solution to our problem comes to us via a more thorough consideration of the relation between object-level thoughts and higher-level thoughts (i.e., between thoughts of the form *I showered in water this morning* and self-attributions of thoughts, like *I am thinking that I showered in water this morning*). The solution is simple and striking: higher-order thoughts inherit their content from the object-level thought.²⁵ Therefore there is simply no room for a semantic slippage between my object-level thoughts and my thoughts about my thoughts that would make for a failure of self-knowledge. In Burge's words, "[o]ne is thinking that *p* in the very event of thinking knowledgeably that one is thinking it" (Burge, "Individualism and Self-Knowledge," p. 654). If I am correctly understood by my peers as thinking that *I showered in XYZ this morning*, then there is no room for puzzling about whether I might mistakenly think that *I am thinking that I showered in water this morning*, because the higher-order thought simply inherits the content of the object-level thought, *whatever that content is*. If the term 'water' denotes water, then what I am thinking about is water, whether at the object-level or at the higher-order level, and *vice versa* if it denotes XYZ. Thus, the subject's presumed inability to discriminate between thoughts about water and thoughts about XYZ has no bearing on the question of her ability to know the content of her occurrent thought.²⁶ The transparency of the logical operator "I am thinking that ...," as applied to one's own occurrent

²⁵ Thus, this account is sometimes labeled the 'inclusion account' (the object-level thought is included in the higher-level thought) or the 'redeployment account' (the object-level thought is redeployed in the higher-level thought)

²⁶ This is sometimes referred to as the problem of *comparative content* (following Kevin Falvey and Joseph Owens, "Externalism, Self-Knowledge and Skepticism"). Talk of 'comparison' or 'discrimination' between contents frequently invites the objection that this attempt to conjure a problem of self-knowledge for externalism must illicitly draw on perceptual or observational models of self-knowledge, precisely the kind of model that externalist philosophers are typically busy arguing the error of. One is, in effect, inviting us to think of the problem of 'discriminating' between two thought contents as analogous to the problem of perceptually discriminating between the real *Nightwatchman* and a well-made forgery. On this, see Richard Moran, *Authority and Estrangement*, pp. 15-16, in response to Paul Boghossian, "Content and Self-Knowledge." Moran, in turn, draws on an example due to Wittgenstein, in *The Blue Book*, p. 39.

thoughts, *guarantees* that there cannot be a failure of self-knowledge in such cases. The only level at which any kind of failure might occur is at the level of being able to pick out the right kind of stuff in the world (say, “I am hereby thinking that this is water,” said by a recent arrival on Twin Earth). But this is an ordinary empirical mistake, and has no bearing on self-knowledge.

These reflections allow Burge to generate a template for what he calls *cogito*-thoughts, which will serve as the paradigm of *basic self-knowledge*. Burge offers the following example:

(1) I am thinking that there are physical entities.

He comments: “To be true, (1) requires only that I am engaging in some thought whose content is that there are physical entities” (Burge, “Our Entitlement to Self-Knowledge,” p. 240). Presumably that is precisely what I am doing when I declare that I am thinking about physical entities. (1) is thereby rendered true by its very performance, and does not rest on observation or perception, either in its inner or outer modes. In this sense, pronouncements of form (1) are “contextually self-verifying” and, indeed, “infallible” (ibid.)

To recap, what I have called the generic problem of externalism and self-knowledge concerns the question of how we can achieve a form of self-knowledge that is independent of sense-experience if the contents of our thoughts are determined at least in part by contingent factors in the external world, factors of which we could know only by way of sense experience. Burge’s *cogito*-argument provides an example of just such self-knowledge: self-ascriptions of occurrent thoughts (“*cogito*-judgments”) are infallible, contextually self-verifying, and independent of sense experience. They constitute basic self-knowledge. All it takes to be entitled to claim such knowledge is to ascribe the thought to oneself in the right kind of way.

I call the challenge to which the *cogito*-argument is an answer the “generic problem” of externalism and self-knowledge in order to point out that it arises from a factor that is not unique to Burgean anti-individualism²⁷ and to differentiate it from the problem of self-knowledge that I will go on to address shortly, which in my view is unique to Burge’s position.

The philosophical assumptions that lead some to think that the generic problem is indeed a problem stem, in my view, from an insufficient appreciation of just how radical the externalist move is. Crudely put, the assumption driving the objection is that externalism simply pushes the internalist conception of the mind one step farther back, but otherwise leaves its structure intact. There would still be, then, some element of the mind, home to a fully self-sufficient Cartesian cognizer, as yet untouched by the externalist argument. If this were so, it would indeed have devastating consequences. On this reading, the kinds of skeptical worries that traditional Cartesianism could formulate regarding the subject’s possibility of gaining knowledge about objects and events in the external world could now be formulated about the subject’s possibility of knowing about his own mental states. But, of course, this is not how the externalist argument operates. Externalism leaves no element of the mind untouched. The objection assumes that there must be some level of the mind where mental content remains constant throughout environmental changes. But this is an illegitimate throwback to the kind of narrow content approach to the philosophy of mind that we dispelled in chapter 1.

In other words, I hold that Burge’s *cogito*-argument is convincing, at least as understood as a purely deflationary argument designed to defeat the ‘achievement’-half of the incompatibilist

²⁷ Indeed, arguments of a very similar form to Burge’s *cogito*-argument had already been formulated by other prominent externalists, such as Donald Davidson (“Knowing One’s Own Mind”) and John Heil (“Privileged Access”). See also Robert Stalnaker, “Narrow Content,” p. 209, for a crisp statement of the position.

objection.²⁸ What, though, of the consequence problem? This, recall, is the claim that *if* externalism could secure a priori self-knowledge, it could only come at the expense of allowing – absurdly – that such knowledge could provide a route also to a priori knowledge of contingent states of affairs in the external world. All in all, Burge has had little to say about this challenge; remarkably so, given how much of the critical literature is concerned with it. Moreover, in my view, what little he has had to say about it threatens to spoil some key anti-individualist theses.²⁹ However, I shall not pursue this matter here, but only note that while the achievement problem is the same in the case of both Burgean anti-individualism and the more generic form of externalism that I favor (and in this sense constitutes a ‘generic’ problem of externalism and self-knowledge), the consequence problem strikes with very different force in the two cases. Indeed, the prominent “generic externalist” Donald Davidson has actually fully embraced the consequence problem and sought to convert it into a new

²⁸ I enter this small reservation because there are occasions on which Burge seems to hold out for the idea that the cogito-argument contains the seeds of some positive insights into the basic structure and workings of the rational mind (cf. Burge, “Mental Agency in Authoritative Self-Knowledge,” p. 419). I have strong doubts about attempts to develop the argument along these lines, and side with Donald Davidson when he argues that Burge’s approach here runs the danger of becoming, in a word, “individualistic” (Davidson, “Reply to Burge,” p. 664). In an insightful article, Victoria McGeer expresses a general reservation which I share, concerning externalist invocations of higher-order thoughts, as they “[make it] seem as though being consciously aware of, and even expressing a first-order state – for example, there is water in the glass – depends on having formed (perhaps unconsciously) a second-order belief about that first-order state.” This runs afoul of what is in her view “a fundamental insight of the ‘externalist turn’,” namely “that thinking (and talking) about the world consciously does not consist in an unconscious, second-order process of thinking about thoughts about the world. Rather, it is just a process of thinking (and talking) about the world” (McGeer, “Is ‘Self-Knowledge’ an Empirical Problem?,” p. 502). Consult Sarah Sawyer’s “In Defense of Burge’s Thesis” for more on why we should understand the *cogito*-argument as having only a deflationary application. On this view, generating positive insights into self-knowledge and the structure of rational agency becomes entirely the job of Burge’s other argument, the argument from critical reasoning.

²⁹ I have in mind passages in which Burge, in an effort to deflect the consequence problem, will say that there is no problem for the anti-individualist in allowing that someone may be capable of thinking, say, water-thoughts even in water-less environments. (For instance, “Other Bodies,” pp. 94-98, “Reply to Martin Davies,” pp. 262-263.) What is problematic about this move is that it jeopardizes anti-individualism’s distinctive metaphysical component, according to which there is a necessary connection between the kinds of thoughts a subject can token and the kinds of physical substances that exist in his environment. If Burge were to follow this move to its conclusion, anti-individualism would simply collapse into a quite generic form of externalism: all we should say now is that an individual, in order to have any thoughts at all, must find himself in *some* physical environment or other. (Cf. Burge, *Foundations of Mind*, Introduction, p. 9: “for an individual to have any propositional attitudes at all, he must have *de re* states or attitudes.”) This significantly weaker construal avoids, as we shall see, the consequence problem, but also fails to support the kinds of counterfactuals that Burge’s thought-experiments operate on.

argument against skepticism.³⁰ Davidson thinks, that is to say, that certain empirical facts, for instance the existence of the external world and the existence of other minds, can be known a priori from mere consideration of the contents of our own thoughts – not, to wit, from the specific contents of our thoughts, but rather from the fact that there is contentful thought at all.

It should also be clear why accepting or even embracing the consequence problem would be much less difficult for a generic externalist like Davidson than for Burge. Davidson's arguments are holistic and work on a global level. What they allow us to deduce is an answer to the skeptic (and hence the mere existence of external world which is "by and large" as we conceive it to be, although not necessarily so in any particular aspect). Burge's argument, by contrast, works at the level of individuated *de re* thought. It concerns the metaphysical question of what must be the case for a thinker to instantiate any particular thought content. Thus, the consequence problem that he faces is different from, and much more severe than, that which is faced by the generic externalist. Burge would be forced to countenance not merely an a priori inference to the existence of an external world, but a priori knowledge of particular states of affairs in the world – for instance, the existence of water.³¹ In my view, then, the consequence problem is a further argument in favor of a generic form of externalism as against Burge's anti-individualism, though not one that I will pursue here.

³⁰ Cf., Davidson, "A Coherence Theory of Truth and Knowledge," "Three Varieties of Knowledge."

³¹ None of this is to say, of course, that Davidson's anti-skeptical argument is entirely problem-free. For instance, one is reminded of the following confession from Colin McGinn: "I had the idea of using the anti-skeptical consequences of the principle of charity as a *reductio* before I learned that Davidson regards this as a *virtue* of his account of interpretation" (McGinn, "Radical Interpretation and Epistemology," p. 359n6, cited in Bjørn Ramberg, "What Davidson said to the Skeptic," p. 232n10). In any case, it seems safe to say that what resistance the argument encounters today (cf. Burge, "Some Reflections on Skepticism: Reply to Stroud") has to do with whether it actually has the powerful anti-skeptical force that Davidson thinks it has, not with whether its being a consequence of Davidson's externalism reduces that externalism to absurdity.

4.3.5 Philosophical limitations of the *cogito*-argument

Even as we find reason to agree with the *cogito*-argument, however, there are legitimate concerns about just how widely the underlying model applies and so about how much philosophical burden the argument can shoulder. Right away, we can note the following kinds of limitations: first, *cogito*-judgments lack, precisely through the performative character that gives them their particular flavor, any appreciable temporal dimension. The argument applies only in the very act of thinking the thought in question. Thus, the *cogito*-argument seems unable to properly connect what I am knowledgeably thinking *now* with what I was thinking just a few moments ago. We are not infallible in recollection, even though we do seem to enjoy a legitimate presumption of first-person authority even here. Responding to pressure on these matters, Burge has developed an account of our entitlement to the products of what he calls preservative memory.³² I think this account is extremely interesting and quite plausible. However, it seems that an appeal to the function of preservative memory will never have quite the intuitive force of the *cogito*-argument itself. Insofar, then, as the *cogito*-argument requires supplementation with an argument concerning preservative memory, this constitutes a notable limitation on its scope.

Second, since the *cogito*-argument focuses only on occurrent thoughts, it is unclear that it has any bearing on our self-knowledge of standing states, such as beliefs and desires. As social psychologists take great pleasure in showing, we have far from infallible insight into several of our standing states. These manifest themselves in our behavior in ways that may only be apparent to us upon empirical investigation.³³ Nonetheless, even while we are not infallible, we do possess a

³² Cf. Burge, “Memory and Self-Knowledge,” “Content Preservation.”

³³ See, for instance, Nisbett and Wilson’s classic “Telling More than We can Know” and the currently popular ‘Implicit Association Test,’ which the designers introduce with the following words: “It is well known that people don’t always

remarkably high degree of non-empirical self-knowledge even on these matters, and it seems to behoove an account of our self-knowledge to have something to say about this. The cogito-argument cannot help us here, so we will have to look elsewhere.

The third point is that while the cogito-argument may be able to explain how it is that we cannot be mistaken about the *content* – *that water is wet* – of our object-level propositional attitudes when self-ascribing those attitudes, it seems unable to explain how I can grasp my particular attitude toward this content – *believing* that water is wet, *doubting* that water is wet, and so on. Our knowledge of our attitudes simply falls outside the scope of the cogito-argument. Yet grasping the particular attitudinal force of one’s various attitudes is properly thought of as a matter of self-knowledge.³⁴

All of these limitations are widely noted in the critical literature, and also acknowledged by Burge. Together, they conspire to throw doubt on whether the cogito-model can serve as the paradigm of anything like a comprehensive model of self-knowledge. In particular, whatever else one might say about such self-ascription of thoughts, one cannot help but feel that appealing to it isn’t the least bit illuminating when time comes round to account for the kind of self-knowledge that we exercise in reasoning. For this kind of reasoning involves things like assessing what is a reason for what; and thus involves relations between several beliefs, not just self-ascriptions of beliefs, one by one, which is all that the *cogito*-model can account for.

‘speak their minds’, and it is suspected that people don’t always ‘know their minds’. Understanding such divergences is important to scientific psychology.” For more, go to <https://implicit.harvard.edu/implicit/>.

³⁴ Cf. Burge, “Mental Agency in Authoritative Self-Knowledge,” pp. 426-427, Sven Bernecker, “Externalism and the Attitudinal Component of Self-Knowledge,” John Gibbons, “Externalism and Knowledge of the Attitudes.” There is an important question in this vicinity that, unfortunately, I cannot go further into here. This is the question of whether a properly post-Cartesian philosophical psychology should hold that knowledge of attitudinal force is properly part of privileged self-knowledge at all. Thus, Bernecker, “Self-Knowledge and Closure,” p. 347: “Privileged access to the attitudinal components of one’s thoughts is one of the Cartesian superstitions that the inclusion theory forces us to abandon.” Similar views are expressed in Fred Dretske, *Naturalizing the Mind*, pp. 54-55. A more thorough survey of such ideas in recent philosophy and psychology can be found in ch. 9 of Alvin Goldman’s *Simulating Minds*. On my view, this systematic dissociation of knowledge of content from knowledge of attitude seems, so far from a final liberation from a Cartesian picture of our cognitive operations, precisely a throwback to it.

Presumably, then, it is in order to get the right handle on the notion of self-knowledge involved here that Burge turns to the argument from critical reasoning.³⁵ Where the *cogito*-argument was capable of securing only a deflationary notion of self-knowledge, the argument from critical reasoning promises to anchor what I have called a substantive notion of self-knowledge, with application to a much wider range of states than the *cogito*-argument was capable of securing. We shall now turn to reviewing the idea of critical reasoning in more detail.

4.4 The idea of critical reasoning and its connection with self-knowledge

I remarked above that the argument from critical reasoning marks something of a transcendental approach to the question of self-knowledge. By this I mean that the claim that we are in fact critical reasoners is simply assumed and never defended in detail. The same holds for Burge's claims about the specific character of our capacity for critical reasoning: he sketches an account of this capacity which is certainly plausible, but which is also never defended in detail.³⁶ Against the background of these assumptions, Burge can trace out the conditions of possibility for this kind of critical

³⁵ Actually, Burge has never spoken out on what he takes the division of labor between the two arguments to be, even in those discussions where he acknowledges the inherent limitations of the *cogito*-model (cf. "Mental Agency in Authoritative Self-Knowledge: Reply to Kobes"). On the other hand, Sarah Sawyer ("In Defense of Burge's Thesis") plausibly argues that we should construe the division of labor in the way that I follow here.

³⁶ It is worth noting, for instance, that Burge never addresses in any detail the challenge of psychological experiments which claim to empirically demonstrate systematic shortcomings in our ability to gauge the processes that lead to the formation of our attitudes, and thereby, one would think, similar shortcomings in our ability to exert rational control over these attitudes. Famous examples include Nisbett and Wilson, "Telling More than We can Know," Nisbett, *Strangers to Ourselves*, and Gopnik, "How We Know Our Minds: The Illusion of First-Person Knowledge of Intentionality". Indeed, one would think that these findings are especially pernicious for Burge, in that they make no claims about our ability to know the occurrent thoughts (i.e., they leave untouched the conclusion of the argument from *cogito*-thoughts). Indeed, our ability to know and accurately report what is on our minds is very much presupposed in this literature. Rather, they target the more ambitious argument from critical reasoning, in that they ask questions about our supposed ability to exert rational control over the formation of these attitudes. In "Reason and the First Person" (p. 255), Burge does say that our status as critical reasoners is compatible with the occasional lapse of this kind of self-knowledge. However, what goes unremarked in this brief response is precisely the apparently *systematic* nature of the kinds of lapses that these psychologists claim to find.

reasoning. Among these conditions will be a certain kind of self-knowledge. We can then draw the conclusion that we possess the requisite kind of self-knowledge, insofar as we are critical reasoners.³⁷

Our status as critical reasoners, in a nutshell, involves our ability to assume an attitude toward our attitudes *qua* attitudes. Attitudes are the kind of thing for which one might legitimately inquire into reasons, and so the posture we adopt is, as the name suggests, one of *critical review* – we adopt it for the purpose of checking our beliefs, their internal coherence, their epistemic credentials, and so on. Thus the argument from critical reasoning brings to light the peculiarly *normative* standpoint rational beings possess with respect to their own mental lives. It starts from the observation that we stand to our attitudes not as we stand to a realm of objective facts. Rather, our attitudes are *ours* – in a very real sense, we have a responsibility for what we think that we do not toward an objective realm of fact (even as what we *ought* to think remains very much constrained by these objective facts).³⁸ Unlike many other philosophers, Burge does not hesitate to say that higher animals and young children have propositional attitudes.³⁹ However, what they do not have, and which thereby differentiates them from rational human adults, is this ability to survey and assume a critical stance toward their attitudes as *their* attitudes, as things over which they possess an appreciable measure of rational control.⁴⁰

³⁷ Note the connection with Sydney Shoemaker’s argument designed to demonstrate the impossibility of the self-blind person (“Self-Knowledge and ‘Inner Sense’”), i.e., a person who possesses the normal range of abilities with respect to the mental lives of others, but who has no immediate / non-inferential awareness of her own mental life.

³⁸ In this way, there is a clear affinity between Burge’s arguments concerning the connection between critical reasoning and self-knowledge and the approach developed by Richard Moran in his *Authority and Estrangement: An Essay on Self-Knowledge*.

³⁹ Cf. Burge, “Perceptual Entitlement,” “Social Anti-Individualism, Objective Reference,” pp. 312-313. Donald Davidson (“Rational Animals”) and John McDowell (*Mind and World*, ch. VI) are among the philosophers who have expressed reservations about attributing propositional attitudes to animals. However, their reasons for this are complex, and we would be quite wrong to assume that they thereby mean simply to dismiss the concerns and insights that motivate Burge’s positive affirmation. McDowell and Davidson discuss this matter most recently in McDowell, “Subjective, Intersubjective, Objective” and Davidson, “Replies to Barry Stroud, John McDowell, and Tyler Burge,” pp. 696-698.

⁴⁰ See, for instance, Burge, “Reason and the First Person,” pp. 252-253, and “Perceptual Entitlement,” p. 521.

Critical reasoning is thus essentially a posture of normative review.⁴¹ One surveys one's attitudes *as* attitudes, evaluating their epistemic credentials and checking for possible conflicts among them. A crucial factor appears when we consider the *implementation* of the results of critical reasoning. For when I find an attitude of mine wanting in some respect in light of the ideal of reason, I thereby possess an immediate reason to change my attitude. This stands in a sharp contrast with the implementation of the results of my critical review of *another's* attitudes. For we often perform critical reviews also of the attitudes of others. There is no doubt a sense in which if you perform a critical review of my attitudes and find them wanting, I thereby also possess a (*prima facie*) reason to change my mind. But the implementation of this reason is not immediate, nor is the reason decisive.⁴²

Burge suggests two ways of cashing out this difference. One is that reasons are tied to persons, or, in a metaphor that we shall explore in more detail later, to *points of view*. In Burge's words:

Since mismatches in information on which reasons are based are always possible, no rational evaluation that is not universally self-evident, however reasonable, has rationally immediate application, with consequences for immediate implementation, across persons or points of view. As long as the attitude is not taken to be one's own, there is always the possibility of a gap, and filling that gap requires a rational step. (Burge, "Reason and the First Person," p. 255)

The reason, he goes on to say, why the results of critical reasoning will provide immediate incentive for implementation "is because the first- and the second-order perspectives are the *same* point of view" (Burge, "Our Entitlement to Self-Knowledge," p. 258). More elaborately:

⁴¹ Burge argues, moreover, that the status of critical reasoner cannot be based merely on a *potential* for critical reasoning, but must be based on *actual* critical reasoning. This is a central theme for instance in "Reason and the First Person," pp. 250-251.

⁴² See, in particular, Burge, "Reason and the First Person," pp. 254-255.

If the reviewing knowledge is to be integral to critical reasoning, if it is to provide immediate rational ground for change in the reviewed material, the review must take up the same perspective or point of view as the act under review – the reasoner’s own object-level point of view. [...] the point of view of the review and that of the attitude must be the same and must be first-personal. (Burge, “Our Entitlement to Self-Knowledge,” p. 262)⁴³

A second point underwriting the difference between my relation to my own attitudes and my relation to the attitudes of others has to do with the different mechanisms by which a change in attitude would come about in each case. In my own case, the conclusion of the rational review is ordinarily quite enough for the change to take place – as the conclusion is drawn, so my attitude changes. This is indeed part of what we mean by saying that the implementation is immediate. By contrast, in reviewing another’s attitudes, implementation of the change will have to proceed via *some adopted means* (paradigmatically verbal ones).⁴⁴ While this second difference is interesting in its own right, my focus in the following will be on the first claim.

The idea of critical reasoning, then, pivots on the idea of the immediacy of the implementation of its results. For such critical reasoning to obtain, there needs to be a very robust connection between our attitudes and our ability to know them. It is central to Burge’s argument that no theory which would model our self-knowledge on some form of ‘inner observation’ or

⁴³ A similar line of thought is also suggested in the earlier essay “Individualism and Self-Knowledge,” p. 658, although this is before the development of the concept of critical reasoning and the distinction between entitlement and justification. Here Burge writes: “The special epistemic status of these cases depends on the judgments’ being made simultaneously from and about one’s first-person point of view. The point of view and time of the judgment must be the same as that of the thought being judged to occur.”

⁴⁴ Burge, “Reason and the First Person,” p. 255. Burge notes (ibid.) that there may also be attitudes of one’s own that show a certain kind of recalcitrance, such that they resist immediate review and one might have to adopt means resembling those one might adopt in trying to persuade another. But these will necessarily have to be the exception, lest one not count as a reasoner at all. If such recalcitrant attitudes were seen to proliferate, one would have to doubt that the person is home to a single, rational point of view at all. Daniel Dennett appears to take a rather different view of this matter, and speaks of “induc[ing] oneself to change” by various means such as “persuasions, arguments, threats, bribes” (Dennett, “Conditions of Personhood,” pp. 284-285). Richard Moran discusses Dennett’s position in *Authority and Estrangement*, p. 119n6, arriving at a conclusion that closely resembles that of Burge.

introspection will provide for a sufficiently robust connection. For any model of observation or introspection must allow that such modes of knowing are compatible with what Burge calls *brute error*. A brute error is defined as an error that does not indicate a malfunction in the system in which the error obtains. Examples of brute error in ordinary perception are part and parcel of the philosopher's stock in trade: shrewdly place a mirror in front of my gaze, and I may think that the apple on the table is in front of me, when, in fact it is behind me. We can readily make sense of such errors in perception without questioning the well-functioning of the mechanism that delivers the perception.⁴⁵

According to Burge, self-knowledge is importantly different from perceptual knowledge in just this respect. We precisely cannot make sense of someone misidentifying the content of their own first-order mental states without also taking this as indicating some malfunctioning of their cognitive abilities. This form of malfunctioning is one facet of what we call *irrationality*. Further, the fact that such malfunctioning is immediately recognizable as a form of irrationality suggests an intimate connection between rationality and self-knowledge.⁴⁶

The best we can expect from an observational theory of self-knowledge is a contingent connection between our attitudes and our knowledge of our attitudes. Such contingent connections, too, come in various degrees of robustness. But even the most robust of such connections remains contingent, and would thus in principle be compatible with the occurrence of brute error.⁴⁷ This, according to Burge, is not strong enough for critical reasoning. The only thing that will exclude the possibility of brute error is a direct and non-inferential form of knowledge.

⁴⁵ Cf. Burge, "Perceptual Entitlement," p. 507.

⁴⁶ For Burge's most protracted discussion of self-knowledge and brute error, see "Our Entitlement to Self-Knowledge," pp. 251-256.

⁴⁷ To my knowledge, Burge does not provide an argument for this claim. However, the idea is certainly plausible, and sees much play in the current debates, in the form of argument (often attributed to Sydney Shoemaker) that it belongs to the nature of a perceptual mechanism that it be fallible. Therefore, infallible self-knowledge cannot be delivered by anything like a perceptual mechanism.

4.5 Deflationary and substantive self-knowledge

A brief reminder of where we have come so far might be helpful: according to the doctrine of incomplete concept mastery, there is nothing wrong with attributing to a subject thoughts containing concepts he does not fully master. Supposedly, then, the subject who utters the sentence “I have arthritis in my thigh” evinces an incomplete grasp of the concept *arthritis*. According to the line of thought that we are considering, however, this has no bearing on whether it is nonetheless the concept *arthritis* that he is thinking with, whether this is the concept that captures his cognitive perspective.

It is important to note here that Burge’s contention is not the relatively uncontroversial point that in spite of his conceptual confusion, we bystanders may conveniently use the word ‘arthritis’ in describing his thought contents, say, for lack of a better term. Rather, it is *the concept arthritis* that he is thinking with, even as his utterance clearly displays his incompetence with this concept. It is this concept which captures his cognitive perspective, which is to say, among other things, that this is the concept that should figure crucially in explanations of his actions and assessments of his rationality.

The doctrine of incomplete concept mastery has as its corollary the idea of incompletely understood *thoughts*: a subject can be rightly credited with thinking thoughts that he cannot, for instance, correctly spell out the consequences of, not because he is stupid or otherwise hasn’t given much thought to the matter, but because he thinks the concepts that figure in his thoughts apply to things that they don’t. We can easily imagine our subject spelling out what he takes to be the consequences of his thought, yet that he would be systematically mistaken in his endeavor. Someone else might be in a much better position than he is to spell out what he is actually thinking. On

Burge's view, however, this should not deter us from saying that it is this very thought that he is thinking.

In my view, this constitutes a second problem of self-knowledge, a problem unique to Burge's anti-individualism. But it is plain to see that this problem must involve a different *notion* of self-knowledge altogether than does the generic problem. For our sense that if the anti-individualist thought attribution were correct, then this person would not know what he is thinking persists even in the face of Burge's assurance that he can attribute the thought to himself in the right kind of way by saying "I think, with this very thought, that *p*." What we are asking about now, in other words, goes well beyond the truth- or verification conditions of a certain kind of self-ascriptive statement. Rather, we are asking about the subject's *understanding* of the thoughts that he thus ascribes to himself. The intuition is that failing to grasp, for instance, the basic consequences of one's thoughts is, in a relevant sense, to fail to know what one is thinking. If incomplete concept mastery is, as Burge claims, "common or even normal," this should have dramatic consequences for our view of ourselves as by and large rational agents, possessed of the right kind of cognitive control over our mental lives. Burge, however, denies that this is a consequence of his view, and, as we have seen, his own argument from critical reasoning is premised on people having just such control.

It seems, then, that something will have to give. Burge himself is, of course, aware that the term 'self-knowledge' is open to various interpretations. In "Individualism and Self-Knowledge," he introduces a version of what I have called a distinction between two notions of self-knowledge – a deflationary notion and a substantive notion. He writes:

There is no apparent reason to assume that, in general, one must be able to explicate one's thoughts correctly in order to know that one is thinking them.

Further:

It is a truism that to think one's thoughts, and to think cogito-like thoughts, one must understand what one is thinking well enough to think it. But it does not follow that such understanding carries with it an ability to explicate correctly one's thoughts or concepts via other thoughts or concepts; nor does it carry an immunity to failures of explication. So one can know what one's thoughts are even while one understands one's thoughts only partially, in the sense that one gives incomplete or mistaken explications of one's thoughts or concepts. One should not assimilate 'knowing what one's thoughts are' in the sense of basic self-knowledge to 'knowing what one's thoughts are' in the sense of being able to explicate them correctly – being able to delineate their constitutive relations to other thoughts. (Burge, "Individualism and Psychology," p. 662).

Donald Davidson was among those who objected to the doctrine of incomplete concept mastery on the grounds that it seems to raise this peculiar problem of self-knowledge.⁴⁸ The passage from Burge that I just read is in part meant as a rebuttal of Davidson on this score. Writing as recently as 2003, Burge repeats essentially the same line of criticism. Here he claims that Davidson has never explained what the supposed problem consists in, and hypothesizes that the sense that there is a problem here at all "rests on a failure to distinguish knowing what one's thoughts are in the sense of being able to understand them well enough to think them and self-attribute them, and knowing what one's thoughts are in the sense of being able to give correct explications of them. Davidson's indiscriminate claim that on my view the relevant agents do not know what they mean and think suggests this conflation" (Burge, "Davidson and Forms of Anti-Individualism: Reply to Hahn," p. 357).

To my mind, this is a very puzzling argument for Burge to offer as a rebuttal of Davidson. Presumably Davidson understands the distinction that Burge is pointing to. Presumably he does not hold that the subjects of Burge's thought-experiments are incapable of attributing to themselves the

⁴⁸ Davidson, "Knowing One's Own Mind," pp. 25-26, "Epistemology Externalized," pp. 197-199..

relevant thoughts in the way that the cogito-argument describes.⁴⁹ So we have reason to think that Davidson is not merely *conflating* two senses of self-knowledge. Rather, I think we should understand him as claiming that securing the first sense of self-knowledge is not enough. We also need to account for the richer, more demanding sense of self-knowledge, and this is what it seems that the doctrine of incomplete concept mastery does not allow us to do.

I take it, though, that when Burge draws this distinction, he does mean for it to be implying an argument of sorts. Presumably, he means to say that this richer, more demanding sense of self-knowledge is *not* a legitimate notion of self-knowledge at all. From what I can tell, he never provides an argument for this contention, other than to call on the precisely the same intuitions that were supposed to provide support for the anti-individualist thought-experiments in the first place. Thus, he will sometimes say that in pressing for this richer, more demanding requirement on self-knowledge, we are guilty of simply begging the question against the doctrine of incomplete concept mastery.⁵⁰

We have, then, from Burge's own hand, an appeal to a distinction between a *deflationary* and a *substantive* notion of self-knowledge. The first notion is deflationary, because it does not require us to say more about self-knowledge than that the subject needs to "understand what he is thinking well enough to think it," and this much is clearly assumed in Burge's description of the thought experiment. I call the other notion substantive, because it makes substantive demands on the subject's explicatory abilities, his ability to connect his thoughts with each other, or as I shall say more generally, his ability to *do* things with that thought, beyond merely attributing it to himself.

I will now argue that the deflationary notion of self-knowledge is not sufficient. Anti-individualism also needs to incorporate what I have called a substantive notion of self-knowledge if

⁴⁹ As I pointed out above, Davidson himself has employed a similar argument. See "Knowing One's Own Mind," pp. 34-38, "What is Present to the Mind?" p. 61.

⁵⁰ Cf. Burge, "Individualism and the Mental," p. 131.

it is to sustain its claim to be able to capture the subject's cognitive perspective. Yet such a substantive notion of self-knowledge is what anti-individualism cannot have, if it is to hold on to the doctrine of incomplete concept mastery. The connection between thought attribution and rationalization and action explanation is premised on the subject possessing self-knowledge of this kind. In other words, a concession to substantive self-knowledge is structurally integral to the theory and practice of thought attribution, in a way that is at odds with the doctrine of incomplete concept mastery.

4.6 Critical reasoning requires substantive self-knowledge

Since the anti-individualist is likely to claim that in arguing for the importance of this substantive notion of self-knowledge, we are guilty of begging the question against his position, it would be useful if we could find a way of motivating our claim from within Burge's own project. In my view, Burge's resort to the argument from critical reasoning provides a leverage point for just the kind of internal critique that we need.

As we saw above, it is reasonable to assume that Burge turns to the argument from critical reasoning because he sees that what is secured with the *cogito*-argument is too narrow to serve as a full account of self-knowledge. But here lies the snag: in turning to the argument from critical reasoning, Burge appears to simply assume that he has already shown, with the help of the *cogito*-argument, that anti-individualism is compatible with self-knowledge. On this assumption, the argument from critical reasoning would merely be a further elaboration of a result that he has already secured through that other argument. I think this is wrong. For the argument from critical reasoning makes substantial demands on self-knowledge in a way that the *cogito*-argument does not. The two arguments address different notions of self-knowledge altogether: the *cogito*-argument

involves the deflationary notion, while the argument from critical reasoning involves the substantive notion. A substantive notion of self-knowledge, I will try to show, is something that anti-individualism must have if it is to sustain its claim to capture the subject's cognitive perspective, but which it cannot have as long as it holds onto the doctrine of incomplete concept mastery.

The problem with the cogito-model is not just that it doesn't cover enough cases.⁵¹ It would seem that even for the cases that it does cover, it does not say all that needs to be said. Remember that the cogito-model provides an answer to the following question: assume, as externalists do, that a subject's thought contents may shift with changes in the environment. How, then, can a subject of such environmental switches know which of these many possible thoughts he is thinking, without first determining which of these many possible environments he finds himself in? The Burgean answer, simple and elegant, is that he can demonstrate this knowledge simply by saying "I believe, with this very thought, that *p*." Relative to the challenge that it seeks to defuse, this seems an appropriate answer. But this answer can hardly represent more than one dimension of the question of self-knowledge. For even in the case of occurrent thoughts, it is relevant to ask not just whether the subject can pick out the right thought in a *de re* fashion, which is all that the cogito-model really comes down to. When the "objects" we inquiring into are the subject's own thoughts, it seems relevant to also ask about her *understanding* (her ability, as it were, not just to refer *de re* to the thought, but to 'unpack its content'). Such understanding – what I have called substantive self-knowledge – is what seems to be ruled out by the doctrine of incomplete concept mastery, according to which a subject can be found to have these thoughts – in all relevant senses – without understanding them.⁵²

⁵¹ See section 4.3.5 for more detail.

⁵² In a discussion of Burge on self-knowledge, Sanford Goldberg proposes, in the spirit of Burge's appeal to the distinction between what I have called deflationary and substantive self-knowledge, that "one can count as knowing *that* one is thinking that *p* without knowing *what* one is thinking in (knowing that one is) thinking that *p*" (Goldberg, "What

The doctrine of incomplete concept mastery is in tension with substantive self-knowledge; this much is effectively conceded by Burge in the quotes I gave above. Being able to “explicate one’s thoughts correctly” and to “delineate their constitutive relation to other thoughts” is precisely the stuff of substantive self-knowledge.⁵³ But the tenor of the argument he gives against Davidson is clearly that if we were to insist that such explicatory abilities are integral to self-knowledge, we would simply be wrong.

The question is whether this is a line that Burge can maintain once the argument from critical reasoning comes on the stage. For the kind of self-knowledge that is involved in critical reasoning is precisely the substantive variety. That is, the kind of self-knowledge in play in critical reasoning and in rational agency quite generally goes far beyond anything that can be delivered by the cogito-argument, both in terms of scope and in terms of depth. It requires an account of self-knowledge that goes beyond merely occurrent thoughts to also encompass standing states such as beliefs and desires, and it requires an account of the subject’s ability to evaluate the epistemic standing of her beliefs, the conceptual connections between them, and so on. But to see such a requirement as foundational to the theory of thought-ascription is flatly inconsistent with anti-individualism’s doctrine of incomplete concept mastery.⁵⁴

Do You Know When You Know Your Own Thoughts?” p. 254). Goldberg argues that Burge should simply concede that anti-individualism can make no sense of substantive self-knowledge (what Goldberg calls ‘conceptual self-knowledge’) so as to consolidate the territorial gains had through the argument from *cogito*-thoughts: the unimpeachable status as knowledge of self-ascriptions of present tense occurrent thoughts. While this move is certainly consistent with the tone of Burge’s response to Davidson, I don’t think it is a strategy that he can adopt without major loss. As I will argue, without assumptions about substantive self-knowledge, our thought attributions can get no grip on the subject’s patterns of reasoning and his actions. Conceding this is to give up on anti-individualism as a theory of thought attribution.

⁵³ For more detail, see Jessica Brown, “Critical Reasoning, Understanding, and Self-Knowledge,” Åsa Wikforss, “Self-Knowledge and Knowledge of Content.”

⁵⁴ Jessica Brown (“Critical Reasoning, Understanding, and Self-Knowledge”) acknowledges that there is a tension between the doctrine of incomplete concept mastery and the idea that we should view critical reasoning capabilities as foundational to the cognitive lives of rational human adults. On her view, this insight forces the realization that such incomplete concept mastery cannot be as rampant as Burge assumes. In her view, critical reasoners must be “mostly right about most of the concepts they possess; for most of the concepts they think with, they must understand the core of those concepts, and any incompleteness in their understanding must relate to only relatively minor and peripheral

In short, then, Burge turns to the argument from critical reasoning to gain an account of self-knowledge that can improve on the cogito-model's crucial shortcomings in terms both of scope and depth. I agree with Burge that an account of the requirements of critical reasoning and of rational agency more broadly has the right shape for such a theory of self-knowledge. The trouble is, however, that Burge never considers whether the doctrine of incomplete concept mastery is compatible with our status as critical reasoners and as self-knowing rational agents in the requisite sense. He seems to simply assume that he has shown, with the help of the cogito-argument, that anti-individualism is compatible with self-knowledge. I have argued how this will not do, since the notion of self-knowledge at stake in the argument from critical reasoning is very different from the kind at stake in the cogito-argument. If I am right, then, an account of substantive self-knowledge is something that anti-individualism must have if it is to provide a full and satisfying account of the conditions of rational agency, but which simultaneously, it cannot have, if it is to hold on to the doctrine of incomplete concept mastery.

matters" (p. 675). However, she argues that this territorial concession will allow Burge to retain the ground that should really matter to him, namely that the idea of incomplete concept mastery is metaphysically cogent and describes a real phenomenon. As I shall elaborate in the next section, this is a Pyrrhic victory. I have no doubt that there could be cases of incomplete concept mastery in the sense that Brown describes. But Burge's cases are not among these. One would possess incomplete concept mastery in Brown's sense if one possessed a faulty grasp of the real implications of one's own thinking. One would in this sense fail to exercise rational control over one's own patterns of thought. By contrast, Burge's subjects fail to grasp, or have false beliefs about, the conventional implications of the words they use. They can be as consistent and as thoughtful in their thinking as we like, and still possess incomplete concept mastery in Burge's sense. It is the idea that the conventional semantic values of terms should pull this kind of weight in thought attribution that I am arguing against. As I shall explain in the next section, if we take this view, we would deprive ourselves of the possibility of distinguishing properly between cases of genuine inconsistency and putative cases of incomplete concept mastery. It follows that such attributions could not work properly in assessments of the subject's rationality and explanations of his actions, as Burge thinks they can.

4.7 Two kinds of cases

I have argued that it is essential to our picture of people as rational agents that when we attribute thoughts to them, we simultaneously credit them with an appreciable measure of what I have called substantive self-knowledge; an ability not only to ascribe these thoughts to themselves in a deflationary fashion, but to *understand* these thoughts. Understanding here is, at least in relevant part, a matter of correctly relating one's various thoughts to each other. This form of understanding, for instance like that which we exercise in critical reasoning, essentially involves stringing together multiple beliefs and appreciating the relations between them.

I believe this can shine important light on the question that we started out with, namely that of our ability to capture the subject's cognitive perspective. My claim has been that anti-individualism's doctrine of incomplete concept mastery prevents us from catching the subject in the exercise of this kind of substantive self-knowledge. We should now be able to see that this claim is just the mirror image of the claim that anti-individualist thought attributions will be unable, despite Burge's protestations, to capture the subject's cognitive perspective. For it is in the exercise of substantive self-knowledge – in critical reasoning – that the subject's cognitive perspective takes form.

In ascribing thoughts to some subject, we want to know not only *what beliefs he has*, but also *how he understands those beliefs*. This is a matter of capturing what, in the subject's point of view, is a reason for what. If there is any substance at all to the notion of a cognitive perspective, this must be it. Thus, I have argued that cognitive perspective is a holistic, not an atomistic notion, which is to say that it essentially ranges over multiple beliefs, and cannot plausibly be anchored in beliefs one by one.

Earlier in this chapter, I argued that the implicitly atomistic methodology of the anti-individualist thought experiments gave a false sheen of plausibility to the claim that attributions of

thoughts containing incompletely understood concepts can serve to capture the subject's cognitive perspective. I argued, instead, that thought attribution must always seek to place the candidate thought in a context of the subject's other beliefs. It is the rational connection between these various beliefs that yields understanding and explanatory power.

In order to gain critical leverage on Burge's assumption that we can make sense of the idea of cognitive perspective as anchored in beliefs one by one, it will be helpful to consider two kinds of cases; on the one hand, cases of what I will call genuine inconsistency, and, on the other, the cases that Burge calls instances of incomplete concept mastery. An instance of the kind of inconsistency I have in mind would be to believe, on the one hand, that arthritis can only ever occur in the joints, but to believe, on the other hand, that the pains I am feeling in my thigh one morning may be a symptom of arthritis. I take it as obvious that, except perhaps in the rare case that someone would affirm an outright contradiction in a single breath, convicting someone of inconsistency is a matter of seeing the problematic attitude in the context of what else we know about this person's attitudes. That is, we cannot convict someone of inconsistency by looking at his beliefs one by one, but only by looking at larger sets of the subject's beliefs.

I venture the following two claims: first, as regards cognitive perspective, someone who is genuinely confused in this way precisely does not have a cognitive perspective on the matter at hand, or at the very least, he does not have a unique cognitive perspective. Similarly for self-knowledge: this person, in a fundamental sense, does not know what he is thinking, if only because there is no one thing that he is thinking. In the case of a genuinely inconsistent subject, then, the dual constraints of capturing the subject's cognitive perspective and preserving self-knowledge do not apply in any straightforward way. But we can only know this by considering a larger set of the subject's beliefs, not by looking at his beliefs one by one.

Contrast this now with what Burge calls cases of incomplete concept mastery. The question I want to press is whether an anti-individualist can make meaningful sense of a distinction between cases of genuine inconsistency and cases of putative incomplete concept mastery. Some critics (notably, Akeel Bilgrami)⁵⁵ hold that the attribution of thoughts such as “I have arthritis in my thigh” would *ipso facto* be an attribution of an inconsistency to the subject, thereby violating the principle of charity. I think this objection needs to be given a rather different spin if it is to be effective. This argument truly *would* beg the question against anti-individualism, in just the way that Burge claims: anti-individualism *presupposes* that one can possess a concept without fully mastering it. That is, in order to possess and employ the concept *arthritis* one need not know – and might actually be wrong about – the definitional truths about arthritis. If this is right, one need not know that arthritis can only occur in the joints to be thinking with the concept *arthritis*, and so it follows that an attribution of the thought “I have arthritis in my thigh” need not be an attribution of a self-contradiction.⁵⁶ This direct line of attack simply denies that premise outright, and so cannot constitute an argument against anti-individualism.

I think that a more circumspect strategy is required to show what is wrong with such attributions. A better strategy, I believe, is to ask instead how the anti-individualist proposes to distinguish between cases of incomplete concept mastery and what I just now described as cases of genuine inconsistency. Distinguishing between these cases is absolutely vital, for as Burge and we both agree, the subject with the putative incomplete concept mastery is not being inconsistent in any straightforward sense, and so his case should not be assimilated to that of the genuinely confused person.

⁵⁵ Bilgrami, “A Trilemma for Redeployment,” “Can Externalism be Reconciled with Self-Knowledge?”

⁵⁶ See section 3.2 for more.

Now, the only way I can think of to distinguish between the two kinds of cases is to widen our focus and look at the context supplied by the rest of each subject's beliefs. Unlike those of the genuinely confused person, the beliefs of the person putatively suffering from incomplete concept mastery form a coherent whole. There is a certain rational systematicity to the way in which he applies the problematic belief which we cannot find in the case of the genuinely inconsistent person. In other words, the subject of Burge's thought-experiments *does* have a cognitive perspective on the matter in a way that I argued that the inconsistent subject does not. Moreover, the subject of Burge's thought-experiments possesses substantive self-knowledge, not just the deflationary self-knowledge that the inconsistent person can claim. That is, he can draw all manner of inferences from his belief. For instance, we can imagine that he arrives at his belief that he has arthritis in his thigh through a piece of rather sophisticated reasoning. In this sense, he arrives at a false belief by way of what we ought to recognize as a rational process of deliberation.⁵⁷

The problem is that the inferences that he commits himself to are not ones that we can properly appreciate the rational force of if we persist in saying that it is the concept *arthritis* that he is thinking with. And if we cannot see the rational force of his reasoning, then we cannot claim to have captured his cognitive perspective. I remarked earlier that capturing the subject's cognitive perspective is a matter of appreciating what, from that subject's point of view, is a reason for what. It should be clear that for our subject, feeling a characteristic pain in his thigh one morning *is* a reason to think that his arthritis might have spread beyond his joints. By contrast, it is emphatically not a reason for his doctor to think the same. Thus, the patient and his doctor are clearly thinking about this "in different ways," and so we must conclude that they inhabit different cognitive perspectives on the matter. If concept and cognitive perspective go hand in hand in the way that

⁵⁷ This raises the question of how one might retain the sense that the belief he arrives at is indeed a false belief, if we take my route and insist that he is thinking of the matter from a different cognitive perspective than, say, the experts in his community. I will return to this question in ch. 5.

Burge assumes, then we should say that the patient is not thinking of this using the concept *arthritis* at all, and so, *a fortiori*, that he is not thinking of it with an incompletely understood concept *arthritis*. If we insist that it is indeed the concept *arthritis* that he is thinking with, then we will be unable to tell the difference between him and the genuinely inconsistent person. And so we will have foregone our possibility of seeing his thoughts and deliberations as the expressions of a rational point of view, which it was precisely the purpose of the cognitive perspective metaphor to capture.

It should be easy to see how the failure of anti-individualist thought-attributions in capturing and conveying the rational force of the subject's reasoning will also affect its ability to feature in explanations of that subject's actions.

The kinds of action explanations we are interested in here take the form of rationalizations.⁵⁸ While Burge has little to say about the character of action explanation, we have every reason to believe that he concurs with this basic model. But if action explanations are (or centrally involve) rationalizations, it should be clear why an argument to the effect that anti-individualist thought-attributions cannot capture the rational force of a subject's reasoning can also serve as an argument to the effect that they cannot aid us in action explanations.

A different angle on the problem will essentially confirm this diagnosis. Action explanations operate on the consequences of the beliefs we attribute to the agent. Predictions of what an agent will do work on assumptions of the general form "if one believes that *p*, one ought to do *A*, circumstances permitting." When the agent performs the action, we have simultaneously an explanation of the action and a reason to believe that we attributed the right attitude to the agent to the agent in the first place. But, of course, these practical consequences are just the kinds of things that, according to anti-individualism, the subject may fail to know, even as he can still be rightly

⁵⁸ This view is, of course, mostly associated with a series of articles by Donald Davidson, starting with "Actions, Reasons, and Causes." However, it is important to note that Davidson here taps into a philosophical tradition that stretches back (at least) to Aristotle's practical syllogism.

counted as thinking the thought in question. Since the doctrine of incomplete concept mastery undermines the connection between the attribution of a belief-content to some subject and the assumption that the subject will endorse the practical consequences of that belief, anti-individualism thereby severs the reliable connection between attitude-attribution and action explanation. Under the doctrine of incomplete concept mastery, attitude attribution and action explanation cannot have the reliable, evidentiary connections that they have under the model of action explanation as rationalization.⁵⁹

4.8 Closing remarks

I have argued that in order to capture the subject's cognitive perspective we cannot consider the subject's beliefs one by one, but must look at larger sets of beliefs, and the rational transitions between these beliefs that the subject is disposed to make. This is the only way that we can distinguish between the genuinely inconsistent person and the person putatively suffering from incomplete concept mastery. With the genuinely inconsistent person, we will be unable to find any such rational pattern. With the person putatively suffering from incomplete concept mastery we may find that he arrives even at a false belief in a rational fashion, by rationally reflecting on and drawing consequences from his beliefs. Since the belief he thus rationally arrives at is recognizably false, we

⁵⁹ We can see this quite clearly if we conjoin the following two arguments in "Individualism and the Mental": first, "If there were not deep and important connections between propositional attitudes and understanding, one could hardly expect attributions of mental content to facilitate reliable predictions of what a person will say, do, or think" (Burge, "Individualism and the Mental," p. 118); second, "to a fair degree, mentalistic attribution rests not on the subject's having mastered the contents of the attribution, and not on his having behavioral dispositions peculiarly relevant to those contents, but on his having responsibility to communal conventions governing, and conceptions associated with, symbols that he is disposed to use" (ibid., p. 148). Of course, Burge's "to a fair degree" is no doubt intentionally vague, but one cannot escape the sense that any attempt to precisify it would tend to undermine his point. I shall, by the way, have more to say about the idea of a "responsibility to communal conventions" in my next chapter.

ought to conclude that the cognitive perspective from which he reflects on these matters is importantly different from that of, say, his doctor.

Similar considerations apply to the question of self-knowledge. From the anti-individualist perspective the genuinely inconsistent person and the person putatively suffering from incomplete concept mastery will be in exactly the same situation. They both possess self-knowledge in the deflationary sense, insofar as they can both attribute to themselves the thought in the manner prescribed by the cogito-model. But again, there has to be more to it than this. As I have argued, the genuinely inconsistent subject is confused, and so does not know what he is thinking, whether or not he can pass the cogito-test. By contrast, the subject of the anti-individualist thought-experiments does know what he is thinking in a richer sense, in that he is busy engaging in rational reflection, drawing consequences of his beliefs, and so on. But again, this is not a form of self-knowledge that we will be able to appreciate unless we expand our considerations beyond the candidate belief attribution and try to gauge the relation between his several beliefs. But once we do this, we will see that the thought that our subject knows himself as thinking is not the one that the anti-individualist advocates that we should attribute to him.

These observations shed new light on the questions that occupied us throughout chapter 3. Burge supports his argument on the observation that attribution of thoughts containing incompletely understood thoughts is “common practice” in belief attribution. The burden of proof must weigh heavily on the philosopher who wishes to argue that we are wrong in this practice, when the practice seems to serve us so well in our everyday interactions. We can now see what is wrong in this contention. Burge rightly observes, of course, that it is common practice to use the problematic *word* in the attribution. We typically do not, that is to say, go to the effort of coining a new term every time someone misapplies a word. This much is not in question. What is in question, of course, is whether employing a particular word marks an attribution of the *concept* that is conventionally

associated with that word. Burge gives us no reason to think that this is the case. In fact, now that we have drawn the connection between concept and cognitive perspective – with textual support from Burge – we are in a position to turn the tables on anti-individualism: what Burge explicitly says about the matter confirms our view, not his. If, as Burge puts it, we can “simply bear in mind” that the subject thinks the word applies to different things than the rest of us do, then we have *ipso facto* attributed a different concept to him, whether or not we persist in using that particular syntactic shape in our attribution. Burge’s claim that it is the conventional concept we attribute, even though we make room for the fact that the subject “understands” the concept differently than the rest of us do, can only be understood as a recipe for decisively dissociating concepts from “ways of thinking” about objects⁶⁰ and thus from cognitive perspective and the explanatory tasks that fall on that notion.

Burge can, of course, continue to insist that the conventional concept is involved in some way or other in the attribution. But if so, then none of the explanatory burden that we have associated with the notion of cognitive perspective is carried by that concept. It is now the subject’s “understanding” of that concept which enters into assessments of his rationality and explanations of his actions. As we noted in ch. 3, the appeal to the conventional concept becomes an idle wheel, deprived of any real explanatory purpose.

⁶⁰ Adapted from Burge, “Other Bodies,” p. 92.

5

NORMATIVITY

5.1 Introductory

I ended chapter 4 with the following claim: if a serious consideration in the practice of thought attribution is to capture the subject's cognitive perspective – it is only our success on this front which will enable us to use our attributions to appreciate and convey the force of the subject's reasoning and to explain his actions – then the kind of incompletely understood conventional concepts that Burgean anti-individualism works from can have no serious role to play. We may, of course, continue to use the *word* that is conventionally associated with that concept in our attributions. There may even be some rarefied technical sense in which we could say that the concept in question is 'involved' in the attribution. (As we shall see, we may continue to say that the public concept embodies the knowledge by reference to which we judge the subject's thinking to be in error.) But so long as we recognize that the subject's *understanding* of the word departs from its conventional meaning in significant and systematic ways, then we have *ipso facto* already attributed a different concept to him. To say, as Burge does, that we can continue to operate with a literal interpretation of the subject's words but simply "keep in mind" his misunderstanding of those words, is to gesture toward a distinction that we will be unable to make sense of, namely a distinction between how a person is thinking about some subject matter and what concept he is thinking with. In such cases, the anti-individualist covertly relies on the pattern of the subject's misunderstanding to do all the explanatory work for him. To "keep in mind" the subject's

misunderstanding in our attribution is just what it is to attribute a different concept, a concept that matches or reflects his misunderstanding. It could not be otherwise if concept and cognitive perspective are connected in the way that we have been led to assume, and thought attributions are to be saddled with the explanatory burden that I laid out in ch. 3.

Apart from some remarks in ch. 3, I have throughout most of this dissertation evaded one overarching topic of great importance to Burge and many other thinkers in the post-Fregean and post-Wittgensteinian traditions,¹ namely *normativity*. Proper appreciation of the importance of normativity to the study of mind and language might seem to offer a rough and ready sketch of an answer to the challenge I presented at the end of chapter 3: norms are fundamental to mindedness and the kinds of activities that properly go along with mindedness. The function of the appeal to the conventional concept in the kinds of cases from which Burge is working, is to preserve the normative grip that such attributions give us on our subject's thinking. The anti-individualist, then, rightly seeks to preserve the following intuition: we should not construe our subject as thinking correctly, albeit with a non-standard concept. But the anti-individualist goes on to argue that it is only by invoking such conventional, normatively infused concepts in our attributions that we can sustain our sense that there is something *wrong* with the subject's thinking.

This opens up a topic of much greater scope and importance than I can fully do justice to here. Nonetheless, in the brief space available I will try to outline what I think is right and what is wrong about this picture. I will argue that norms are indeed important to our notion of mindedness and to the theory of thought attribution, but that they cannot function in anything quite like the way that the anti-individualist supposes.

¹ See, not least, the debates on "rule-following" that arose in the wake of Saul Kripke's book *Wittgenstein on Rules and Private Language*. Due to space limitations, I will have very little to say about Wittgenstein and the concerns specific to the rule-following debates in this chapter (although see my brief remarks in ch. 3). Instead I will concentrate specifically on the invocation of normativity-concerns to bolster the case for anti-individualism.

5.2 Putnam on the division of linguistic labor

Before looking in detail at how the theme of normativity develops in Burge's anti-individualism, it will be helpful to consider Putnam's hypothesis concerning the division of linguistic labor. This idea has had an enormous influence not only on the development of anti-individualism, but also on philosophical discussions of the social dimension of language in general. Putnam introduces the idea of a social division of linguistic labor by way of an alternative "*non-science-fiction*" thought-experiment which has become almost as famous as that involving Twin Earth:

Suppose you are like me and cannot tell an elm from a beech tree. We still say that the extension of 'elm' in my idiolect is the same as the extension of 'elm' in anyone else's, viz., the set of all elm trees, and that the set of all beech trees is the extension of 'beech' in *both* of our idiolects. Thus 'elm' in my idiolect has a different extension from 'beech' in your idiolect (as it should). Is it really credible that this difference in extension is brought about by some difference in our *concepts*? My concept of an elm tree is exactly the same as my concept of a beech tree (I blush to confess). (Putnam, "The Meaning of 'Meaning'," p. 226)

We note first a potential source of terminological confusion:² Putnam says that his concepts *elm* and *beech* will, in a case like this, be the same. This is precisely not what Burge's analysis entails. On the anti-individualist view, we can tell that these are different concepts, precisely because they have different extensions. Instead, if we are to follow Burge here, we should say that Putnam's *grasp* of these concepts is so deficient or incomplete that his understanding or *conception* of each concept is the same.³

² I will argue in section 4 of this chapter that more than a mere terminological confusion is at stake. But before going on to show this, I will try to portray anti-individualism at its strongest and most plausible.

³ Strictly speaking, this too would be an oversimplification, since Putnam would presumably know that the two words refer to different things, even if he cannot state the criteria of application or perceptually discriminate between the two kinds of trees. However, I will let this pass here.

The point of the elm-beech thought experiment is to highlight the fact that Putnam can refer differentially to elms and beeches even though he himself could not tell which is which. It is this observation that supports the hypothesis concerning the social division of linguistic labor. The fact that Putnam can refer differentially to elms and beeches without knowing the difference between them demands an explanation. The explanation that Putnam offers is that he can rely on others more knowledgeable in matters arboreal to fix the reference for him. In a word that has become commonplace in these discussions, Putnam can *defer* to the expert as regards the meanings of his words. Thus, for example:

everyone to whom gold is important for any reason has to *acquire* the word ‘gold’; but he does not have to acquire the *method of recognizing* if something is or is not gold. He can rely on a special subclass of speakers. (Putnam, “The Meaning of ‘Meaning’,” pp. 227-228)

And more generally:

Whenever a term is subject to the division of linguistic labor, the ‘average’ speaker who acquires it does not acquire anything that fixes its extension. In particular, his individual psychological state *certainly* does not fix its extension; it is only the sociolinguistic state of the collective linguistic body to which the speaker belongs that fixes the extension. (Putnam, “The Meaning of ‘Meaning’,” p. 229)

In brief, then, in complex societies at least,⁴ to be a member of a linguistic community is to be a member of a cooperative unit which functions in such a way that a member might successfully refer to a certain class of objects by the use of a word, even though he himself is unable to pick out the

⁴ Both Putnam and Burge speculate that extremely simple societies might not yet have developed such division of linguistic labor. See Putnam, “The Meaning of ‘Meaning’,” p. 229; Burge, “Postscript to ‘Individualism and the Mental’,” p. 175.

class of objects in question. He can simply defer to the experts on the objects in question, whose knowledge does fix the extension of the word he uses.⁵

We can see how the Burgean anti-individualist cannot simply appropriate these points as-is to support his theory. He faces three distinct tasks. First, he must show that the division of linguistic labor operates not only on natural kind terms (as Putnam seems to hold), but on a broad range of social kind terms as well.⁶ Second, he must show why these *prima facie* merely semantic matters should also impact on psychological attribution.⁷ Finally, he must cast light on the putative normative significance of this division of labor. That is, he must show that the possibility of deferring is not merely a courtesy extended to the speaker as a member of a linguistic community; there is also some sense in which we are bound by the norms that are embodied in the conventions governing the words we are disposed to use.

5.3 Normativity in “Individualism and the Mental”

Let us turn to see how Burge might meet these challenges. The theme of normativity does not permeate the argument of “Individualism and the Mental” quite to the extent that subsequent discussions might lead one to assume. Explicit invocations of normativity to bolster the article’s

⁵ It is a fine question whether this can be right, even by Putnam’s own lights. It seems clear enough that Putnam should precisely not want to say that expert usage fixes the extensions of, say, our scientific terminology, lest we also conclude that scientists prior to the Chemical Revolution (who could not discriminate between water and XYZ) were not talking about water when they said ‘water.’ It might seem that the case is importantly different with so-called social kind terms, where it seems plausible to say that the extension of a given term (say, ‘contract’) cannot be other than what the experts agree that it is. I think this view, too, leads to certain revealing problems, which I will investigate in section 6 of this chapter.

⁶ Bear in mind here how expansive is the set of kinds of terms to which Burge thinks the thought-experiment might apply: “We could have used an artifact term, an ordinary natural kind word, a color adjective, a social role term, a term for a historical style, an abstract noun, an action verb, a physical movement verb, or any of various other sorts of words” (Burge, “Individualism and the Mental,” pp. 106-7). Note, by contrast, that Putnam specifically mentions ‘chair’ as a term that does not evince a division of linguistic labor (Putnam, “The Meaning of ‘Meaning,’” p. 228). ‘Chair’ is among Burge’s examples in “Individualism and the Mental.”

⁷ In opposition to Putnam’s claim that it would be ‘absurd’ to say that there would be any psychological difference between the two cases. See Putnam, “The Meaning of ‘Meaning,’” p. 227.

main conclusion occur twice in the text. Relatively early in the article, in the context of a first discussion of whether the kind of apparent conceptual confusion on display in the thought-experiments should occasion “automatic reinterpretation,” Burge writes as follows:

People are frequently held, and hold themselves, to the standards of their community when misuse or misunderstanding is at issue. One should distinguish these cases, which seem to depend on a certain *responsibility* to communal practice, from cases of automatic reinterpretation. (Burge, “Individualism and the Mental,” p. 119)

There certainly are cases, then, in which a subject’s utterance of a particular word should not lead us to attribute the concept conventionally associated with that word. But these are, as it were, exceptions that demonstrate the underlying rule. This rule stipulates that members of a linguistic community are responsible to certain communal standards or norms. The exceptions just noted mark the leeway that we grant, for instance, to foreigners and children as they are in the process of being inducted into the language community. They are in the process of acquiring the language, but have not yet achieved the requisite degree of mastery upon which the responsibility to the communal standard is premised, and which marks full-fledged membership in the linguistic community. Thus, we can note two things: (i) a certain level of mastery or competence is required for talk of responsibility to get any real purchase, but presumably not complete mastery, since (a) very few of us have complete mastery of any large range of the concepts that nonetheless appear to be at our disposal, and (b) if complete mastery were required, then the current discussion would be pointless, since it concerns precisely how to deal with cases of incomplete mastery; (ii) Burge here appears hesitant on a key point, namely whether what triggers the responsibility is merely a subject’s being held by his peers to the communal standard, or whether he also needs to hold himself to such a standard.

A fuller discussion of the theme of normativity occurs toward the end of the article. Here Burge writes:

Crudely put, wherever the subject has attained a certain competence in large relevant parts of his language and has (implicitly) assumed a certain general commitment or responsibility to the communal conventions governing the language's symbols, the expressions the subject uses take on a certain inertia in determining attributions of mental content to him. In particular, the expressions the subject uses sometimes provide the content of his mental states or events even though he only partially understands, or even misunderstands, some of them. Global coherence and responsibility seem sometimes to override localized incompetence. (Burge, "Individualism and the Mental," p. 147)

This essentially maintains the line adopted in the earlier discussion, albeit with two specifications worthy of note: (i) Burge here seems to allow that the responsibility that supports the invocation of normativity may rest on a merely *implicit* assumption of such responsibility on the part of the subject. The implicit assumption of responsibility comes, it would seem, as part and parcel simply of being inducted into a language community in the first place;⁸ (ii) to make his point, Burge relies on terms such as 'inertia' and 'override'. I will later say how these terms might reveal something important about Burge's thinking about these matters.

Further:

⁸ In much the same sense, presumably, as induction into any community will bring with it an implicit assumption of responsibility to the norms that govern that community. Why, one might ask, should we speak here of the individual *assuming* this responsibility (however much the assumption is merely implicit; however much it must have been undertaken prior to the age at which we are usually held fully accountable for our actions), rather than merely say we are subjected to these norms as we are inducted into society? This points in the direction of a fascinating discussion in political philosophy, dating back at least to Hobbes's *Leviathan*. The founding insight of the Hobbesian revolution in political thought is that there cannot be any true norm-giving authority to which the subject has not consented to being subjected. Since it is extremely doubtful that we can model political authority in any extant society on the *actual* such consent of all its members, subsequent writers would rather say that it rested on an *implicit* such consent. The implicit consent was taken to rest on some good the pursuit of which the subject could not rationally forswear, yet which could not be had except through membership in the political community. In turn, of course, this political community could not exist but for the subjection of its members to a certain set of norms. If this analogy with social contract theory is appropriate, then one would be tempted to say that anti-individualism, at least as standardly construed, must constitute something of a Rousseauian moment, emphasizing the binding force and transformative powers of the General Will. I think there is something to this, and moreover, that a Rousseauian moment in analytic philosophy of mind and language is long overdue. Nonetheless, I would maintain that Burge draws the wrong consequences from it.

What I want to stress is that to a fair degree, mentalistic attribution rests not on the subject's having mastered the content of the attribution, and not on his having behavioral dispositions peculiarly relevant to those contents, but on his having a certain responsibility to communal conventions governing, and conventions associated with, symbols that he is disposed to use. It is this feature that must be incorporated into an improved model of the mental. (Burge, "Individualism and the Mental," p. 148)

The clause "to a fair degree" is, of course, inherently imprecise. I think we should read it as signaling a concession that the binding force of norms is not absolute: not merely are children and foreigners excused from the inertial force of such norms; sometimes even the performance of an otherwise competent subject falls so far short of the norms that we must doubt whether literal interpretation can yield the right attribution.⁹ In short, literal interpretation is not automatic either, even for those who we should ordinarily say are bound by the norms Burge speaks of. There will be, we may presume, a set of cases in which neither the literalist strategy nor the reinterpretation strategy can claim automatic privilege, and it must remain an open question which would yield the right interpretation on the occasion. Enlisted on the literalist side is responsibility to communal norms and the maximization of interpersonal comparability. Enlisted on the reinterpretivist's side is the individual's behavioral dispositions and verbal testimony concerning what he takes the extension of his words to be. Burge's language of 'overriding' and 'inertia' suggests that there is a real conflict here. I agree with this: indeed, my point throughout has been that there is no way to resolve this conflict in favor of literalism without giving up on the idea that our attributions of attitudes can play the systematic role in rationalization and action explanation that we have sought for them. (As we have seen, Burge will claim that there is one segment of the subject's behavior which favors the

⁹ Cf. Burge, "Individualism and the Mental," p. 118: "If there were not deep and important connections between propositional attitudes and understanding, one could hardly expect attributions of mental content to facilitate reliable predictions of what a person will say, do, or think."

literalist strategy, namely his disposition to self-correction. In sections 3.5.4 and 3.5.5, I argued that, contrary to Burge, this does not actually support the literalist strategy. I will revisit the point below.)

However, there is also another, more surprising sense in which the normative force of linguistic conventions is not absolute. Quite in contrast with how things are usually taken to be with the moral and legal conventions that govern our community at large, individuals may legitimately ‘opt out’ of their linguistic community by simply renegeing on their responsibility to these norms.

A person born and bred in the parent community might simply decide (unilaterally) [...] to fashion his own usage with regard to particular words, self-consciously opting out of the parent community’s conventions in these particulars. In such a case, members of the parent community would not, and should not, attribute mental contents to him on the basis of homophonic construal of his words. Here the individual’s intentions or attitudes toward communal conventions and communal conceptions seem more important than the causal antecedents of his transactions with a word. (Burge, “Individualism and the Mental,” pp. 147-148)

In brief, governance by norms is the default situation, but we are free to ‘opt out’ by some more or less explicit decree of our own. It is telling, however, that Burge holds that reinterpretation in such cases is motivated by considerations of the subject’s intentions and attitudes *towards these communal conventions and conceptions*, rather than considerations of his intentions and attitudes toward the worldly phenomena that we should otherwise intuitively say that he is thinking of. In light of assumptions Burge and I share concerning the priority of *de re* attitude attributions, I find this qualification quite puzzling, but will not pursue it further here.

5.4 Elms and beeches redux

I said above that a juxtaposition of Putnam and Burge on these matters must take care to avoid a potential source of terminological confusion. According to Putnam, his concept *elm* and his concept *beech* are the same concept. Putnam's usage here ties 'concept' to the subject's perceptual-discriminative abilities. Putnam presumes that these discriminative abilities indicate something about the psychological state he is in when thinking about elms or beeches; viz., that he is in the same psychological state in both cases. Since his words 'elm' and 'beech' nonetheless have different extensions, Putnam concludes that his psychological state on the occasion does not determine (in the Fregean sense) the extensions of these words.

I have agreed with Burge that this separation of semantic ascription and psychological attribution is inherently problematic. It is incompatible with our commonsense supposition that correct apprehensions of what a subject is saying on some occasion provide material for psychological attributions that can reliably play a certain role in explanations of his actions. In order to retain the connection between semantic ascription and psychological attribution, Burge sides with Putnam's analysis of what we should take the meanings of each of the speaker's words to be (viz., that which conforms to the extant linguistic conventions), and concludes accordingly that the subject is thinking with different concepts when using the two words, even though his grasp of each concept is incomplete (in fact, so incomplete that he could not differentiate between them).¹⁰

But it should be obvious by now that this is not the only way to retain the intuitive connection between semantics and psychology. We could also agree with Putnam that whatever we would want to say about the speaker's concepts, it would have to be anchored in his discriminatory capacities, his inferential and behavioral dispositions, etc. Taking this line will require us to say that

¹⁰ Again, it can hardly be overemphasized how awkward such constructions are. It is *kinds of trees* that he is unable to distinguish between, not concepts.

the proper interpretation of the speaker's words will not (in all cases) be the interpretation that accords with the semantic conventions. But this is just what I have been arguing all along.

So it is clear that Burge cannot simply appropriate the conclusion of Putnam's thought-experiment to support his own argument. In fact, closer inspection reveals that, if stripped of the assumption that semantic content and psychological content are categorically distinct matters, the considerations that underlie the thought-experiment would support my conclusion, not his. The reason Putnam's conclusion preserves common sense – the conclusion that we should take Putnam's utterances of the words 'beech' and 'elm' to have the semantic values that they have when the experts utter them (in spite of the fact that he does not know the difference) – is precisely the presumption that we are *not* to take the semantic interpretation to reveal Putnam's psychological content.¹¹ *If* we took correct semantic interpretation to bear on psychological attribution in the way that both Burge and I suppose, then it is quite clear that the conclusion we should draw from Putnam's thought experiment would be different.

Thus, if we run the thought experiment again, but without the assumption that semantic content and psychological content can come apart, we will see that the anti-individualist diagnosis is intuitively false. It is true, perhaps, that in speaking in vague generalities about elms or beeches ("elm trees are magnificent"), the expert will probably take Putnam's word 'elm' to refer to elms, and so will ascribe to Putnam the belief that elm trees are magnificent. (I can allow this even in a case where the expert knew of Putnam's inability to perceptually distinguish between elms and beeches. But this is only because the semantic content of the utterance is suitably unspecific and the expert lacks any contextual clues as to what else Putnam might be thinking of.) By contrast, consider Putnam and the expert taking a leisurely stroll through a beech forest, and Putnam says "elm trees are magnificent."

¹¹ Recall Putnam's claim ("The Meaning of 'Meaning,'" p. 227) that it would be "absurd" to say that his psychological state in the two cases is one bit different.

In this context, the expert has contextual clues that would dictate that he interpret Putnam's "elm trees" as referring to beech trees. Taking Putnam to be talking about elm trees on the occasion would be inexplicable. Knowledge of Putnam's inability to distinguish elms from beeches thus provides the expert with grounds to override the linguistic conventions on the occasion. (Assume, by contrast, that Putnam goes on to say "much nicer than these paltry beech trees, at any rate," and proceeds to tell the expert how he has been reading up on the local flora. Now the case would be different again.)

A further example might help drive the point home. Say Putnam and I are jointly acquainted with two identical twins Peter and Paul, and I am aware of Putnam's inability to tell them apart. Putnam tells me he thinks Paul is a nice guy. I have no particular reason to suppose that Putnam might be confused on this occasion, and so I take him to be referring to and thinking about Paul. By contrast, say Putnam tells me that the reason he thinks Paul is a nice guy is that he ran into him the other day while he was doing volunteer work for a local charity. I happen to know, however, that Paul is abroad, whereas Peter is in town. In this context, I understand Putnam's "Paul" as referring to "Peter."

The conclusions I propose here certainly appear to contradict those drawn by Putnam and Burge. The conflict with Putnam, however, is merely apparent. Putnam's conclusion preserves common sense only on the assumption that the semantic ascription (which is all Putnam takes to be at stake in the thought-experiment) does not license the corresponding psychological attribution. This is precisely the assumption that Burge and I put to the question. Thus, the conflict with Burge is real. His view must be that we might, even in my revised version of the elm-beech example, take Putnam to be thinking and talking about whatever accords with the literal interpretation of his

words.¹² But what could ground such a claim? This is precisely what the invocation of the normative force of linguistic conventions is supposed to accomplish.

5.5 What is the explanatory force of linguistic norms?

We saw that on Burge's view, the responsibilities that full-fledged members of a linguistic community have toward the conventions governing their language make it so that "the expressions the subject uses take on a certain inertia in determining attributions of mental content to him." Thus, considerations of "[g]lobal coherence and responsibility seem sometimes to override localized incompetence" (Burge, "Individualism and the Mental," p. 147). These metaphors of 'inertia' and 'override' may seem drastic, chosen for dramatic effect rather than philosophical cogency. On the contrary, I think they perfectly capture the drift of the anti-individualist position on these matters. Closer inspection will reveal just how implausible this position is, and, in particular, how it must involve a covert rejection of the idea that psychological attributions play a fundamental role in assessments of the subject's rationality and explanations of his actions.

Metaphors such as 'inertia' and 'override' bespeak tension or resistance. So we must ask what it is that is being overridden by this inertial force, what offers the resistance. I think the only plausible reading is that what would be overridden in these cases is the evidence that suggests that the subject was thinking some other thought than what a literal interpretation of his words yields. If

¹² Subject, of course, to an indefinite number of qualifications. Thus, Burge will say that "[l]iteral interpretation is *ceteris paribus* preferred" (Burge, "Individualism and the Mental," p. 116). He never says much about how the *ceteris-paribus* clause is to be filled in, and it is hard to see how it could be filled in without jeopardizing the force of the argument. Thus, we might assume that literal interpretation is preferred *unless* there are obvious contextual clues as to what else he might be thinking of; *unless* his thinking is so far out of tune with the conventions that a literal attribution could not help explain his actions, etc. But if these are the kinds of factors that hide under the *ceteris-paribus* clause, then Burge has merely named an epistemic predicament besetting the interpreter. He could not assume, without begging the question, that in cases where there were no, say, obvious contextual clues as to what else a subject might be thinking of, the subject's thought contents are precisely those that are captured by the literal interpretation of his words.

this inertial force seems arbitrary or aimless, the anti-individualist will assure us that it is not. It is *appropriate* to allow the inertial force to run its course, because it is the expression of a legitimate social norm.

This is certainly a possible line to take. But it does demand an honest recognition that in such cases, the attributions yielded by the literal interpretation will not systematically explain the subject's actions in the required way, because such attributions do not capture the force and character of his reasoning in these cases.¹³ If this is so, then there should be some other content attribution (even if it is not captured by any sentence of his conventional language on its literal interpretation) that would explain the action. Why *this* should not be the attitude content that our attribution is supposed to capture is not clear. Minimally, this does seem like an implicit admission that we would have to keep two books in such cases; one in which we note the content that enters into explanations of the subject's actions, and another in which we note the content that accords with the linguistic conventions.¹⁴ This strategy, while no doubt consistent, must still answer the question in what sense what is noted in this second book should count as *mental* content.

This concession might be seen as a preface to a theoretical reconsolidation on the part of the anti-individualist, which might proceed in something like the following way: the point of attitude attribution is not *only* to explain the subject's actions and assess his rationality. True, that is one aim. But another aim is to facilitate information flow between speakers and to aid them in sundry practical cooperative endeavors. Sometimes these two desiderata may come apart, as when the subject's thinking evinces the kind of conceptual confusion that we see in the thought-experiments.

¹³ Bear in mind that it will not do for Burge to say that in order to explain the subject's pattern of reasoning and his actions we simply "keep in mind" his aberrant beliefs about the word we use in our attribution. For then we would already have attributed to him a different concept, a different 'way of thinking.'

¹⁴ Akeel Bilgrami makes this point forcefully in his *Belief & Meaning*, ch. 2.

Since the subject is in some real sense at fault for his confusions (he is, after all, responsible to the norms), we can rightly ‘override’ his confusion and treat him as though he were competently thinking the thought in question. In this we assert a privilege stemming from the role that such norms have precisely in grounding the possibility of such information flow and cooperative endeavors.

In this vein, Burge writes:

Symbolic expressions are the overwhelmingly dominant source of detailed information about what people think, intend, and so forth. Such detail is essential not only to much explanation and prediction, but also to fulfilling many of our cooperative enterprises and to relying on one another for second-hand information. Words interpreted in conventionally established ways are familiar, palpable, and public. They are a common coin, a relatively stable currency. These features are crucial to achieving the ends of mentalistic attribution just cited. They are also critical in maximizing interpersonal comparability. And they yield a bias toward taking others at their word and avoiding *ad hoc* reinterpretation, once overall agreement in usage and commitment to communal norms can be assumed. (Burge, “Individualism and the Mental,” p. 149)

A few quick remarks are in order before we embark on a closer analysis of this passage. First, Burge says that these claims hold when “overall agreement in usage and commitment to communal norms can be assumed.” I have already said a few things about such “commitment” or “assumption of responsibility” (section 5.3), and refer the reader back to that discussion. Second, “overall agreement in usage” must be construed as a rather loose notion, since we are precisely addressing the question of how to deal with *disagreement* in usage. Finally, I remind the reader that we should in no way take Burge’s injunction against ‘*ad hoc* reinterpretation’ to suggest that such reinterpretation would be *arbitrary*, as in not grounded in philosophical principle or concerns of a more general nature.¹⁵

¹⁵ For further remarks on this, see section 3.6.3.

Moving on to the substance of the quote, we note that Burge invokes the aims of prediction and explanation. But as I have stressed several times now, the concern we are expressing at this point is precisely that words interpreted in their conventional way cannot be employed in predictions and explanations, *if the subject systematically misunderstands the meanings of those words*. So this cannot be what grounds the normative bias in favor of literal interpretation. Thus, I want to spend the rest of this section investigating whether these other aims that Burge mentions – maximizing interpersonal comparability, guiding us in our cooperative practical endeavors, facilitating information flow between speakers – can supply the missing grounds. Could it be that these other aims still require literal interpretation, such that these concerns can “override” the concern with prediction and explanation on occasions where they cannot be jointly satisfied?

As I will argue presently, this cannot be right either. Moreover, I think we should be deeply suspicious of the idea that there can be any real conflict between the kind of understanding that we seek for rationalization and action explanation, on the one hand, and the kind of understanding that grounds our various cooperative endeavors, on the other.

Let us begin with the question of acquisition of knowledge by the testimony of others. Little reflection is required to get a sense of just how important this is to the cognitive lives of human beings. Moreover, the topic is obviously of great importance to anyone who wishes to investigate the social dimension of mindedness, as both Burge and I do. Finally, regardless of one’s stance on the question of whether language is required for propositional thought, there cannot be any doubt that the social practice of giving and taking knowledge by testimony depends on the development of

a complex language, and that this in turn is essential to the development of human society as we know it.¹⁶

The question before us now, however, is of a different and much more local sort: does privileging literal interpretation in cases of putative incomplete concept mastery aid or in any other way enhance our possibility of giving and taking knowledge by the testimony of others? (No doubt, testimonial belief acquisition is hugely facilitated by *the fact that* people can largely agree on the meanings of words and use them accordingly. But this is not what is in question here, which is rather how we should understand people when they do not.) In particular, we should ask whether testimony could be as reliable and rewarding a source of knowledge as it is if incomplete concept mastery is as widespread as Burge holds and we must maintain a firm bias in favor of literal interpretation in the encounter with such incomplete concept mastery.

Consider, for instance, what Burge in his landmark article “Content Preservation” calls the Acceptance Principle: “*A person is entitled to accept as true something that is presented as true and that is intelligible to him, unless there are stronger reasons not to do so*” (Burge, “Content Preservation,” p. 467). Among the countervailing reasons that might undermine local employment of the Acceptance Principle could be well-founded knowledge contrary to the testimony, knowledge that the testifier has deceived us on previous occasions, and so on. The problem I want to highlight here, however, is that if incomplete concept mastery is as widespread as Burge says that it is, then we already have a weak but quite general reason *not* to take as true that which is presented to us as true and which seems intelligible to us. The reason is simple: people who do not know what their words mean cannot always be taken to be reliable testifiers.

As I said, this countervailing reason is quite weak and does nothing to undermine the social practice of giving and taking testimony at large. In this sense, acknowledging the ubiquity of the kind

¹⁶ Cf. Jared Diamond, *The Third Chimpanzee: The Evolution and Future of the Human Animal*, pp. 54-56.

of verbal confusion that Burge calls ‘incomplete concept mastery’ is compatible with endorsing the practice of testimony (which is at it should be). But that is not the point: assume instead that we know of *this* particular person that he is prone to *this* particular verbal confusion. This knowledge would constitute a strong reason *not* to accept his testimony, if the testimony includes the word in question. In my view, this diagnosis is only partly right, however, and the way in which it is wrong can be seen to undermine Burge’s argument. As I shall explain shortly, we may still, in such a case, have strong reasons to accept his testimony, but not as testimony to the truth of the proposition yielded by the literal interpretation of his words.

Burge’s view might work, I think, if the practice of testimony were restricted to passing from teacher to student, from someone comparatively more knowledgeable to someone comparatively less knowledgeable. This would fit nicely with Burge’s assumptions concerning the ubiquity and importance of conceptual deference, and would go some way toward explaining why we should say that the student is “thinking the same thought” as his teacher, in spite of the fact that former’s grasp of the involved concepts is seriously incomplete.¹⁷ But this would be an unjustified *ad hoc* restriction. Testimonial knowledge passes as naturally and effortlessly upwards and sideways as it does downwards.

In brief, Burge’s view might work well if we assume that those whose thinking evinces incomplete concept mastery are always or ordinarily on the receiving end of testimonial belief-acquisition. But this would neglect the fact that if incomplete concept mastery is as widespread as he thinks, then they will also be on the giving end of that relation. This creates a problem for the anti-individualist. For in many of the cases we are considering, there *is* a belief that I could justifiably take over from my informant, but it is *not* the belief that is captured on the literal interpretation of his

¹⁷ Testimonial belief acquisition is a matter of “preserving content” across individuals (cf. Burge, “Content Preservation”). It seems, in short, that the unique epistemic entitlements that come with testimonial belief acquisition depend on its being ‘the same content’ on both ends of the relation.

words. *That* belief content I have no justification for taking over if it is known to me that my conversation partner misunderstands his word in certain systematic ways.

Consider, for example, Putnam returning from a trip to Norway and enthusiastically relating that “there are lots of beautiful elm trees in Norway.” Knowing of Putnam’s inability to distinguish between elms and beeches, I am not justified in acquiring the belief content *there are lots of beautiful elm trees in Norway*. But there is some belief that I could justifiably acquire from this bit of communicative interaction, perhaps that *there are lots of beautiful trees – either elms or beeches – in Norway*. Suppose I know that there are not, in fact, many elm trees in Norway: it now seems that I could justifiably acquire the belief *there are lots of beautiful beech trees in Norway* from Putnam’s testimony. But according to the anti-individualist, this is not the belief content that we should ascribe to Putnam. Thus, if we maintain a firm bias in favor of literal interpretation, it seems that we would be at a loss to explain how I can be justified in acquiring this belief second-hand through Putnam. But this is an absurd predicament, since it was precisely in part the possibility of such testimonial belief acquisition that the prerogative of literal interpretation was supposed to illuminate. If belief acquisition by testimony requires that it be a unique belief content which is imparted from speaker to hearer, then we must say either that I was not justified in acquiring this or any other belief from Putnam’s testimony (and so on for all the cases in which incomplete concept mastery is involved and is relevant to the case at hand), or that the literal interpretation does not capture what Putnam was thinking, whereas the reinterpretation does. The fact that I do seem to be justified in acquiring the belief attributed on the reinterpretation strategy would certainly favor the latter.

This lesson applies to the arthritis-example as well. It is clear that the doctor cannot justifiably acquire the belief *NN has arthritis in his thigh* based on his patient's testimony.¹⁸ He can, however, justifiably acquire the belief *NN has rheumatosis in his thigh*, as well he should, being NN's doctor. But again this would require him precisely to reinterpret the subject's word 'arthritis,' in violation of the reigning linguistic conventions. And so I conclude that the normative privilege of linguistic conventions does not help explain the possibility of acquiring knowledge by testimony.¹⁹

Let us turn to examining what role literal interpretation may have in helping us secure agreement in practical cooperative endeavors. This is obviously a larger question than I can fully cover here. Nonetheless, the following remarks will hopefully suffice to shake the anti-individualist's confidence that he can claim this turf as his own. Take, as an example, the task of securing political agreement on the proposition *all wars of aggression are to be banned henceforth*. It seems that our effort to reach a (genuine) agreement on this proposition would not be aided by privileging the literal interpretation of the words involved. The reason is clear: one could secure an *apparent agreement* on the proposition, yet it might turn out that the different parties to this agreement do not even agree as to the basic consequences of what they have assented to. Why is this? Simply because they might 'mean different things' by the term 'wars of aggression.' In order to secure a *genuine agreement* – the kind of agreement around which we could properly structure our collective actions – we would have

¹⁸ In fact, the present quarrel concerns whether the doctor can even justifiably acquire the belief that *NN believes he has arthritis in his thigh* from his patient's testimony, if, as Burge claims, the employment of the word 'arthritis' is supposed to indicate something vital about the subject's cognitive perspective.

¹⁹ I imagine the following objection: it is true that the doctor may justifiably acquire the belief that his patient has rheumatosis in his thigh. But this belief is not connected to the *content* of the utterance in the right kind of way for it to constitute testimonial knowledge proper. It is rather a case of inference by (pragmatic) implication (cf. Grice, "Logic and Conversation"), as in the case where I propose a particular restaurant to a friend, and my friend responds "I hear their food is terrible." I now justifiably acquire the belief *my friend does not want to go to this restaurant*. This might have been (pragmatically) implied by what he said, and I correctly apprehended the implication and drew the inference. But this does not constitute testimonial knowledge. (For considerations for and against such restrictions, see Coady, *Testimony: A Philosophical Study*, ch. 2, especially pp. 48-49.) By way of a brief response, this does not seem right: as applied to the arthritis-example or the elm-beech example, the anti-individualist would beg the question were he to simply assume that the semantic content of the speaker's utterance is what is yielded by the literal interpretation of his words.

to ensure that all parties to the agreement understand the words in the same way, as entailing the same rights and responsibilities. But this shows that we should not take a given party as having agreed to the proposition *all wars of aggression are to be banned henceforth*, if there is still disagreement on exactly what is being banned. Are humanitarian interventions to be counted as wars of aggression? What about regime change actions to overthrow a corrupt tyranny? What about wars by proxy? These are questions that we would never get around to asking if we could in the first instance rightly represent the various parties as having all assented to the same sentence on its literal interpretation.²⁰ In short, we should not count everyone who will assent to a certain proposition as *thinking the same thought* if they cannot agree to even the most basic consequences of what they thereby have assented to. Our cooperative enterprises would go nowhere if we did, and the maximization of interpersonal comparability that we would thereby obtain is achieved only by overlooking the actual differences between what the different subjects are in fact thinking.

I conclude, then, that granting a normative prerogative to linguistic conventions in the problematic kinds of cases we are considering not only renders our attributions incapable of entering into assessments of the subject's rationality and explanations of his actions, but also does nothing to illuminate the possibility, based on communicative interaction, of coordinating our practical endeavors and acquiring knowledge by each other's testimony.

It seems worthwhile also to point out that Burge's proposed application of these communal norms finds little support in mainstream normative theory. In particular, it seems to run afoul of common assumptions concerning the relation between culpability and ignorance. Consider, for

²⁰ These cases are instructively different from cases in which someone tries to exculpate himself on the basis of a semantic technicality. Say a CIA officer agrees to the proposition *the use of torture by all US officials and their affiliates is to be banned henceforth*. Upon being caught in the act, he says that he did not understand waterboarding and other 'enhanced interrogation techniques' as falling under the concept *torture*. This would be not a matter of thinking with a different concept, but of lying and deceiving.

instance, an example I briefly introduced in chapter 2: *public servants should not take bribes*. Assume that two public servants, **PS1** and **PS2**, both pledge by this proposition upon entering office. Assume that **PS1** is fully conversant with the meaning of the term ‘bribe,’ whereas **PS2** is not: he is quite certain that one should not accept money gifts from interested parties, less certain about accepting weekend trips to Paris (family invited) to learn more about the interested party’s expansion plans abroad, and quite certain that it is OK to accept invitations to a 3-star restaurant to meet the interested party’s CEO. Let us assume, however, that all of these things do in fact constitute bribes. **PS1** knows this, whereas **PS2** does not. Let us say with Burge, then, that **PS2** has an incomplete mastery of the concept *bribe*.

Now let us assume that both **PS1** and **PS2** accept the invitation to the 3-star restaurant. **PS1** reasons as follows: “I know I said I shouldn’t take bribes, but this is a minor matter, and nobody fusses about these things any more.” **PS2**, by contrast, reasons as follows: “I know I said I shouldn’t take bribes, but I’m pretty sure this wouldn’t constitute a bribe.” Should we say that when each says “I know I said I shouldn’t take bribes,” they are giving expression to the same thought?²¹ To be sure, they are both culpable for their actions, but, it would seem, differently so. **PS1** *knowingly and intentionally* broke his pledge. **PS2**, by contrast, is *culpably ignorant*, on at least two counts: as a public servant he should know the norms that attach to his office, and quite generally, agents should know the content and consequences of what they make themselves responsible for.

Now, if **PS2** is otherwise honest, he will accept his punishment and will self-correct in just the ways that Burge argues.²² In this sense, he clearly sees himself as beholden to the norms embodied in the public concept *bribe*, even though he did not fully master this concept at the time, again just as Burge says. But we can see now that we can and should separate the question of

²¹ Modulo, of course, the indexicals.

²² See chapter 3.5.4 for more details.

whether **PS2** sees himself as beholden to the norms from the question of how these norms should affect our description of his thought contents and process of reasoning at the time. For it is clear that morally (if not always legally), we hold **PS1** and **PS2** to different standards, and we find it natural to trace this differential culpability back to our assessment of what (or how) each person was thinking at the time of the transgression. In this sense, we take the state of each person's knowledge at the time as relevant to determining what they were thinking and so to determining their degree of responsibility to the norms in question. In short, **PS1** and **PS2** were not committing the same transgression, even though they performed what was (under some description) the same action. The difference traces back to what each was thinking of himself as doing at the time: **PS1** was thinking of himself as violating his pledge, whereas **PS2** was not.

This stands in sharp contrast to the anti-individualist conclusion, according to which we uphold the communal norms precisely by saying that **PS1** and **PS2** were thinking the same thought, despite their difference in knowledge. To be sure, everyone will admit that there is a sense in which **PS2** also assented to the proposition *public servants should not take bribes* on taking his pledge. But this is, one would want to say, a predominantly *legalistic* rather than *moral* sense of assent. We do not need to represent **PS2** as having tokened the unique thought content *public servants should not take bribes* in order to explain why he is nonetheless accountable to the norm expressed by his words. In fact, the particular way in which we hold him accountable – seeing him as having committed a lesser transgression than **PS1** – suggests that we do not. If we took **PS2** to have been in the same psychological state as **PS1** at the time of the transgression, we could not have been able to claim the kind of understanding of his way of thinking that expresses itself in our partial exculpation of him.

Thus, the anti-individualist's invocation of linguistic norms fails to justify a systematic preference for literal interpretation in the kinds of cases we are considering here. Shared conventional concepts are

not essential to communication. Invocation of such norms will tend to erode our ability to use our attributions to capture and convey the force of the subject's reasoning and to explain his actions. Invocation of such norms does not, in these cases, illuminate the possibility of acquiring knowledge by the testimony of others; but rather the reverse. Finally, invocation of such norms does not help us achieve, or help explain how we do achieve, our various cooperative ends for which communicative interaction is necessary.

In all of this, one is tempted to say that Burge's ascription of an inertial force to communal linguistic norms, such that they can determine attributions of thought contents to a subject even when that subject is incognizant of the norms in question, constitutes something like an attempt to cross the descriptive-normative divide in the opposite direction of what most philosophers have attempted before. As such, it is hardly a surprise that anti-individualist thought-attributions must fail to give proper descriptions of what the subjects of the thought experiments are *actually* thinking. We might concede something like the following formula: the anti-individualist thought attribution provides an accurate description of what the subject *would* have been thinking if, contrary to fact, he had none of these false beliefs about the meanings of his words. In this sense, there are norms in play: this is what one *ought* to mean by uttering these words, and one is in a real sense responsible for having uttered them. It does not follow that the subject in question *is* thinking that thought on the occasion, i.e., the same thought that someone who did possess 'full mastery' would express with those words.

Perhaps the idea of deference is supposed to supply the missing piece here. If the subject defers with regard to the meaning of his words to people more knowledgeable than he, is there not some sense in which we could say that he is thinking, albeit by deference, the same thought as the more knowledgeable person is? No. Such deference is what we do when we do not know what to think. It would be quite disingenuous (to use one of Burge's terms) of our subject to claim that he is

thinking whatever the expert is thinking on this occasion.²³ Certainly, our attribution of that thought to him would not help us get any grip on his pattern of reasoning and would not help explain his actions. Moreover, to allow such deference to take the place of genuine thinking seems inconsistent with our taking that kind of rational control over our cognitive activity which is demanded by the ideal of critical reasoning. If this is what we get from the division of linguistic labor, one should say that such division is (as Marx insisted of all division of labor) apt to produce a certain kind of alienation – an alienation of the cognitive subject from the contents of his own thoughts; an alienation which I described in ch. 4 as a failure of self-knowledge.

5.6 Can a reinterpreter account for the subject's error?

There is one final issue to consider. It might seem that my theory leaves us unable to account for our conviction that there is *error* involved in these putative cases of conceptual confusion. On this view, it is only by reference to the norms inherent in such public, conventional concepts that we can say that the subject is at fault. Thus, in considering the reinterpretation strategy, Burge says that it must involve ascribing a “true object level belief” to the subject.²⁴ One reason Burge supposes this is the fact that the reinterpreter will typically bolster his case with vague invocations of the principle of charity.

This presents a real challenge to my position. My argument in ch. 4 proceeded precisely on the assumption that we could attribute genuine error to the subject. This argument traded on the claim that we must recognize that the subject arrives at a false belief, though by a rational process of reflection. There were two aspects to this claim. On the one hand, I argued that the fact that we can

²³ Contrast this disingenuousness with the ironic tone I might take when, say, a helpful soul saves the day after I have botched a philosophical point in class, and I say “that’s *exactly* what I was thinking!”

²⁴ Cf. Burge, “Individualism and the Mental,” p. 124.

apprehend a certain kind of rational structure in his thinking forces us to conclude that he is not conceptually confused. On the other hand, the fact that he arrives at a recognizably *false* belief through this process of reasoning forces us to acknowledge that he must be thinking with a different concept than the rest of us are. But what entitles me in the first instance to say that the belief I advocate attributing is a false belief, or at any rate, a false belief about arthritis?²⁵

I reply that the reinterpreter's invocation of the principle of charity does not, in fact, dictate the attribution of a true belief to the subject, but rather the attribution of a false belief *described in such a way* as to render the error explicable and the attribution fit for figuring in assessments of the subject's rationality and in explanations of his actions. Charity does not require, in particular local cases, that we attribute true beliefs to the subject. Attributions of false beliefs are allowable so long as they find their place in an otherwise cogent and rational cognitive perspective on the world. The point of the 'reinterpretation' is precisely to describe the (false) belief in such a way as to make its location in the rational pattern of the subject's cognitive perspective as transparent as possible.

Thus, what is at stake here is my entitlement to say simply – and naïvely, perhaps, but with philosophical conviction – that the subject's belief is false. His 'way of thinking' about arthritis is consistent, but wrong. In this final section I will defend my naïve entitlement to say just this by excavating and undermining the philosophical intuitions that would seem to speak against it. I will show that the intuitions that seem to speak against this entitlement are the same that underlie what I in ch. 2 called the Concept-Cognition Thesis, according to which a subject's ability to be in any kind of cognitive touch with this particular region of reality requires that he possess the concept *arthritis*.

²⁵ There is, as we have already seen, one kind of error that the anti-individualist and I will agree is made by our subject; viz, a metalinguistic error. However, we also agree that a mere reference to a metalinguistic mistake is not sufficient to preserve our sense of the specific character of the subject's error, which is that he is in an important sense wrong about the world, not merely about words.

At the end of this section, we will be able to see, in a more general way than we were in ch. 2, what is wrong with this view.

Before embarking on that, however, it will be useful to ask whether the anti-individualist really fares any better with the task of preserving our sense that the subject is in error. As Donald Davidson pointed out,²⁶ any externalist position faces a peculiar problem of accounting for error – for representational failure –, a problem that is in many ways parallel to the internalist’s problems with accounting for representational success. The anti-individualist is of course right to assume that pointing to communal standards of correctness is a plausible first step toward saying how the subject might be guilty of error. He might even be right to say that these standards of correctness constitute something like a communal concept, in the present case, the communal concept *arthritis*. But this is not what is in question here. True, we know that our subject is in error by pointing to the communal standards of correctness, the communal concept *arthritis*. But in what sense does saying this require us also to say that the subject is *thinking with* that very concept on the occasion of committing the error? In my view (and as I suggested toward the end of ch. 4), this would amount to a serious *mislocation* of the error. For we have seen that our subject is not at all confused about his concept, if his concept has anything to do with his ‘way of thinking.’ If, on the other hand, our subject’s concept swings free of his way of thinking, then we are still left with the problem of saying precisely how *his thinking* is in error. In short, I surmise that the anti-individualist too wants to get himself into the position of being entitled to say that the subject is just plain wrong about how the world is. The anti-individualist supposes, however, that the only way to capture this error is by saying that he is wrong about his concept. I agree, of course, that it is by reference to the (presumed) wisdom embodied in the public concept that we convict him of error. But saying that he is already

²⁶ Davidson, “Externalisms.”

cognitively beholden in a certain way to that public concept – that he is thinking with it, albeit incompetently – does not contribute to our understanding of the nature and content of his error.

I will argue that the philosophical intuitions that underlie the anti-individualist's claim that the reinterpretivist is not entitled to convict the subject of an error about arthritis simply do not stand up to scrutiny, for they can be seen to trade on a certain kind of conceptual relativism, at least as regards so-called 'social kinds.' Recall that the anti-individualist's claim is that unless we ascribe the concept *arthritis* to our subject, we are not even portraying him as being in cognitive touch with the relevant region of reality, much less as thinking true or false thoughts about it. In my view, there is no other way to read this claim than as expressing the relativist conviction that the truth or falsity of a belief is entirely relative to the concepts deployed in that belief. Our subject's belief may *seem* to be pointing in the same general direction as arthritis-beliefs are, but unless it involves the concept *arthritis*, his belief is not capable of being so much as false about arthritis. It is instead true-or-false about something else (say, *tharthritis*).

This, I claim, is just conceptual relativism under a new guise. Before going on to elaborate on this allegation, however, it is important that we understand just how philosophers might end up painting themselves into this corner (as we shall see, Burge is not alone). As I understand the underlying line of thought, it typically starts with drawing a distinction between natural kinds and social kinds. Instances of natural kinds (e.g., water, gold) are, one might say, *attitude-independent* – they are what they are independently of being thought about and classified in a certain way by sentient beings such as ourselves. This makes it plausible to suppose that we possess a clear and objective epistemic standard for thinking about them: that our 'way of thinking' about them – our conceptualizations – must reflect how these things are in themselves. Thus, while there may be many ways of conceptualizing and classifying natural phenomena, we possess a clear standard of

correctness that is quite independent of any particular conceptualization, namely the very nature of the phenomena in question. Thus, we are free to say with respect to natural kinds that different systems of conceptualization are different conceptualizations of *the same thing*, and, moreover, that they are better or worse conceptualizations of the same thing, according to how well they capture the essential or defining properties of the thing in question.

The matter seems importantly different with social kinds (e.g., contract, arthritis). Pertinently, instances of these kinds do not seem to be attitude-independent. Until there was a certain kind of social practice in play, there simply were not any contracts around. The social practice supplies the norms required for objectivity and defines the subject matter in the same motion. The norms are embodied in concepts, and in order to get at the concepts we look to linguistic conventions. A different set of norms – a different set of concepts – would define a different subject matter. Thus, without being beholden to the norms in the right kind of way, there would be no way to make true or false claims about the subject matter, because one would not be able to direct one's thoughts at that region of reality at all. Thus, to return to where we started, it seems that without portraying our subject as being beholden to the concept *arthritis* in the right kind of way, I have no right to say that he is wrong about arthritis.

As natural and as plausible as this picture seems, it cannot contain the whole truth. In fact, it cannot even support the anti-individualist's claim to supply foundations for the objectivity of discourse *within* the community whose social practice is embodied in these concepts. As I said, Burge is certainly not alone in subscribing to a picture of this sort. Other writers in the post-Wittgensteinian tradition, such as Crispin Wright and Saul Kripke,²⁷ exhibit many of the same tendencies. Robert Brandom summarizes Crispin Wright's account of the normativity and objectivity of social discourse in the following way:

²⁷ Cf. Wright, *Wittgenstein on the Foundations of Mathematics*, Kripke, *Wittgenstein on Rules and Private Language*.

[Wright identifies] the normative status of being a correct application of a concept with being taken to be such a correct application, not by an individual, but by the whole community. According to this story, although individual performances can be correct or incorrect, and assessments of correctness by individuals can be correct or incorrect, no such difference applies to communal assessments. The community is incorrigible about what is a proper application of a concept and what is not. Communally endorsed applications of a rule [...] cannot be mistaken. “For the community itself there is no authority, so no standard to meet”. (Brandom, *Making it Explicit*, p. 53; the embedded quote is from Wright, *Wittgenstein on the Foundations of Mathematics*, p. 220)

Brandom perceptively goes on to argue:

There clearly are socially instituted norms of this sort. Whatever the Kwakiutl treat as an appropriate greeting gesture for their tribe, or a correctly constructed ceremonial hut, is one; it makes no sense to suppose that they could be collectively wrong about this sort of thing. The question is whether *conceptual* norms ought to be understood as being of this type. There is good reason to think they ought not. It is a fundamental feature of our understanding of our concepts that they incorporate *objective* commitments. (Brandom, *Making it Explicit*, p. 53)

Wright, on Brandom’s retelling, draws the proper consequence of this form of social constructivism and concedes that there is no objectivity to be had here: “the objectivity we take our conceptual norms to have is an illusion that must be surrendered if they are to be properly understood” (Brandom, *Making it Explicit*, pp. 53-54). Burge draws no such conclusion, but, as I shall argue shortly, it is the proper consequence of his views.²⁸

²⁸ Another very interesting take on these tendencies in contemporary philosophy can be found in John Haugeland’s claim that they express a form of *Social Cartesianism*. True, Kripke, Burge, Wright, et al., are vehemently anti-Cartesian in one important respect, in that they deny “the possibility of legitimate, determinate meanings for individual people considered in isolation” (Haugeland, “Social Cartesianism,” p. 224). But for all that, they retain an allegiance to a deeper and more insidious Cartesian heritage, namely the underlying assumption that “if our ideas exist at all, they are fully determinate in and of themselves, quite apart from anything else that there may or may not be” (ibid., p. 223). In this, the new picture “neglects an essential element in the determinacy-cum-legitimacy of these admittedly public meanings.

Brandom's diagnosis does well to tap into the metaphysical and epistemological roots of the anti-individualist's troubles. But more might be learned from seeing the distinctively practical problems that such a view of social discourse leads to. I will outline two ways to approach the nature of the philosophical problem in question. First, the kind of argument I marshaled in ch. 2 can be generalized to show that whatever success the anti-individualist can claim in laying foundations for the objectivity of intra-communal discourse comes at the cost of losing *any* foundation whatsoever for objective and rational comparison in an inter-communal setting. The problem here is not merely the familiar one that people thinking with different concepts may not possess demonstrative criteria for deciding whose 'way of thinking' is right and whose wrong (if indeed, they are not both wrong). Rather, what is lost is the much more fundamental possibility of saying that they are so much as thinking and talking about the same things at all, if the subject matter of discourse is defined by the concepts in question, if making reference to arthritis *depends* on one's possessing and deploying that concept.

Yet more devastating, however, is the fact that the same problem arises for the anti-individualist in accounting for rational comparison across temporal stages of our own communal linguistic practice. Take the example of 'contract,' a paradigm social kind term in Burge's sense. Many philosophers are given to the idea that the extension of 'contract' just is whatever the legal experts take it to be: it is thought to be inconceivable that we as a collective body could fail to properly grasp what a contract is, since the very existence of contracts somehow depends on our agreement.²⁹ Clearly, there is something to this view. But as it stands, it has serious problems.

That further essential element is the world itself. The irony is that, inasmuch as they systematically deny the world a role they do not so much demolish the Cartesian barrier as merely shift it 'outward' a notch. That expanded inner/outer perimeter now encloses not only private thoughts and feelings, but also public language and social norms; but it still *excludes* the real things and happenings that those thoughts and conversations are primarily about" (ibid., p. 224).

²⁹ Cf. Brandom's recapitulation of Wright's view above.

Pertinently, such a view cannot account for continuity through conceptual change. Thus, if this is the anti-individualist's view, it should be clear that he cannot account for the possibility of conceptual improvement over time.

It is an indisputable fact that the extensions of these various social kind terms change over time,³⁰ and perhaps even that they undergo continuous change. On the view we are currently considering, recognizing this fact would force us to postulate a new concept for every significantly different stage in this development. There might even, it seems, be large stretches of time during which it was unclear *which concept* was in play. In such cases, it would be pure stipulation to say that there is a single unique concept in play throughout this development, and that all the disagreement we observe about the precise extension of 'contract' is merely a disagreement between those who have a worse and those who have a better grasp of what that concept is (or perhaps, between people who have different, but equally bad grasps of this concept).

To make sense of conceptual change here, Burge would have to allow that people who are thinking with different concepts can nonetheless make rational contact with each others' points of view and see themselves as talking about the same things. But this concession is just what I needed in order to maintain my claim at the start of this section, namely that I am entitled to say that our subject is simply wrong about arthritis, without thereby attributing to him the concept *arthritis*. Thus, Burgean anti-individualism must fail, because it must either hold that the subject matter is indeed fully defined by the concept, and then be forced to give up the hope of making rational sense of the emergence of our current concepts; or he must join us in saying that if there are norms in play at all, they must somehow lie in the subject matter itself, a subject matter which can be jointly reached by multiple subjects who do not share a particular conceptualization of it.

³⁰ Of course, this is not unique to social kind terms. *Pace* Putnam on natural kind terms, a clear-headed view of the situation requires our acknowledging that, say, the scientific term 'planet' has changed its extension (in fact, as recently as August 16, 2006).

I conclude, then, that the anti-individualist has presented no argument to undermine my entitlement to say that our subject thinks falsely about arthritis. The sketch of an argument that the anti-individualist does present backfires, and ultimately undermines his claim to have secured the normative foundations for even an arbitrarily delimited communal context such as *our legal practice right now*. If there is no way of even making reference to the subject matter of that discourse without deploying the concepts that define it, we would have no way to rationally persuade anyone to adopt those concepts and take on the wisdom we claim is embodied in them. What is left is merely a sham form of normativity, a pure conventionalism.

In other places, Burge displays more sensitivity to these problems.³¹ Thus, for instance, in “Intellectual Norms and the Foundations of Mind” he is concerned to account for the fact that we can *improve our concepts* in important ways (and not merely improve *our grasp* of these concepts). Accounting for such improvement requires that we be able to define our concepts independently of the linguistic norms in play at any given time. As we saw in ch. 2, Addendum, Burge’s strategy in this article is to tie a speaker’s concept to the *cognitive value* of a term in that speaker’s mental repertoire, rather than, as in “Individualism and the Mental,” to the *conventional meaning* of that term in the public language. A motivation for this major theoretical overhaul is the realization that tying concepts to conventional meanings leaves us, as I argued above, unable to account for continuity through change in our thinking – how we can see ourselves as talking about the same subject matter, albeit in a different way.³²

Needless to say, I endorse this move. There remains a crucial worry, however, as Burge has yet to draw the full consequence of this liberation of our thinking from the shackles of convention.

³¹ Although, it should be said, he never acknowledges them as problems resulting from the view defended in “Individualism and the Mental.”

³² Burge, “Intellectual Norms and the Foundations of Mind,” p. 271.

For while Burge now acknowledges that our cognitive possibilities *as thinkers* swing free of the linguistic conventions to which they were tied in “Individualism and the Mental,”³³ he still defends a view according to which our cognitive possibilities *as understanders* are still bound by these conventions.³⁴ Throughout these pages, I have argued strenuously against the idea that what we are capable of thinking and thinking about is in any important sense constrained by the conventions governing our language community. But that idea, however flawed, at least provided an argument as to why our hermeneutic efforts should grant a systematic privilege to interpretation in accordance with those conventions. But even now that this underlying idea is brought into question, Burge still holds on to the supposition that what we can *present each other as thinking and thinking about* is constrained by these linguistic conventions. This is an unattractive view in its own right: even when I understand you optimally well, I may yet not understand you. The view is merely self-defeating, however, when it is offered as the conclusion of a sustained inquiry into the social dimension of mindedness, into the conditions of possibility for communication and shared knowledge. Such a view could certainly no longer be defended by reference to a supposed normative prerogative that such conventions enjoy.

³³ Cf. Burge, “Intellectual Norms and the Foundations of Mind,” p. 272: “[O]ur cognitive commitments and potential go beyond the boundaries set by conventional (or idiolectic) linguistic meaning.”

³⁴ See Burge, “Postscript to ‘Belief *de re*’,” p. 80, and further analysis in ch. 2, Addendum.

REFERENCES

- Anscombe, G.E.M. *Intention*. Ithaca, NY: Cornell University Press, 1957.
- Antony, Louise M. "Rabbit-Pots and Supernovas: On the Relevance of Psychological Data to Linguistic Theory." In Alex Barber, ed., *The Epistemology of Language*, Oxford University Press, 2003, pp. 48-68.
- Bach, Kent. "Burge's New Thought-Experiment: Back to the Drawing Room." *Journal of Philosophy*, Vol. 85, No. 2 (Feb. 1988), pp. 88-97.
- , *Thought and Reference*. Oxford: Clarendon Press, 1987.
- Balsvik, Eivind. "Davidsons forklaring på førstepersons autoritet." *Norsk filosofiske tidsskrift*, Vol. 2, No. 1-2, pp. 121-137.
- Beaney, Michael. *Frege: Making Sense*. London: Duckworth, 1996.
- Bernecker, Sven. "Externalism and the Attitudinal Component of Self-Knowledge." *Nous*, Vol. 30, No. 2 (June 1996), pp. 262-275.
- , "Self-Knowledge and Closure." In P. Ludlow & N. Martin, eds., *Externalism and Self-Knowledge*, Stanford, CA: CSLI Publications, 1998, pp. 333-349.
- Bilgrami, Akeel. "A Trilemma for Redeployment." *Philosophical Issues*, Vol. 13, No. 1, 2003, pp. 22-30.
- , *Belief & Meaning: The Unity and Locality of Mental Content*. Oxford: Blackwell, 1992.
- , "Can Externalism be reconciled with Self-Knowledge?" Reprinted in A. Pessin and S. Goldberg, eds., *The Twin Earth Chronicles: Twenty Years of Reflection on Hilary Putnam's 'The Meaning of Meaning'*, Armonk, NY: M.E. Sharpe, 1996, pp. 362-393.
- Boghossian, Paul. "Content and Self-Knowledge." Reprinted in Peter Ludlow and Norah Martin, eds., *Externalism and Self-Knowledge*, Stanford, CA: CSLI Publications, 1998, pp. 149-174.
- , "What the Externalist can know a priori." In C. Wright, B.C. Smith, and C. Macdonald, eds., *Knowing Our own Minds*, Oxford University Press, 1998, pp. 271-284.
- Brandom, Robert B. "Inferentialism and some of its Challenges." *Philosophy and Phenomenological Research*, Vol. 74, No. 3 (May 2007), pp. 651-676.

- . *Making it Explicit: Reasoning, Representing, and Discursive Commitment*. Cambridge, MA: Harvard University Press, 1994.
- Brown, Jessica. *Anti-individualism and Knowledge*. Cambridge, MA: The MIT Press, 2004.
- . "Critical Reasoning, Understanding, and Self-Knowledge." *Philosophy and Phenomenological Research*, Vol. 61, No. 3 (Nov. 2000), pp. 659-676.
- . "Externalism and the Fregean Tradition." In Alex Barber, ed., *Epistemology of Language*, Oxford University Press, 2003, pp. 431-455.
- . "The Incompatibility of Anti-Individualism and Privileged Access." *Analysis*, Vol. 55, No. 3 (Jul. 1995), pp. 149-156.
- Burge, Tyler. "A Century of Deflation and a Moment about Self-Knowledge." *Proceedings and Addresses of the American Philosophical Association*, Vol. 73, No. 2 (Nov. 1999), pp. 25-46.
- . "Belief *de re*." Reprinted in *Foundations of Mind*, Oxford: Clarendon Press, 2007, pp. 44-64.
- . "Cartesian Error and the Objectivity of Perception." Reprinted in *Foundations of Mind*, Oxford: Clarendon Press, 2007, pp. 192-207.
- . "Comprehension and Interpretation." In L.E. Hahn, ed., *The Philosophy of Donald Davidson*, LaSalle, IL: Open Court Publishing, 1999, pp. 229-250.
- . "Concepts, Definitions, and Meaning." Reprinted in *Foundations of Mind*, Oxford: Clarendon Press, 2007, pp. 291-306.
- . "Content Preservation." *The Philosophical Review*, Vol. 102, No. 4 (Oct., 1993), pp. 457-488.
- . "Descartes on Anti-individualism." Reprinted in *Foundations of Mind*, Oxford: Clarendon Press, 2007, pp. 420-439.
- . *Foundations of Mind: Philosophical Essays, Vol. 2*. Oxford: Clarendon Press, 2007.
- . "Frege on Extensions of Concepts, from 1884 to 1903." Reprinted in *Truth, Thought, and Reason: Essays on Frege*, Oxford: Clarendon Press, 2007, pp. 273-298.
- . "Frege on Knowing the Third Realm." Reprinted in *Truth, Thought, and Reason: Essays on Frege*, Oxford: Clarendon Press, 2007, pp. 299-316.

- "Frege on Sense and Linguistic Meaning." Reprinted in *Truth, Thought, and Reason: Essays on Frege*, Oxford: Clarendon Press, 2007, pp. 242-269.
- "Individualism and Psychology." Reprinted in *Foundations of Mind*, Oxford: Clarendon Press, 2007, pp. 221-253.
- "Individualism and the Mental." Reprinted in *Foundations of Mind*, Oxford: Clarendon Press, 2007, pp. 100-150.
- "Intellectual Norms and the Foundations of Mind." Reprinted in *Foundations of Mind*, Oxford: Clarendon Press, 2007, pp. 254-274.
- "Memory and Self-Knowledge." In Peter Ludlow and Norah Martin, eds., *Externalism and Self-Knowledge*, Stanford, CA: CSLI Publications, 1998, pp. 351-370.
- "Mental Agency in Authoritative Self-Knowledge: Reply to Kobes." In Martin Hahn and Bjørn Ramberg, eds., *Reflections and Replies: Essays on the Philosophy of Tyler Burge*, Cambridge, MA: The MIT Press, 2003, pp. 417-433.
- "Other Bodies." Reprinted in *Foundations of Mind*, Oxford: Clarendon Press, 2007, pp. 82-99.
- "Our Entitlement to Self-Knowledge." Reprinted in Peter Ludlow and Norah Martin, eds., *Externalism and Self-Knowledge*, Stanford, CA: CSLI Publications, 1998, pp. 239-263.
- "Perceptual Entitlement." *Philosophy and Phenomenological Research*, Vol. 67, No. 3 (Nov. 2003), pp. 503-548.
- "Phenomenality and Reference: Reply to Loar." In Martin Hahn and Bjørn Ramberg, eds., *Reflections and Replies: Essays on the Philosophy of Tyler Burge*, Cambridge, MA: The MIT Press, 2003, pp. 435-450.
- "Philosophy of Language and Mind, 1950-2000." Reprinted in *Foundations of Mind*, Oxford: Clarendon Press, 2007, pp. 440-464.
- "Postscript to 'Individualism and the Mental'." In *Foundations of Mind*, Oxford: Clarendon Press, 2007, pp. 151-181.
- "Postscript to 'Belief *de re*'." In *Foundations of Mind*, Oxford: Clarendon Press, 2007, pp. 65-81.
- "Reply to Christopher Gauker." In M. J. Frappolli and E. Romero, eds., *Meaning, Basic Self-Knowledge, and Mind: Essays on Tyler Burge*, Stanford, CA: CSLI Publications, 2003, pp. 244-250.

- . "Reply to Martin Davies." In M. J. Frappolli and E. Romero, eds., *Meaning, Basic Self-Knowledge, and Mind: Essays on Tyler Burge*, Stanford, CA: CSLI Publications, 2003, pp. 261-268.
- . *Self and Self Understanding*. John Dewey lectures, delivered at Columbia University, December 2007.
- . "Sinning against Frege." Reprinted in *Truth, Thought, and Reason: Essays on Frege*, Oxford: Clarendon Press, 2007, pp. 213-239.
- . "Some Reflections on Skepticism: Reply to Stroud." In Martin Hahn and Bjørn Ramberg, eds., *Reflections and Replies: Essays on the Philosophy of Tyler Burge*, Cambridge, MA: The MIT Press, 2003, pp. 335-346.
- . *Truth, Thought, and Reason: Essays on Frege*. Oxford: Clarendon Press.
- . "Two Thought Experiments Reviewed." Reprinted in *Foundations of Mind*, Oxford: Clarendon Press, 2007, pp.182-191.
- . "Wherein is Language Social?" Reprinted in *Foundations of Mind*, Oxford: Clarendon Press, 2007, pp. 275-290.
- Cappelen, Herman, and Ernest Lepore. *Insensitive Semantics: In Defense of Semantic Minimalism and Speech Act Pluralism*. Oxford: Blackwell, 2004.
- Carston, Robyn. *Thoughts and Utterances: The Pragmatics of Explicit Communication*. Oxford: Blackwell, 2002.
- Chomsky, Noam. *Syntactic Structures*. The Hague: Mouton & Co., 1957.
- . *New Horizons in the Study of Language and Mind*. Cambridge University Press, 2000.
- Coady, C.A.J. *Testimony: A Philosophical Study*. Oxford: Clarendon Press, 1992.
- Davidson, Donald. "A Coherence Theory of Truth and Knowledge." Reprinted in *Subjective, Intersubjective, Objective*, Oxford: Clarendon Press, 2001, pp. 137-157.
- . "A Nice Derangement of Epitaphs." Reprinted in *Truth, Language, and History*, Oxford: Clarendon Press, 2005, pp. 89-107.
- . "Actions, Reasons, and Causes." Reprinted in *Essays on Actions and Events*, Oxford: Clarendon Press, 1980, pp. 3-19.

- . "Dialectic and Dialogue." Reprinted in *Truth, Language, and History*, Oxford: Clarendon Press, 2005, pp. 251-259.
- . "Epistemology Externalized." Reprinted in *Subjective, Intersubjective, Objective*, Oxford: Clarendon Press, 2001, pp. 192-204.
- . "Externalisms." In P. Kotatko, P. Pagin, and G. Segal, eds., *Interpreting Davidson*, Stanford, CA: CSLI Publications, 2001, pp. 1-16.
- . "First Person Authority." Reprinted in *Subjective, Intersubjective, Objective*, Oxford: Clarendon Press, 2001, pp. 3-14.
- . *Inquiries into Truth and Interpretation*. Oxford: Clarendon Press, 1984.
- . "Knowing One's Own Mind." Reprinted in *Subjective, Intersubjective, Objective*, Oxford: Clarendon Press, 2001, pp. 15-38.
- . "Mental Events." Reprinted in *Essays on Actions and Events*, Oxford: Clarendon Press, 1980, pp. 207-225.
- . "On the Very Idea of a Conceptual Scheme." Reprinted in *Inquiries into Truth and Interpretation*, Oxford: Clarendon Press, 1984, pp. 183-198.
- . "Psychology as Philosophy." Reprinted in *Essays on Actions and Events*, Oxford: Clarendon Press, 1980, pp. 229-239.
- . "Radical Interpretation." Reprinted in *Inquiries into Truth and Interpretation*, Oxford: Clarendon Press, 1984, pp. 125-139.
- . "Rational Animals." Reprinted in *Subjective, Intersubjective, Objective*, Oxford: Clarendon Press, 2001, pp. 95-105.
- . "Replies to Barry Stroud, John McDowell, and Tyler Burge." *Philosophy and Phenomenological Research*, Vol. 67, No. 3 (Nov. 2003), pp. 691-699
- . "The Inscrutability of Reference." Reprinted in *Inquiries into Truth and Interpretation*, Oxford: Clarendon Press, 1984, pp. 227-241.
- . "The Myth of the Subjective." Reprinted in *Subjective, Intersubjective, Objective*, Oxford: Clarendon Press, 2001, pp. 39-52.

- . “The Social Aspect of Language.” Reprinted in *Truth, Language, and History*, Oxford: Clarendon Press, 2005, pp. 109-125.
- . “Three Varieties of Knowledge.” Reprinted in *Subjective, Intersubjective, Objective*, Oxford: Clarendon Press, 2001, pp. 205-220.
- . “What is Present to the Mind?” Reprinted in *Subjective, Intersubjective, Objective*, Oxford: Clarendon Press, 2001, pp. 53-67.
- Davies, Martin. “Externalism, Architecturalism, and Epistemic Warrant.” In C. Wright, B.C. Smith, and C. Macdonald, eds., *Knowing Our Own Minds*, Oxford University Press, 1998, pp. 321-61.
- . “Externalism, Self-Knowledge, and Transmission of Warrant.” In M.J. Frapolli and E. Romero, eds., *Meaning, Basic Self-Knowledge, and Mind: Essays on Tyler Burge*, Stanford, CA: CSLI Publications, 2003, pp. 99-124.
- . “The Problem of Armchair Knowledge.” In S. Nuccetelli, ed., *New Essays on Semantic Externalism and Self-Knowledge*, Cambridge, MA: The MIT Press, 2003, pp. 23-55.
- Dennett, Daniel C. “Beyond Belief.” Reprinted in *The Intentional Stance*, Cambridge, MA: The MIT Press, 1987, pp. 117-202.
- . “Conditions of Personhood.” Reprinted in *Brainstorms: Philosophical Essays on Mind and Psychology*, Cambridge, MA: The MIT Press, 1978, pp. 267-285.
- . “Intentional Systems.” Reprinted in *Brainstorms: Philosophical Essays on Mind and Psychology*, Cambridge, MA: The MIT Press, 1978, pp. 1-22.
- . “True Believers: The Intentional Strategy and why it works.” Reprinted in *The Intentional Stance*, Cambridge, MA: The MIT Press, 1987, pp. 13-35.
- Devitt, Michael. “Linguistics is not Psychology.” In Alex Barber, ed., *The Epistemology of Language*, Oxford University Press, 2003, pp. 106-139.
- Diamond, Jared. *The Third Chimpanzee: The Evolution and Future of the Human Animal*. New York: Harper Perennial, 1992.
- Donnellan, Keith. “Burge Thought Experiments.” In Martin Hahn and Bjørn Ramberg, eds., *Reflections and Replies: Essays on the Philosophy of Tyler Burge*, Cambridge, MA: The MIT Press, 2003, pp. 59-75.

----- . “Reference and Definite Descriptions.” *The Philosophical Review*, Vol. 75, No. 3 (Jul., 1966), pp. 281-304.

Dretske, Fred. *Naturalizing the Mind*. Cambridge, MA: The MIT Press, 1995.

Dummett, Michael. *Frege: Philosophy of Language*. 2nd edition. Cambridge, MA: Harvard University Press, 1981. [1st edition, 1973.

----- . *Origins of Analytical Philosophy*. London: Duckworth, 1993.

Ebbs, Gary. *Rule Following and Realism*. Cambridge, MA: Harvard University Press, 1997.

Falvey, Kevin and Joseph Owens. “Externalism, Self-Knowledge and Skepticism.” *The Philosophical Review*, Vol. 103, No. 1 (Jan., 1994), pp. 107-137.

Fodor, Jerry. *Concepts: Where Cognitive Science Went Wrong*. Oxford University Press, 1998.

----- . “Methodological Solipsism considered as a Research Strategy in Cognitive Psychology.” *Behavioral and Brain Sciences*, Vol. 3, No. 1, 1980, pp. 63-72..

----- . *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*. Cambridge, MA: The MIT Press, 1987.

Fodor, Jerry and Ernest Lepore. “Brandom Beleaguered.” *Philosophy and Phenomenological Research*, Vol. 74, No. 3 (May 2007), pp. 677–691.

----- . *Holism: A Shopper’s Guide*. Cambridge, MA: Blackwell, 1992.

Frege, Gottlob. “On Concept and Object.” Translated by Peter Geach. Reprinted in Michael Beaney, ed., *The Frege Reader*, Oxford: Blackwell, 1997, pp. 181-193. [Original German publication, 1892.]

----- . *Foundations of Arithmetic*. Translated by J.L. Austin. Oxford: Blackwell, 1950. [German original, 1884.]

----- . “Letter to Husserl 24.5.1891.” Translated by Hans Kaal. Reprinted in Michael Beaney, ed., *The Frege Reader*, Oxford: Blackwell, 1997, pp. 149-50.

----- . “Letters to Husserl 1906.” Translated by Hans Kaal. Reprinted in Michael Beaney, ed., *The Frege Reader*, Oxford: Blackwell, 1997, pp. 301-307.

----- . “On *Sinn* and *Bedeutung*.” Translated by Max Black. Reprinted in Michael Beaney, ed., *The Frege Reader*, Oxford: Blackwell, 1997, pp. 151-171. [Original German publication, 1892.]

- . “The Thought.” Translated by Peter Geach and R.T. Stoothoff. Reprinted in Michael Beaney, ed., *The Frege Reader*, Oxford: Blackwell, 1997, pp. 325-345. [Original German publication, 1918.]
- Gadamer, Hans-Georg. *Truth and Method*. 2nd edition. Translation revised by Joel Weinsheimer and Donald G. Marshall. New York: Continuum Publishing, 1997. [Original German publication, 1960.]
- . “Wie weit schreibt Sprache des Denken vor?” Reprinted in *Gesammelte Werke 2, Hermeneutik II*, Tübingen: J.C.B. Mohr (Paul Siebeck), 1993. [Original publication, 1970.]
- Gallois, Andre. *The World Without, The Mind Within: An Essay on First-Person Authority*. Cambridge University Press, 2004.
- Gardner, Howard. *The Mind's New Science: A History of the Cognitive Revolution*. New York: Basic Books, 1985.
- Gauker, Christopher. “Social Externalism and Linguistic Communication.” In M. J. Frappolli and E. Romero, eds., *Meaning, Basic Self-Knowledge, and Mind: Essays on Tyler Burge*, Stanford, CA: CSLI Publications, 2003, pp. 1-33.
- Geertz, Clifford. “Anti Anti-Relativism.” *American Anthropologist*, New Series, Vol. 86, No. 2 (Jun., 1984), pp. 263-278.
- Gibbons, John. “Externalism and Knowledge of the Attitudes.” *Philosophical Quarterly*, Vol. 51, No. 202 (Jan. 2001), pp. 13-28.
- Goldberg, Sanford. “What do you know when you know your own thoughts?” In Susana Nuccetelli, *New Essays on Semantic Externalism and Self-Knowledge*, Cambridge, MA: The MIT Press, 2003, pp. 241-256
- Goldman, Alvin I. *Simulating Minds: The Philosophy, Psychology, and Neuroscience of Mindreading*. Oxford University Press, 2006.
- Gopnik, Allison. “How We Know our Minds: The Illusion of First-Person Knowledge of Intentionality.” *Behavioral and Brain Sciences* 16 (1993), pp. 1-14.
- Grice, Paul. “Logic and Conversation.” Reprinted in *Studies in the Way of Words*, Cambridge, MA: Harvard University Press, pp. 3-143.
- Griffiths, Paul. “Molecular and Developmental Biology.” In P. Machamer and M. Silberstein, *The Blackwell Guide to the Philosophy of Science*, Oxford: Blackwell, 2002, pp. 252-271.

- Hahn, Martin. "When Swampmen get Arthritis: 'Externalism' in Burge and Davidson." In Martin Hahn and Bjørn Ramberg, eds., *Reflections and Replies: Essays on the Philosophy of Tyler Burge*, Cambridge, MA: The MIT Press, 2003, pp. 29-58.
- Haugeland, John. "Social Cartesianism." In R. Schantz, ed., *The Externalist Challenge*, Berlin: Walter de Gruyter, 2004, pp. 213-225.
- Hobbes, Thomas. *Leviathan, or The Matter, Forme and Power of a Common Wealth Ecclesiasticall and Civil*. Edwin Curley, ed., Indianapolis, IN: Hackett Publishing Company, 1994. [Original edition, 1651.]
- Hornsby, Jennifer. "Semantic Innocence and Psychological Understanding." In *Simple Mindedness: In Defense of Naïve Naturalism in the Philosophy of Mind*, Cambridge, MA: Harvard University Press, 1997, pp. 195-220.
- James, William. *Principles of Psychology*. 2 vols. Reprint edition, New York: Dover Publications, 1950. [Original publication, 1890.]
- Kant, Immanuel. *Critique of Pure Reason*. Translation by Norman Kemp Smith, Boston & New York: Bedford/St. Martin's, 1965. [Original German publication, 1781, 2nd edition, 1787.]
- Kobes, Bernard. "Mental Content and Hot Self-Knowledge." In Martin Hahn and Bjørn Ramberg, eds., *Reflections and Replies: Essays on the Philosophy of Tyler Burge*, Cambridge, MA: The MIT Press, 2003, pp. 201-227.
- Kripke, Saul. *Naming and Necessity*. Cambridge, MA: Harvard University Press, 1980.
- . *Wittgenstein on Rules and Private Language: an Elementary Exposition*. Cambridge, MA: Harvard University Press, 1982.
- Kuhn, Thomas S. *The Structure of Scientific Revolutions*. 3rd edition. University of Chicago Press, 1996. [1st edition, 1962.]
- Laurence, Stephen. "Is Linguistics a Branch of Psychology?" In Alex Barber, ed., *The Epistemology of Language*, Oxford University Press, 2003, pp. 69-104.
- and Eric Margolis. "Concepts and Cognitive Science." In E. Margolis and S. Laurence, *Concepts: Core Readings*, Cambridge, MA: The MIT Press, 1999, pp. 3-81.
- Leibniz, G.W.F. *Discourse on Metaphysics*. Translation by George Montgomery. In *Discourse on Metaphysics, Correspondence with Arnauld, and Monadology*, LaSalle, IL: Open Court Publishing, 1902.

- . *Monadology*. Translation by Nicholas Rescher. In N. Rescher, ed., *G.W. Leibniz's Monadology: An Edition for Students*, Pittsburgh, PA: University of Pittsburgh Press, 1991.
- Loar, Brian. "Social Content and Psychological Content." In Robert H. Grimm and Donald D. Merrill, eds., *Contents of Thought*, Tucson, AZ: The University of Arizona Press, 1988, pp. 99-110.
- Ludlow, Peter and Norah Martin, eds., *Externalism and Self-Knowledge*. Stanford, CA: CSLI Publications, 1998.
- Ludwig, Kirk. "The Myth of Social Content." Retrieved from <http://web.clas.ufl.edu/users/kludwig/web/users/kludwig/Papers/The%20Myth%20of%20Social%20Content.htm>
- McDowell, John. *Mind and World*. 2nd edition. Cambridge, MA: Harvard University Press, 1994.
- . "Putnam on Mind and Meaning." Reprinted in *Meaning, Knowledge, and Reality*, Cambridge, MA: Harvard University Press, 1998, pp. 275-291.
- . "Subjective, Intersubjective, Objective." *Philosophy and Phenomenological Research*, Vol. 67, No. 3 (Nov. 2003), pp. 675–681.
- McGeer, Victoria, "Is 'Self-Knowledge' an Empirical Problem? Renegotiating the Space of Philosophical Explanation." *Journal of Philosophy*, Vol. 93, No. 10 (Oct., 1996), pp. 483-515.
- McGinn, Colin. "The Structure of Content." In A. Woodfield, ed., *Thought and Object: Essays on Intentionality*, Oxford: Clarendon Press, 1982, pp. 207-258.
- . "Radical Interpretation and Epistemology." In Ernie Lepore, ed., *Truth and Interpretation: Essays on the Philosophy of Donald Davidson*, Oxford: Blackwell, 1986, pp. 356-368.
- McKinsey, Michael. "Anti-individualism and Privileged Access." *Analysis*, Vol. 51, No. 1 (Jan., 1991), pp. 9-16.
- McLaughlin, Brian. "McKinsey's Challenge, Warrant Transmission, and Skepticism." In S. Nuccetelli, ed., *New Essays on Semantic Externalism and Self-Knowledge*, Cambridge, MA: The MIT Press, 2003, pp. 79-96.
- Moran, Richard. *Authority and Estrangement: An Essay on Self-Knowledge*. Cambridge, MA: Harvard University Press, 2001.

- , "Interpretation Theory and the First Person." *The Philosophical Quarterly*, Vol. 44, No. 175 (Apr., 1994), pp. 154-173.
- Nichols, Shaun and Stephen Stich. *Mindreading: An Integrated Account of Pretence, Self-Awareness, and Understanding of Other Minds*, Oxford University Press, 2003.
- Nisbett, Richard and Timothy D. Wilson. "Telling More than we can Know." *Psychological Review* Vol.84, No.3 (May 1977), pp. 231-259.
- Nuccetelli, Susana. "Knowing that One knows What one is Thinking about." In S. Nuccetelli, ed., *New Essays on Semantic Externalism and Self-Knowledge*, Cambridge, MA: The MIT Press, 2003, pp. 169-184.
- , ed. *New Essays on Semantic Externalism and Self-Knowledge*. Cambridge, MA: The MIT Press, 2003.
- Peacocke, Christopher. *A Study of Concepts*. Cambridge, MA: The MIT Press, 1992.
- Pessin, Andrew and Sanford Goldberg, eds. *The Twin Earth Chronicles: Twenty Years of Reflection on Hilary Putnam's "The Meaning of 'Meaning'."* Armonk, NY: M.E. Sharpe, 1996.
- Preyer, Gerhard, and Georg Peter, eds. *Context-Sensitivity and Semantic Minimalism: New Essays on Semantics and Pragmatics*. Oxford University Press, 2007.
- Prinz, Jesse. *Furnishing the Mind: Concepts and their Perceptual Basis*. Cambridge, MA: The MIT Press, 2002.
- Putnam, Hilary. "The Meaning of 'Meaning'." Reprinted in *Mind, Language, and Reality: Philosophical Papers, Vol. 2*, Cambridge University Press, 1975, pp. 215-271.
- Quine, W.v.O. "Ontological Relativity." In *Ontological Relativity & Other Essays*, New York: Columbia University Press, 1969, pp. 26-68.
- , "Two Dogmas of Empiricism." *The Philosophical Review*, Vol. 60, No. 1 (Jan., 1951), pp. 20-43.
- , *Word & Object*. Cambridge, MA: The MIT Press, 1960.
- Ramberg, Bjørn T. *Donald Davidson's Philosophy of Language: An Introduction*. Oxford: Blackwell, 1989.
- , "What Davidson said to the Skeptic or: Anti-Representationalism, Triangulation, and the Naturalization of the Subjective." In P. Kotatko, P. Pagin, and G. Segal, eds., *Interpreting Davidson*, Stanford, CA: CSLI Publications, 2001, pp. 213-236.
- Recanati, François. *Literal Meaning*. Cambridge University Press, 2004.

- Rousseau, Jean-Jacques. *The Social Contract*. In Viktor Gourevitch, ed., *The Social Contract, and Other Later Political Writings*, Cambridge University Press, 1997. [Original French edition, 1762.]
- Russell, Bertrand. *Mysticism and Logic*. London: Unwin Hyman, 1959.
- Ryle, Gilbert. *The Concept of Mind*. Reprint edition. Chicago, IL: University of Chicago Press, 2002. [Original publication, 1949.]
- Salmon, Nathan. "Two Conceptions of Semantics." In Z.G. Szabo, ed., *Semantics versus Pragmatics*, Oxford: Clarendon Press, 2005, pp. 317-328.
- Sawyer, Sarah. "Conceptual Errors and Social Externalism." *Philosophical Quarterly*, Vol. 53, No. 211 (2003), pp. 265-273.
- "In Defence of Burge's Thesis." *Philosophical Studies*, Vol. 107, No. 2 (2002), pp. 109-128.
- Schleiermacher, Friedrich. *Hermeneutics and Criticism, and Other Writings*. Edited by Andrew Bowie. Cambridge University Press, 1998.
- Segal, Gabriel. *A Slim Book about Narrow Content*. Cambridge, MA: The MIT Press, 2000.
- Sellars, Wilfrid. "Empiricism and the Philosophy of Mind." Reprinted in *Science, Perception, and Reality*. Atascadero, CA: Ridgeview Publishing Company, 1963, pp. 127-196.
- Shoemaker, Sydney. "Self-Knowledge and 'Inner Sense.'" *Philosophy and Phenomenological Research*, Vol. 54, No. 2 (Jun., 1994), pp. 249-314.
- *The First-Person Perspective, and Other Essays*. Cambridge University Press, 1996.
- Sperber, Dan and Deirdre Wilson. *Relevance: Communication and Cognition*. Oxford: Blackwell, 1986.
- "Relevance Theory." In L. Horn and G. Ward, eds., *The Handbook of Pragmatics*. Oxford: Blackwell, 2004, pp. 607-632.
- Stalnaker, Robert. "Narrow Content." Reprinted in *Context and Content: Essays on Intentionality in Speech and Thought*, Oxford University Press, 1999, pp. 194-209.
- Sterelny, Kim and Paul E. Griffiths. *Sex and Death: An Introduction to Philosophy of Biology*. Chicago, IL: University of Chicago Press, 1999.
- Stich, Stephen. "Autonomous Psychology and the Belief-Desire Thesis." *The Monist*, Vol. 61, No. 4 (1978), pp. 573-591.

- . *From Folk Psychology to Cognitive Science: The Case against Belief*. Cambridge, MA: The MIT Press, 1983.
- Warfield, Ted A. "Tyler Burge's Self-Knowledge." *Grazer Philosophische Studien* 70 (2005), pp. 169-178.
- Wikforss, Åsa. "Content Externalism and the Fregean Tradition." In Tomas Marvan, ed., *What Determines Content? The Internalism/Externalism Dispute*, Cambridge Scholars Press, 2006, pp. 163-179.
- . "Social Externalism and Conceptual Errors." *The Philosophical Quarterly*, Vol. 51, No. 203 (April, 2001), pp. 217-237.
- . "Self-Knowledge and Knowledge of Content." Forthcoming in *Canadian Journal of Philosophy*.
- Williams, Michael. "Bilgrami on Belief and Meaning: Is Fregean Externalism Possible?" *Philosophy and Phenomenological Research*, Vol. 58, No. 3 (Sep., 1998), pp. 607-611.
- Wilson, Timothy D. *Strangers to Ourselves: Discovering the Adaptive Unconscious*. Cambridge, MA: Harvard University Press, 2002.
- Wittgenstein, Ludwig. *Philosophical Investigations*. Translated by G.E.M. Anscombe. Oxford: Blackwell, 1953.
- . *The Blue and Brown Books: Preliminary Studies for the 'Philosophical Investigations'*. New York: Harper and Row Publishers, 1958.
- Wright, Crispin. "Cogency and Question Begging: Some Reflections on McKinsey's Challenge and Putnam's Proof." *Nous*, Vol. 34, Suppl. 1 (Oct. 2000), pp. 140-163.
- . "Some Reflections on the Acquisition of Warrant by Inference." In S. Nuccetelli, ed., *New Essays on Semantic Externalism and Self-Knowledge*, Cambridge, MA: The MIT Press, 2003, pp. 57-77.
- . *Wittgenstein on the Philosophy of Mathematics*. London: Duckworth, 1980.