

A Defense of Semantic Conventionalism

by

Nancy Davies

A thesis

presented to the University of Waterloo

in fulfillment for the degree of

Doctor of Philosophy

in

Philosophy

Waterloo, Ontario, Canada, 2007

©Nancy Davies, 2007

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

The purpose of this dissertation is to argue that semantic conventionalism of a, more or less, Dummettian variety is unjustly neglected in contemporary philosophy.

The strategy for arguing this is to make a conjecture about why people ignore it; there seem to be two plausible reasons: 1) there are (what people take to be) obviously preferable candidates on offer; 2) there are (what people take to be) knock-down arguments against semantic conventionalism.

In response to 1), I consider intentionalist Gricean semantics, and argue it is *at least* no better off than conventionalist theories. Of course, any number of theories could be used oppose semantic conventionalism. But the Gricean theory is seen as particularly strong, and showing that it is no better off makes my case for the viability of semantic conventionalism all the more compelling. For 2) I consider three possible reasons for thinking that conventionalism has been refuted.

Chapter Three concerns the objection that semantic conventionalism depends on the existence of “luminous” psychological states, of which there are none (according to Williamson’s anti-luminosity argument). I agree with Williamson, and reject luminosity as part of a viable conventionalist theory.

Chapter Four supposes that semantic conventionalist theories depend on the (untenable) analytic/synthetic distinction to avoid collapse into holism. However, I also reject the analytic/synthetic distinction for a more favourable distinction.

In Chapter Five, the objection I consider is that semantic conventionalism involves an epistemically constrained notion of truth and so collapses into incoherence because of the knowability paradox. However, my response to this is that the semantic conventionalist should be happy with such an epistemic account of truth and that it *does not* lead to the knowability para-

dox. The paradox can, and is, resolved in this chapter.

So, (1) and (2) are false. The concluding chapter brings together all that we have learned throughout the dissertation about what a defensible version of conventionalism might look like.

Acknowledgments

Without the help and encouragement of many people, I most likely would neither have started nor finished graduate school. My thanks go especially to Professors David DeVidi, Timothy Kenyon, and Paul Tomassi — all three contributed to my interest in philosophy of language and logic. Indeed, without Paul Tomassi, my very first philosophy instructor, I would not have decided on pursuing philosophy at all.

Tim Kenyon and Dave DeVidi supervised my work at different stages of my Ph.D. By working with them I have gradually been learning what it means to be a dedicated academic and what I need to do to reach my career goals. And, of course, their help, and that of the department in general, in finding sufficient funding kept me in graduate school.

Also, partial funding for this research was provided by the Social Sciences and Humanities Research Council of Canada through a Standard Research Grant to Robert Stainton (University of Western Ontario) for the project "The Semantics-Pragmatics Boundary from the Perspective of Cognitive Science and Pragmatic Deficits in Autism."

Finally, Debbie Dietrich and Linda Daniel have been supportive in many ways — from finding wayward transcripts to indulging in tangential conversations arising from (the occasional spell of) procrastination.

Dedication

For Heather — the best daughter one could ever hope for.

Contents

Abstract	iii
Acknowledgments	v
Dedication	vi
1 Introduction	1
2 Intention-Based Semantics	27
2.1 Introduction	27
2.2 Plausibility	30
2.2.1 The Third Realm	30
2.2.2 Behaviourism	32
2.2.3 Further Motivations	34
2.3 Intention-based Semantics	36
2.3.1 Natural versus Non-Natural Meaning	37
2.3.2 Reductive Definition of Meaning	39
2.4 Objections	41
2.4.1 Timeless Meaning — Humpty Dumpty	42
2.4.2 Speaker Meaning — Post hoc	45
2.4.3 Speaker Meaning — Determining Intentions/Confabulation	48
2.4.4 Occasion meaning — Fine Grained	49

2.4.5	Occasion meaning — Coarse Grained	52
2.5	Conclusion	57
3	Luminosity	59
3.1	Introduction	59
3.2	Luminosity	62
3.3	Plausibility	63
3.4	Williamson’s Anti-Luminosity Argument	64
3.5	Examples	68
3.6	Conditions on the Linguistic Community	72
3.7	Conclusion	75
4	Molecularism	78
4.1	Introduction	78
4.2	Holism and Atomism	84
4.3	Plausibility	86
4.4	Objections to Holism	88
4.5	Objections to Atomism	89
4.6	Inferentialism	92
4.6.1	Material Correctness	95
4.7	Direct/Indirect Inferences	98
4.7.1	Non-logical Inferences	98
4.7.2	Logical vocabulary	104
4.8	Materially Correct and Other Inferences	107
4.9	The Linguistic Community and Knowledge of Language	108
4.10	Conclusion	109
5	The Paradox of Knowability	110
5.1	Introduction	110
5.2	The Paradox of Knowability	116

5.3	The Paraconsistent Solution: Dialetheism	118
5.4	States of Information Semantics	128
5.5	Explication of <i>S-O-I</i> Semantics	129
5.5.1	Possible World Semantics	136
5.5.2	Kvanvig’s and Percival’s Objections	139
5.6	Paraconsistent logics	143
5.6.1	Minimal — Intuitionistic — Classical Logic	144
5.6.2	Jaśkowski’s System	147
5.6.3	Relevance Logic	153
5.7	Conclusion	155
6	Conventionalism	157
6.1	Introduction	157
6.2	Conventionalism — Motivations	160
6.3	Conventionalist Conditions	161
6.3.1	Meaning is intersubjective.	161
6.3.2	Meaning is use-based.	162
6.3.3	Two-aspect theory.	163
6.3.4	Conventions are arbitrary.	164
6.3.5	Manifestability	168
6.4	Conclusion	169

Chapter 1

Introduction

In this dissertation I argue that a particular semantic theory — conventionalism — is given insufficient consideration, particularly in contemporary philosophy of language. The term ‘conventionalism’ is ubiquitous in philosophy. Several examples are:

(1) David Lewis’s theory of meaning: Lewis’s theory of meaning is related to the possibility of change in languages — languages are arbitrary and conventional in that the meaning of lexical items can change.¹

(2) Poincaré’s conventionalism about geometry.²

(3) Einstein’s conventionalism concerning the propagation of light.³

¹See for instance (Pateman 1982, 135-157).

²Poincaré argued that there are both Euclidean and non-Euclidean geometries, which can equally well account for the structure of space. In addition, as we shall see, it impossible to *tell* whether the universe is Euclidean or not. Whichever we choose is a matter of convention. See for instance (Sklar 1977, 123-128).

³Einstein’s conventionalism is related to the speed of light. The idea is that, if one were to bounce a beam of light off an object *A* to object *B*, then, bouncing it back *from B* to *A*, the beam would travel at the same speed, under one possible interpretation. However, it turns out that, using Einsteinian physics, this *need* not the case. Two options are now available — either the speed of light is constant in both direction *or* the speed of light is reduced from *B* to *A*. Whichever interpretation is chosen is a matter of convention. See

(4) The Carnapian principle of tolerance.⁴

So to say that this dissertation concerns conventionalism is not yet to say much. The sort of conventionalism I have in mind is akin to Dummett's view. Dummettian semantic conventionalism is definitely a minority position in contemporary philosophy of language, but I think that there is much to be said for it. For various reasons, though, it is more often than not ignored as a potentially viable semantic theory.

The sort of conventionalism I have in mind is broadly Dummettian in spirit. At the basis of the semantic conventionalist theory lie the maxims that 'meaning is use' and that to be a competent linguistic agent one must 'master the rules of the language game' — both of these are important to the community-based theory of meaning, which is fundamental to semantic conventionalism, since it is from the community that meaning arises. The theory I present is neo-Dummettian for several reasons, as we shall see in a moment.⁵

In addition to the maxims that 'meaning is use' and that to be a competent linguistic agent one must have mastered the rules of the language game, Dummett's semantic conventionalism relies heavily on the notion of for instance: (Sklar 1977, 297-294), (Maund 1974, 394-407), (Kennedy 2003).

⁴The conventionalism of the "Principle of Tolerance": "might perhaps be more exactly the 'principle of the conventionality of language forms.'" (Carnap 1963, 54-55) In regards to formal systems Carnap states: "we do not wish to set up prohibitions, but rather to stipulate conventions." (Carnap 1934, 51)

⁵These maxims, of course, do not originate with Dummett, and are perhaps most famously advocated by Wittgenstein. Moreover, there is a strand in the philosophical literature dealing with his work that takes his view to be a sort of conventionalism. However, since there is very little that can be said about Wittgenstein that is not controversial, and there would be little to gain for the present project in trying to defend the suggestion that he was a conventionalist — that there was anything so systematic as a semantic theory in his work for instance — I shall associate conventionalism with authors whose commitment to the view is not controversial, like Dummett, and seek the indulgence of those who feel that Wittgenstein is slighted by my doing so.

‘thoughts’ (in the Fregean sense):

For Frege, thoughts — the contents of acts of thinking — are not constituents of the stream of consciousness... (Dummett 1993, 22-23)

Thoughts, for Frege, are the meanings (i.e., the senses) of sentences. What Dummett borrows from Frege is an insistence that meaning is not psychological. However, Dummett rejects the notion that meanings of lexical items are objective, as Frege held them to be. For Frege meaning resides in the ‘third realm’ and is not *directly* accessible to the linguistic agent. Dummettian conventionalism and my version of neo-Dummettian conventionalism, in rejecting the third realm *and* psychologism, makes meaning intersubjective. (Dummett 1993)

Of course I will not hope, in the space of this dissertation, to show that conventionalism is the *right* or the *only* viable semantic theory, but I do take it that it is, *prima facie*, a *plausible* theory and therefore should not be ignored by philosophers of language.

One must wonder, then, why semantic conventionalism is supposed to be so useless.⁶ I think likely reasons are that:

- (1) semantic conventionalism is often thought to be an *obviously* lesser candidate for an adequate semantic theory than some others.
- (2) semantic conventionalism is regarded by some as unable to respond to some fundamental, and potentially ruinous, objections.

In response to (1) I consider a particular example of a rival theory that many think is clearly better — namely Gricean intention-based semantics.

⁶Some recent papers go as far as to state that *semantic* conventionalism is ‘obviously’ wrong, despite accepting that there are other types of conventionalism which are worthy of discussion. See, for instance the examples provided by (Kannetzsky 2003) and (Mundy 1985).

Intention-based semantics is interesting to investigate because: it is a particularly popular theory in current philosophy of language; it contrasts with conventionalism in interesting ways; and intention-based semantics seems intuitively correct to many people. However, I will show that intention-based semantics is in no better shape than conventionalism — in chapter two, I raise important objections which will show why intention-based semantics is problematic.

While this is only one rival theory, the chapter can serve as a type of exemplar — other rival theories could be shown to be at least equally problematic to conventionalism in a side-by-side comparison. To take up each possible rival would be too large an undertaking though.⁷

In response to (2) I will consider three objections to conventionalism which are often taken to be knock-down. Each objection turns on a plausible claim that *if* conventionalism is true, *then* a particular (undesirable) feature must follow: the existence of “luminous” mental states, a commitment to the analytic/synthetic distinction, and an epistemically constrained notion of truth. I devote a chapter to each objection, showing in each case that it is not knock-down.

For the first two, I argue that the claims about the commitments of conventionalism are false. For the third, I argue that the supposed undesirable feature is not undesirable at all (in fact, it *follows* from semantic conventionalism.)

The responses to (1) and (2) frame my argument that conventionalism is a serious candidate for being the correct semantic theory. A nice fringe benefit of the approach that I take is that we will learn a lot about what a more plausible conventionalism would have to look like by the end of the

⁷It is worth remembering that it is easy to underestimate the difficulties confronting currently popular theories — so for any other theory, I am assuming that it would be possible to undertake a similar side-by-side comparison with semantic conventionalism.

thesis.

Let me now describe what is included in the various chapters in the thesis in slightly more detail.

In Chapter two I consider a semantic theory which is often thought to be *obviously* stronger than semantic conventionalism — intention-based semantics.

First, I will consider why intention-based semantics might seem to be a preferable theory. But, second, I will show why intention-based semantics is on no stronger ground than conventionalism.

While our prime interest in intention-based semantics is its role as a rival to conventionalism, it is interesting to note that some of the key motivations for intention-based semantics are *shared* with conventionalism.

One major worry for Grice *and* Dummett is the third realm. Since we do not have direct access to the third realm, there must be some other theory that gives us direct access to the meaning of sentence. Second, Grice and Dummett object to a particular theory prevalent during the 1950s, namely behaviourism. In the dissertation, I consider two forms of behaviourism, the radical behaviourism of Skinner and the milder form proposed by Ryle.⁸

The rejection of Frege's third realm shows a theoretical economy, since "thoughts" are abstract entities. Thoughts are meant to be fundamental to meaning, but it is unclear what their nature consist in. For Grice, by replacing thoughts with intentions we have a different way of explaining what meaning *is* and *where it comes from*, thus avoiding the metaphysical muddle of abstract objects in inaccessible realms.

Skinnerian behaviourism rejects mental states entirely. The resulting theory is that semantics should be based on the behaviour of the linguistic

⁸For the relevant articles, see, (Skinner 1953), (Skinner 1957), (Skinner 1974), (Grice 1957), (Grice 1989c), (Ryle 1949), (Grice 1989b), (Grice 1975), and (Grice 1989a).

agents entirely; that is, the *activities* of the agents. Correct language use *means* behaving in an appropriate way. However, it is unclear, at least to me, how one can develop a semantic theory along such radical behaviourist lines. One might assume that what is actually being proposed is some sort of code-theory, where an agent encodes what they intend to convey, they speak, and their interlocutor decodes the utterance.⁹ However, this cannot be the case for Skinner as he obviously rejects states and mental content as contributors to¹⁰ a semantic theory.

Ryle's behaviourism is, I think, a stronger theory, one which intention-based semanticists and semantic conventionalists were quite right to reject — but the details of Ryle's behaviourism show that they were turning away from what appears to be much more convincing than Skinnerian behaviourism.

Ryle's milder form of behaviourism rests on *dispositions* to behave in a certain way. The problem with other semantic theories, he thought, is that non-behaviourist accounts of meaning are developed on the basis of a category mistake. The incorporation of mental states (of a particular kind — namely internalizing them) in a semantic theory is where the category mistake lies. Mental states are usually taken to be internal, but according to Ryle distinguishing between mental states and dispositions is distinguishing what is actually the same thing. So Ryle did not do away with mental states entirely. Dispositions *were taken to be* the mental states.— an extended discussion of this will be taken up in due course. What is of importance is that Grice and Dummett, quite rightly I think, found dispositions to be insufficient to ground a semantic theory or to explain the grasp of meaning by linguistic agents.

⁹What carries semantic content would be the mental representations of the agents.

¹⁰I am not a philosopher of mind, and so I will not discuss whether mental states and mental content are identical or not. For Skinner it boils down to the same thing really — what goes on in the head does not matter, only behaviour does.

Belief, according to Ryle's view, then, is a particular sort of dispositional property. Given that mental states *are* the dispositions to behave in a certain way, the "inner life" of the agent is irrelevant to semantics: a semantic theory, then, is one which can account for the meaning of lexical items in terms of dispositions.

The rejection of behaviourism further shows a strength of intention-based semantics, since we obviously *have* intentions (of one sort or another). But, mental states in intention-based semantics are dependent on *internal mental states*, not to be conflated with dispositions. Without an adequate account of how, or whether, they play a role in semantics must be considered.

Behaviourism was pervasive in philosophy of language in the 1950s, so it was an important step in the progression of philosophy of language to move from behaviour to intentions as the foundation for semantics.

While both conventionalist theories and intention-based theories sought to distance themselves from the third realm and behaviourism, they went off in opposing directions in seeking a non-platonic and non-behaviourist account of meaning. In contrast to semantic conventionalism, Gricean semantics does not rely on intersubjectivity of language, nor does it follow the aforementioned conditions of a community based theory of meaning.

Now, why should we think that a theory based on intentions is correct? A natural reason is that speakers *do* have particular intentions when they make an utterance. We can clearly *do* different things with the same utterance — the literal meaning of 'nice hat' is simply 'nice hat'. But, the utterance can be made in a sarcastic tone; with a tone which makes it clear that the speaker means what he says; the utterance can be interpreted as an interrogative; or the utterance can be an imperative, e.g., to take the hat off during Sunday mass.

But, is this really so — does the sentence have one meaning with multiple

applications? Perhaps one could argue that applying different force to the same utterance actually *changes* the literal meaning.

I think, however, that intention-based semantics is *at least* as problematic as semantic conventionalism. Some problems that will be considered in chapter two are:

- (a) the Humpty Dumpty problem: the Humpty Dumpty problem indicates why one must have a grasp of a language *prior* to being able to grasp the intentions of the speaker. A common example of this problem is from Carroll's *Through the Looking Glass and What Alice Found There* (Carroll 1994).

Humpty Dumpty says to Alice “There’s glory for you.” To which Alice responds with “I don’t know what you mean.” “Of course you don’t” says Humpty Dumpty “because I have not yet told you what it means. I use “there’s glory for you” to stand for “there’s a nice knock-down argument.”

The problem here is that Alice cannot understand what Humpty Dumpty has in mind (what his intentions are) when he says “there’s glory for you,” because his intentions are not sufficient to make the utterance mean what he intends it to mean. Only if “there’s glory for you” conventionally means “there’s a nice knock-down argument” can Alice gauge what Humpty Dumpty means by his utterance. Humpty Dumpty cannot, simply on the basis of his intentions, make the first assertion mean the same thing as the second assertion. Of course, it is frequent for a linguistic agent to misspeak and their interlocutor is still able to grasp what was meant. But, I think that this does not rely on interpreting intentions — rather the interlocutor understands the misspoken utterance on the basis of existing conventions and a pragmatic interpretation of the context of the utterance is what allows one to interpret the utterance properly. But the latter is a matter of *interpretation not meaning*.

The lesson learned here is not just that intentions are *insufficient* to provide a basis of meaning: Alice must rely on her preexisting knowledge of the conventions of English if she has a hope of understanding Humpty Dumpty's use of the sentence. Intentions seem to indicate which grammatical mood is being used, but do not affect the meaning of an utterance.

(b) intention-based semantics is open to the charge that intentions are post hoc.

It is not unusual for someone to be asked why they did, or said, some particular thing. What often springs to mind is not the intention *at the time of utterance*. Rather, the speaker considers what he may (must or most likely) have intended in order for those intentions to have conferred a particular meaning upon their utterance.

But this seems to imply that intentions are frequently or typically post hoc — they are only considered by a speaker if challenged, and thereafter are posited to be *the* intentions of the speaker at the time of utterance — therefore, it would be the post hoc intentions that determine meaningfulness. But one cannot have intentions being retroactively meaning conferring. If intentions are post hoc intentions don't determine the meaning of an utterance (at the time of utterance), since the intentions simply were not *present* at the right time.

(c) intention-based semantics is open to the charge that intentions are indeterminate and confabulatory.

Determinate intentions may be *the exceptional cases*, which are not representative of the majority of cases where intentions are indeterminate.

Since we cannot be sure that cases where intentions are explicit are mirrored by cases where they are implicit, it may be that the explicitly determinate intentions are *abnormal*, and skew the data: explicit intentions may

differ radically in nature from those that are implicit. We simply do not and cannot draw an analogy between implicit and explicit intentions as there is no way to acquire empirical data that would determine whether they have the same qualities — if they are not the same then explicit intentions would be confabulatory.

- (d) intention-based semantics is open to the charge that intentions can either be too fine or coarse grained. For the former consider:

Grice's account of occasion meaning in 'Utterer's Meaning and Intentions' (Grice 1989b):

"*U* meant something by uttering *x*" is true if and only if, for some audience *A*, *U* uttered *x* intending:

- (1) *A* to produce a particular response *r*
 - (2) *A* to think (recognize) that *U* intends (1)
 - (3) *A* to fulfill (1) on the basis [or at least in part] of his fulfillment of (2).
- (Grice 1989b, 151)

Consider: an American soldier, call him Ted, is captured by Italians during the Second World War. He wants the Italians to believe that he is a German officer and so utters the only German phrase that he knows: "Kennst du das Land wo die Zitronen blühen?" (Lycan 2000, 107). The utterance is made with the intention that:

- (1) The Italian soldiers produce a belief that Ted is German.
- (2) The Italian soldiers recognize that Ted intends them to form this belief.
- (3) The Italian soldiers come to believe that Ted is a German on the basis of [or at least in part because] the recognition of Ted's intention for them to form this belief.

Thus, it seems that Ted's utterance means "I am a German soldier." But, this is too fine-grained. Ted's utterance cannot mean "I am a German soldier" anymore than it may convey to his interlocutors that he thinks that he is a pink giraffe or that his officious tone of voice means "Release me now you incompetents" and so forth. His utterance is simply not specific enough.

In other circumstances intentions can also be too coarse-grained to constitute speaker meaning: a parent may say to her young child "It's ten thirty!" The *broad* intention is to get her to go to sleep. The parent knows that the child knows that ten thirty is considerably past bedtime and that the child must go to bed. Of course, the child, understanding the intentions, will most likely draw the correct inference that going to bed would be in her best interest.

In this case, though, we appear to get a multiplicity of intentions. Initial intentions which influence the meaning of the utterance would appear to be: (1) the parent intends the child to think that the parent wants her to go to bed; (2) the parent further intends for the child to understand that this is her intention; and (3) the parent intends for her to go to bed because of understanding the parent's intention. However, the literal meaning of the utterance merely reflects the time, not "Go to bed!" Intentions can be even more coarse-grained than (1) to (3), however.

The parent may have (1') work to get done, (2') the parent further intends for the child to understand that this is her intention; and (3') the parent intends for her to go to bed because of understanding the parent's intention.

Another motivation might be that the parent has (1'') a need for some relaxation time, (2'') the parent further intends for the child to understand that this is her intention; and (3'') the parent intends for her to go to bed because of understanding the parent's intention.

One could easily expand upon this and provide additional examples, this

will do for now, though. What then is the speaker meaning?

(1') to (3') and (1'') to (3'') may *arguably* fit the Gricean analysis so the meaning of the sentence “It’s ten thirty!” would be “Go to bed!” I will consider in Chapter Two, however, whether such a multiplicity of intentions actually does fit the analysis. The analysis may very well provide different meanings for each case.

I will consider whether a reconciliatory move is possible so that (1')-(3'') do fit the analysis on different occasions of utterance, but I will hold off on this discussion until Chapter Two. It seems, *prima facie*, to be the case, though, that, given the analysis, several speaker meanings can be attached to the utterance (sometimes *at the same time*). But having multiple speaker intentions, and the intentions being unknown by the audience, fails the Gricean analysis of intention-based semantics. Ultimately, a reconciliatory move is not possible to fit multiple speaker intentions into the Gricean analysis, and so the problem of coarse grainedness stands.

We turn now to the next three chapters of the dissertation. As previously mentioned they all concern what are taken to be knock-down arguments against semantic conventionalism is clearly wrong. They are similar in their approach. Chapters Three, Four, and Five revolve around claims of the form: *if* conventionalism is true, *then* a particular undesirable feature must follow. *But* there are compelling reasons to reject the features in question — luminosity, the analytic/synthetic distinction (considered in Chapters Three and Four respectively), and an epistemically constrained notion of truth (Chapter Five).

I agree and show that luminosity and the analytic/synthetic distinction must be rejected. But taking semantic conventionalism to imply an epistemically constrained notion of truth is actually something that I advocate. So, the response shall look somewhat different than those of Chapters Three

and Four. The reason is that, if I am right that conventionalism implies an epistemically constrained notion of truth I can show that conventionalism results in epistemic anti-realism. This is precisely what I think conventionalism *should* imply. And, this is what allows me to get around the potentially knock-down argument in Chapter Five.

The present discussion of the next three chapters will not be as extensive as that relating to Chapter Two — the reason being that their similar form allows us to deal with them quicker. What I will do is give an account of each objection, my response, and an indication of how this contributes to the development of the conventionalist semantics I advocate.

In Chapter Three, I consider Williamson’s charge that conventionalism implies manifestability and hence luminosity of certain mental states. The basic structure of this objection is that:

- (1) If conventionalism is correct, then manifestability follows and if manifestability is accepted, then there are non-trivial luminous mental states.
- (2) But, there are no non-trivial luminous mental states.
- (3) So, conventionalism is false.

But, Williamson has an argument that purports to show that there are no non-trivial luminous mental states, so conventionalism cannot be correct.

Williamson, assuming the truth of the conjunction, takes this to be a knock-down argument. I agree with Williamson that luminosity must be rejected, but I think that manifestability must be accepted, so it is the “if manifestability, then luminosity” conjunct I reject.

Before we can evaluate Williamson’s argument, a bit needs to be said about what he means when he calls a mental state ‘luminous’.

Luminosity implies that, upon *adequate (internal) reflection*, the speaker

can come to know that the meanings are the same. However, it may be the case that the speaker attaches meanings to words on the basis of conjecture or speculation that they mean the same, or simple misunderstanding. Does *adequate reflection* mean that he can independently come to know the meanings of the terms? Likely not. As a British and Canadian English speaker I can affirm that the speaker has indeed made an acceptable assertion if by “The pavement is wet,” he means “The sidewalk is wet,”¹¹ the assertions are made interchangeably.

The knowledge of the agent is not transparent in the sense that, upon reflection (of his own mental states), he knows that the term ‘pavement’ means ‘sidewalk’. If he lacked the required knowledge he can come to know that sidewalk and pavement mean the same thing by asking an “authority,” consulting a dictionary, and so forth.

Since luminous mental states are ‘trivial’, according to Williamson, I will briefly indicate what ‘trivial’ and ‘nontrivial’ mean in this context.

A nontrivial state is one in which an agent who is in a condition C can, at some stage, fail to be in C . However, luminous mental states are shown to reduce to triviality: if an agent is in condition C then at some point he is simultaneously in $\neg C$. Hence luminosity implies (a certain type) of triviality.¹² To be explicit, the distinction is this: if an agent is in condition C , then at some point he can fail to be in C . But in a trivial state, an agent eventually finds himself in $(C \wedge \neg C)$.

Why, then, does it seem plausible that a *conventionalist* must hold that there are luminous mental states? The suggestion is that *knowing a meaning* of a lexical item is a luminous state (granting that knowledge is a mental

¹¹Given, of course, adequate translation of the different dialects.

¹²Being in $C \wedge \neg C$ does not lead to explosion as triviality is normally taken to imply. The point is simply that the agent then finds himself in both conditions, which is, of course, absurd.

state), given that to be a competent linguistic agent one must ‘master the rules of the language game’. Now, to master the rules of the game is to know when a sentence is warrantably assertable — what follows is that a speaker needs to know the meanings of the words employed. But, if this is the case, then the speaker must equally know the assertability conditions of sentences with words containing the same meaning. For instance, if one were to assert a sentence such as “The pavement is wet, it must have rained,” or “The sidewalk is wet, it must have rained,” then to know the meanings of the terms is to know when they can be asserted; and they can be asserted in the same contexts, since they must know that pavement is interchangeable with sidewalk (if one is familiar with both British and Canadian English, say).

We turn now to Williamson’s ‘anti-luminosity’ argument, the point of which (that luminous states are trivial) I agree with. But I will then show that his argument does not ring the death knell for semantic conventionalism. The argument against luminosity can be paraphrased as:

- ★ First we assume that there is a non-trivial condition C to be luminous.
- ★ Second, since an agent can both be in C and fail to be in $\neg C$, we assume that there is a transition for the agent between condition C and $\neg C$. The stages of the transition are s_1, \dots, s_n , with C holding in s_1 and $\neg C$ holding in s_n . Since C is luminous, the condition \mathcal{L} applies.
- ★ \mathcal{L} : If an agent is in condition C , then he can know that he is in C .
- ★ \mathcal{M} : If KC at s_i , then an agent is in C at s_{i+1} .
- Since C (at s_1), then KC at s_n by applying \mathcal{L} — which means that an agent at s_1 knows that they are in C .

- Apply \mathcal{M} ¹³ — If KC at s_1 , then the agent is in C at s_2 .
- By repeated application, for every s_i , if you are in condition C , then KC at s_{i+1} , and so in C at s_{i+1} .

Strictly speaking what Williamson shows is that any luminous mental state has to be trivial, whereby trivial he means that being in C and $\neg C$ at the same time is possible — the sorites style argument shows why luminosity results (in this type of) triviality. A non-trivial mental state would mean that an agent never finds themselves in conditions C and $\neg C$ at the same time — there are unquestionably non-trivial mental states, but a condition of such states is that they are non-luminous.

Note that, rather than *objecting* to Williamson’s anti-luminosity argument, I will grant that luminosity results in triviality which means that an agent can be in C and $\neg C$ at the same time.

So, luminosity is not a condition which can be incorporated into a conventionalist semantic theory. But the rejection of luminosity does not mean that manifestability should be rejected — the former implies internal reflection as a sufficient condition for understanding the meanings of lexical items, the latter is a condition of the externality of language and does not relate to internal reflection. It is a mistake, then, to think that a conventionalist must be committed to the luminosity of such states as knowing a meaning of a term.

One need only consider the example given above concerning pavements/sidewalks to see why this is.

So, I have accepted Williamson’s ‘anti-luminosity’ argument, in that luminous mental states reduce to triviality. Luminosity does fail. This does not mean, however, that we must reject of manifestability. Manifestabil-

¹³I will explain in Chapter Three what, precisely, \mathcal{M} is. For the time being it suffices to note that \mathcal{M} is a way of encoding what Williamson calls “margins of error.”

ity, being part of a community-based theory of meaning, does not rely on luminosity. Internal reflection of an agent *is not* necessary or sufficient to ‘exhibit’, as it were, the meanings of terms. This shows that the objection can be side-stepped because (as mentioned) the second conjunct of Williamson’s objection doesn’t follow.¹⁴ It thus fails to be an objection at all — least of all knock-down.

This is one way in which I am a neo-Dummettian — luminosity, traditionally taken to be a feature of semantic conventionalism, does not make it on to the ‘list’ of what I take to be the defining features of conventionalism.

In Chapter Four I propose a type of molecularist inferential semantics conventionalism — an inferentialism that allows me to develop molecularism for semantic conventionalism.

Before indicating precisely what the type of molecularist inferentialism is, I need to deal with the next supposed knock-down argument — as presented by, for instance, Fodor and Lepore (Fodor and Lepore 1992).

The argument is:

- (1) If conventionalism is true, then there must be a principled analytic/synthetic distinction.
- (2) There is no principled analytic/synthetic distinction.
- (3) Therefore conventionalism is false.

(2) is generally accepted for Quinean reasons. I do not propose to tackle this premise in the argument. Rather, I shall show that the first one is mistaken.

First, then, I need to show precisely why it might be thought that this *is* a knock-down argument — what is it that makes the objection appear to be plausible?

¹⁴Remember that the second conjunct is *if* manifestability *then* luminosity.

Fodor and Lepore argue as follows: either semantic *atomism* is correct, or meaning is determined by the role played by a word in relation to other words. *And not both* by definition of atomism. Atomism is inconsistent with the type of conventionalism that I advocate.

But in the absence of atomism, not having an analytic/synthetic distinction implies holism. For if we wish to avoid holism, we need a *principled* way to distinguish which relations confer meaning and which do not if we are to avoid saying that *all* relations between lexical items confer meaning and the analytic/synthetic distinction is the only candidate that seems to be on offer.

Holism and conventionalism have a feature in common: the meaning of a word is its role in the sentences in which it occurs. Meaning is not, and cannot be, determined in isolation from the rest of the language. So, why not just accept holism? Because it is obviously false: it makes language unlearnable and if holism is correct then there is no way of cordoning off sections of the language which confer meaning, for instance.

The assumption built into Fodor and Lepore's argument is that there must be some way to determine the meaning of lexical items. This is a plausible assumption, and is problematic for molecularism: if you don't have an analytic/synthetic distinction and the meanings of words are determined by relationships between them or their relationships to other parts of the language, then you have to say that *all* its relationships go to determining its meaning. Atomism, on the other hand means that words have meaning in isolation, by individual relationships to things with the world.

So, the rejection of atomism and holism means that there must be an alternative. Therefore I argue for molecularist inferentialism.

I will: (1) describe a type of inferentialism which is consistent with conventionalism — one which does not fall into holism — as Fodor and Lepore

argue is inevitable with any theory other than atomism; (2) show that such an inferentialist picture is molecularist.

On the basis of what I call materially correct and non-materially correct inferences (but still permissible inferences) we *can* cordon off sections of the language. Such an inferentialist picture allows me to maintain that it is correct to say that it is relations between lexical items that determine meaning, but only *some*, not all. This molecularist inferentialism gives conventionalists a way to separate off the inferences that count as meaning-constitutive from the ones that aren't.

There are two ways which show that the distinction between meaning-constituting and nonmeaning-constituting inference. The first is that 'materially correct' inferences count as meaning-constitutive and 'formal inferences' don't, although they are still acceptable inferences. The meaning of a word is determined by materially correct inferences. These give the meanings of lexical items because they share (some) conceptual content.

The second way to distinguish meaning-constituting from nonmeaning-constituting inferences is between direct and indirect inferences. A direct inference is where, say B is inferred from A (with no intermediate steps). An indirect inference is one where the conclusion of an inference involves more than one step, e.g., concluding C from A , but by way of a second step — A to B . So, the inference now is A to B to C .

To distinguish direct and indirect inferences I must establish a hierarchical account that does not allow one to jump from A to C . This hierarchical account allows me to show why permissible inferences from A to B are materially correct, and thus meaning constitutive, while inferences from A to C (via B) not meaning constitutive.

We are nearly done with the supposed knock-down arguments. The final one concerns the epistemically constrained notion of truth — to be discussed

in Chapter Five.

The ‘obvious’ knock-down objection is:

- (1) If conventionalism is true, then all truths are known.
- (2) All truths are not known.
- (3) Therefore, conventionalism fails.

My response to this objection is that the epistemic constraint does not imply that all truths are known, but that all truths are *knowable*.

The *paradox of knowability* seemingly applies in the following way: if conventionalism is accepted, then so too is the problematic premise that all truths are known. It is obviously not the case that all truths are known and so the paradox shows that conventionalism reduces to absurdity. An immediate response is that conventionalism does not imply that all truths are *known*, but that they are *knowable*. But to some *known* and *knowable* amount to the same thing.

Wherein, then, does the plausibility lie? Why would we think that (1) is acceptable? (1) might be thought to be plausible given a particular interpretation of knowledge.

The paradox of knowability (also referred to as ‘Fitch’s paradox’ or ‘Fitch style arguments’), which concerns an argument first presented in ‘A Logical Analysis of Some Value Concepts’, (Fitch 1963), has generated a great deal of philosophical discussion. In that paper, Fitch presents an argument that some believe implies that epistemic anti-realism reduces to absurdity. The intuition of many anti-realists is that it is, in some sense, correct to say that for any truth p , it is possible to know that p is true, which one can represent schematically as $(p \rightarrow \Diamond Kp)$. This is often called the verificationist principle, but is more commonly referred to as the anti-realist principle (Williamson

1987), (DeVidi and Solomon 2001).¹⁵

What Fitch seems to have shown is that if one accepts the epistemic anti-realist principle,¹⁶ then one is forced to accept that for all p , if p is true then p is known: $(p \rightarrow Kp)$. Some have taken this to amount to a refutation of epistemically constrained anti-realism, since it is obviously false that all truths are known. Others, of course, claim that something is amiss — either with the proof, e.g., that it rests on unwarranted assumption of classical logic, or with its interpretation of the modality \diamond .

Thus, conventionalism seems to require a way out of this problem. My approach will be to grant that conventionalism implies knowability, but *deny* that Fitch’s argument works. In effect, I shall show that there is *at least one* way to save conventionalism from this challenge.

My approach will be what is known as a “logical revisionist” response to the paradox. In more detail, what I will show is that there are several logical revisionist proposals that can be used to block the paradox — but that only *one* — one which employs a particular interpretation of intuitionistic logic — is not ad hoc, for the reason that the counterexample to the proof is developed on the basis of states of information semantics. This provides the philosophical motivation and prevents the application of intuitionistic logic from being ad-hoc.

¹⁵I prefer the name ‘the anti-realist principle’ as it avoids presupposing that we have a fixed notion of verification. Many anti-realists would suppose that what counts as a verification varies across domains of discourse. To maintain the name ‘the verificationist principle’ would be to invite the assumption that what is under investigation is *Verificationism* as presented by Carnap, Ayer, Schlick, Hempel, etc. of the Vienna Circle.

¹⁶Naming it ‘epistemic anti-realism’ is important, since it distinguishes other types of anti-realism, e.g. error theory. The first is important for a semantic theory, the latter is not. For the most part, unless necessary, I will refer to epistemically constrained anti-realism simply as anti-realism. This is the only type which we will be concerned with throughout the chapter.

It is common to see it argued that a certain sort of anti-realism follows from conventionalism about meaning. The reasoning usually runs more or less as follows: manifestability requires that, one way or another, all aspects of the meaning of all lexical items are learnable, and can be made publicly available. Thus, for any given declarative sentence S , i.e., for any candidate for being true, there is nothing about the meaning of S that is hidden. But in particular, this means, or at least plausibly suggests, that all the conditions that go into determining the truth or falsity of S are similarly publicly available, and so in principle it should be determinable whether those conditions are satisfied or not. Hence, it seems, S must be *knowable* — though, of course, for contingent, practical reasons this will often be only knowability in principle. This bit of reasoning suggests that conventionalism about meaning leads to a certain sort of anti-realism, often called semantic anti-realism — the view that truth is *epistemically constrained* in the sense that every truth is knowable.

My argument is that if conventionalism is correct then anti-realism follows (the type that on first glance appears to be susceptible to the paradox at least). The correct logic for anti-realism is intuitionistic logic, and the correct semantics is states of information semantics. States of information semantics is epistemically anti-realist — if an agent accesses a state of information at which there is sufficient warrant to establish x , he *could* come to know that x . But it is not a requirement that *any* individual agent need ever reach this state (so, of course, the information would not be accessed by the community at large).

Before explicating states of information semantic in Chapter Five, I consider a relatively new suggestion for blocking the proof, namely dialetheism (the view that there are true contradictions) — which blocks the proof by

rejecting *reductio ad absurdum*.(Beall 2000)¹⁷ The dialethic strategy, however, is shown to be wrong, on the basis of several devastating objections.

The major point, as mentioned, is that states of information semantics provides a philosophical motivation for using intuitionistic logic. However, I use states of information for a further purpose. Since paraconsistent logics are becoming increasingly favoured by some authors, I will show that several paraconsistent logics *can* provide counterexamples to the paradox of knowability, but they only provide a *motivated* solution if states of information semantics are applied — and hence the counterexamples rely on the commonalities they share with intuitionistic logic.

I cannot, of course, consider all systems of paraconsistent logics, as there are simply too many of them. But I do consider three interesting ones: minimal logic, a non-adjunctive system, and the Routley-Meyers paraconsistent logic.¹⁸ Paraconsistent logics seem *prima facie* somewhat odd. But, they are gaining momentum in the field of formal logic — and hence demand serious attention.

So, conventionalism leads to anti-realism. Anti-realism is susceptible to the paradox of knowability. A way of getting around the paradox is to apply intuitionistic logic, but this appears to be *ad hoc*. The philosophical motivation comes from the development and application of states of information. In addition, whilst some paraconsistent systems *can* provide counterexamples to Fitch's paradox, they also do so only on the basis of states of information semantics. The latter is taken to be a particularly well-suited semantics for intuitionistic logic. In the examples I discuss I show that it is not paraconsistency itself that solves the paradox, but the possibility of applying states of

¹⁷Explosion is standardly accepted as requirement for any strong system of logic — the rejection of RAA, on the other hand, prevents this rule of inference.

¹⁸As with Grice, considering the fact that there is neither time nor space to consider all alternatives, I will treat this as a case study.

information semantics. Hence, these counterexamples rely on commonalities they share with intuitionistic logic.

Chapter Six will bring all of the above together and give a bird's-eye view of the version of conventionalism I am developing. In the introduction we learned that semantic conventionalism is taken to be refuted on the basis of two reasons. In the introduction, I suggested two sorts of reasons that semantic conventionalism might be thought not to deserve the attention of contemporary philosophers of language. The first reason is that other theories are thought to be obviously stronger. The second reason is that, in light of various supposedly knock-down arguments conventionalism is not only weaker, but *wrong*.

As just described, one of the main tasks in the main body of the thesis is to show both reasons to be mistaken. While making this case, though, we will learn valuable lessons as to what a viable semantic conventionalism should look like.

Chapter Six shows explicitly that the maxims that meaning is use and that to be a competent linguistic agent one must master the rules of the language games that make up the language are the *overarching* conditions of the type of conventionalism I advocate. All the other conditions of semantic conventionalism fit (and are necessary parts of) the theory precisely because of the way that they dovetail with these maxims.

Concerning the conditions of semantic conventionalism, I will show that there are some fundamental ones in addition to the two *overarching* maxims — thus developing a more robust theory of semantic conventionalism. A key lesson, for instance, will be the commitment to *manifestability*. Manifestability of meaning is easily misunderstood — it need not entail that individual speakers have ready access to all aspects of the meanings of words, and is compatible with the division of linguistic labour. However, it is tempting to

think otherwise.

Before turning to the main business of the thesis, it will be useful to have a brief list of a few important features of conventionalism in one place. One might think of this as the rough version of conventionalism, to which the refinements described in Chapter Six apply. Note that I am not claiming that I have *discovered* these conditions of semantic conventionalism — Dummettian semantic conventionalism, for instance, incorporates the same conditions, and indeed, anything that qualifies as semantic conventionalism will at least incorporate most of them in some form. Throughout the dissertation, however, it will become *clearer* how these conditions should be understood.¹⁹ For now, it should be enough to consider various criteria in “sound bite” form. All of which grow out of the idea that meaning is determined by the rules for the use of language in a community.

- meaning is neither platonic nor subjective — rather it is intersubjective. That is, the semantic conventionalist contends that meanings have an ontological status other than the mind-independent objective existence attributed to them by Platonists such as Frege without falling into crude psychologism about meanings. How exactly this works is, of course, a philosophically fraught matter.

- meaning is use.

- meaning *public*. That is, there is no aspect of meaning that is in

¹⁹It may be, and probably is, the case that the conditions are explicated somewhat differently than they would be by other conventionalists. In other respects they may be very similar. This is not really pertinent to the current dissertation, however. Although I do identify the semantic conventionalism presented here as neo-Dummettian, what I aim to do is to draw out the conditions through the consideration of conventionalism faced with the objections. The conditions, then, come to light in broad strokes. I think that they are necessary for a viable semantic conventionalism, but I do not claim that they are representative of any one person’s theory.

principle private to any individual.

- meaning is *manifestable*. This is a sort of corollary to the public nature of knowledge. Since any aspect of meaning is public, there is no aspect of meaning that cannot be presented to a potential linguistic agent. So, for instance, there cannot be a distinction in meaning between a pair of linguistic terms unless a difference in the applicability of the two terms in a particular case can, at least in principle, be made evident.
- meaning is molecular. That is, the meaning of a given linguistic item cannot be given in isolation from the role that item plays in a particular “language game,” but a language game is smaller than an entire language. Two people can share exactly the same meaning for a word like “two” even if one can play the game of advanced calculus and the other cannot, since the meaning of “two” is fixed by the role it plays in statements about small collections of things, for instance.
- meaning is conventional — that is, it’s arbitrary in some important sense that we operate with the concepts we in fact do; there are alternative conceptual schemes which would have allowed us to function in the world, and it is not a matter of correctness that determines the choice between them.

Other conditions will be discussed in Chapter Six and the explanation of each will be expanded upon. But for the moment this list is sufficient to get the discussion off the ground.

Chapter 2

Intention-Based Semantics

2.1 Introduction

In Chapter Two I consider, as mentioned in Chapter One, a semantic theory which is often thought to be *obviously* stronger than semantic conventionalism, namely intention-based semantics — and so is taken to be a major challenge to semantic conventionalism.

One must not underestimate the influence that the Gricean project has had on present day philosophy of language — leading to a variety of research projects from psycholinguistics, distinguishing between ‘what is said’ and ‘what is implicated’ in speech (leading to modern day discussions concerning the distinction between pragmatics and semantics), and even projects concerning language acquisition (both in children of ‘normal’ linguistic ability and those with language deficits arising from, e.g., autism) and, of course, semantic theories.

The importance of the Gricean project must therefore not be underestimated. Although I will ultimately argue that intention-based *semantics* is not as strong as it appears to be, I do not deny that the distinction between what is said and what is implied should be ignored. Pragmatics is

important to philosophy of language and Grice's projects leads to many interesting investigations in this field. What I maintain, however, which is at odds with many philosophers of language, is that in order to investigate pragmatics properly, we must distinguish two Gricean projects — that of, e.g., 'Meaning' (Grice 1957, 103-109), which is a semantic project, and that of, e.g., 'Logic and Conversation' (Grice 1975, 121-133), which is a pragmatic project.

First, I will consider why intention-based semantics might seem to be a preferable and stronger theory than semantic conventionalism — that is, I will give reasons to indicate *why* Grice¹ would want to turn to intention-based semantics in the first place.

The first two reasons why intention-based semantics seems so strong are that it rejects both the Fregean third realm and behaviourism.

The rejection of the first shows a theoretical economy, since "thoughts" are supposed to be abstract entities. Thoughts are meant to be fundamental to meaning, but it is unclear what their nature consists in. By replacing thoughts with intentions we have a different way of explaining what meaning *is* and *where it comes from*, thus avoiding the metaphysical muddle of abstract objects in inaccessible realms. Since we do not have direct access to the third realm, there must be some other theory that gives us direct access to the meaning of sentence — and Gricean semantics is in the position to do this.

The rejection of behaviourism also shows a strength of intention-based semantics, since we obviously *have* intentions (of one sort or another), whereas,

¹I concern myself with Grice throughout this chapter, since he was the first to develop intention-based semantics. And, further, although there are other intention-based and neo-Gricean theories out there, I am working on the assumption that they are all at least *influenced* by Grice's works. Further, as explained, this theory is taken to be an exemplar and so it is sufficient to consider only this one form of intention-based semantics.

I will argue, it seems implausible to construct a semantic theory from behaviour alone. Behaviourism was pervasive in philosophy of language in the 1950s, so it was an important step in the progression of philosophy of language to move from behaviour to intentions as the foundation for semantics. In the dissertation, I consider two forms of behaviourism, the radical behaviourism of Skinner² and the milder form proposed by Ryle.³

Several other reasons that intention-based semantics is suitable to use as an exemplar are that: it is a particularly popular theory in current philosophy of language; it contrasts with conventionalism in interesting ways; and intention-based semantics seems intuitively correct to many people.⁴

But note that Dummettian semantic conventionalism *equally* sought to reject the third realm and behaviourism. Intention-based semantics and conventionalism simply took different routes.⁵

As mentioned in chapter one, while intention-based semantics is only one rival theory, the chapter can serve as a type of exemplar — other rival theories could be shown to be at least equally problematic to conventionalism in a side-by-side comparison, I am suggesting. To take up each possible rival would be too large an undertaking, but it is worth remembering that it is easy to underestimate the difficulties confronting currently popular theories.

First, I will give an overview of the rejection of the third realm and behaviourism, since it signals the most important strengths of intention-

²See (Skinner 1953), (Skinner 1957), (Skinner 1974).

³See (Ryle 1949).

⁴I will explain later why this latter is the case. To give a brief indication, though, we *know* that we have some sort of intentions and we know that we can *do* different things with the same utterance (which would indicate distinct speaker meanings for the same utterance) — it is fair enough, then, to suppose that intentions are (or at least may be) important to a semantic theory.

⁵For the relevant articles, see (Grice 1957), (Grice 1989c), (Grice 1989b), (Grice 1975), and (Grice 1989a), (Dummett 1993), etc.

based semantics (as mentioned above). The other motivations will also be considered in due course.

Second, I will provide an exposition of intention-based semantics. The purpose is to set the scene to show that intention-based semantics is (at a minimum) in no better shape than the rival theory of conventionalism.

Third, given the objections I discuss, I will show that intention-based semantics at least as problematic as conventionalism. In light of these objections, I argue that conventionalism needs to be reconsidered as a viable semantic theory.

2.2 Plausibility

2.2.1 The Third Realm

‘Third realmers’, as we might call them, were found to be suspicious by many authors in philosophy of language, mathematics, and logic from the 1950s on. A goal of intention-based semantics, which it must be noted is shared by conventionalist theories, is the ‘tidying’ up of Fregean metaphysics. This is in order to gain direct access to the meaning of sentences. While Grice turned to intention-based semantics, semantic conventionalism took a different route, namely accepting that meaning is external, but replacing the objectivity of thoughts with community-based intersubjectivity.⁶

The third realm is where abstract objects ‘reside’, as one might put it. Since Frege maintained that the meanings of sentences have mind-independent existence they could not be posited as a feature of the psychology or intentions of individual linguistic agents.⁷ Furthermore, being abstract and ‘placed’ in a different realm, meanings do not reside in the

⁶As I will later argue, this makes meaning out to be *determined* by the community as a whole, not individual linguistic agents.

⁷That is, the ‘senses’ of sentences are objective and ‘out there’ in the world, as it were.

same ‘reality’ as, say, middle-sized objects such as tables and chairs. Abstract objects do not *exist* in the same way. Concrete objects belong to one realm, abstract objects to another. Meanings must be somewhere — so, as mentioned, they are not dependent on psychology nor intersubjectivity, rather they are objective.

What arises, then, is a somewhat strange metaphysical picture where different types of objects from distinct realms must interact if one wants to provide a semantics for sentences. Propositions, which play a role in providing a semantic theory for sentences, must somehow interact with *this* realm, although they reside in the third realm.

The problem, though, with the Platonic realm of abstract objects is that gaining access to propositions through senses seems mysterious — and does not *explain* how the leap from the actual world to the third realm is made. How senses and propositions contribute to a semantic theory is therefore unclear.

Natural language philosophers were, quite reasonably, not impressed by Frege’s metaphysical view and sought to provide a semantics which avoided it.

There are several ways to respond to the metaphysical stance taken by Frege. Intention-based semantics brings meaning back into the ‘realm’ in which we reside (at least it does so presuming that intentions themselves are to be counted among the furniture of the world). For the Gricean, what needs to be explained is how meaning can be described in terms of the intentions of the *individual* speakers of a linguistic community. This is a discussion which will be taken up in a later section.

Dummett, as already noted, is equally suspicious of the third realm. The conclusion he reaches is *very* different from intention-based semantics. His response to platonism was not to place intentions at the root of a theory

of meaning but to make meanings *intersubjective* and part of the linguistic community as a whole. Meaning, according to Dummett, is determined not by the individual speakers, but the *use* of lexical items.

So much the worse, then, for the third realm.

To conclude, Grice (and Dummett) reject the third realm. One might say that their semantic theories got off the ground *because* of the rejection of this metaphysical stance. However, although they had this, at least, in common, they went off in very different directions precisely as a result of rejecting the third realm. Whilst Grice opted for intention-based semantics — leaving him open to the charge of psychologism⁸ — Dummett, as we shall see in a later chapter, remained closer to the Fregean in spirit. Although the third realm is rejected and the semantic theory is made to hinge on the intersubjectivity of the linguistic community, he still accepted Fregean ‘thoughts’ as important in explicating meaning.

So, we have seen the benefit of the ‘theoretical economy’ of intention-based semantics (and indeed conventionalist semantics). We obviously have some sort of intentions, but “thoughts” (in the Fregean sense) are dubious. By replacing thoughts with intentions, the intention-based semanticist does not expand the metaphysical objects (if they may be called such) that must be incorporated into the theory. The theoretical economy of intention-based semantics, then, simplifies the matter and provides a more plausible foundation for semantics — thus certain, unclear, considerations are avoided.

2.2.2 Behaviourism

Grice objected to behaviourism as he did to the third realm — both the radical behaviourism, proposed by Skinner, and the milder version, proposed

⁸At least, that is how Dummett reads intention-based semantics. Whether this is the case or not will be discussed later.

by Ryle.

Skinnerian behaviourism rejects mental states entirely. As mentioned in the introduction, behaviourism is meant to lead to a semantics based entirely on the behaviour of the linguistic agents; that is, the *activities* of the agents. Correct language use relies on behaving in an appropriate way. However, as mentioned earlier, it is unclear to Grice and other philosophers, how one can develop a semantic theory on such radical behaviourism. Perhaps a code-theory could be applied, where an agent encodes what they intend to convey, they speak, and their interlocutor decodes the utterance.⁹ However, this cannot be the case for Skinner as, we have already mentioned, he sees no room for mental states, so behaviour must carry semantic content, not mental representations, which come about as a *result* of particular behaviour.¹⁰

Ryle's milder form of behaviourism rests on *dispositions* to behave in a certain way. The problem with other semantic theories, he thought, is that non-behaviourist accounts of meaning are developed on the basis of a category mistake. Incorporation of mental states in a semantic theory is where the category mistake lies.

Grice, quite rightly I think, found dispositions to be insufficient to ground a semantic theory or to explain the grasp of meaning by linguistic agents.

Note that Ryle did not do away with mental states entirely, but the dispositions *were taken to be* the mental states. *Belief*, according to this view is a dispositional property. The purely subjective inner feeling or experiences are irrelevant to meaning. Further, given that mental states *are* the dispositions to behave in a certain way, the inner life of the agent is equally

⁹What carries semantic content would be the mental representations of the agents.

¹⁰A charge against Skinner is that he based his semantic theory on the observation of animal behaviour — this seems far removed from being able to develop a theory based on behaviour.

irrelevant to mental states themselves. In relation to a semantic theory, the same holds: understanding lexical items may require particular mental states, but these are still taken to be dispositional; in addition, a semantic theory is one which can account for the meaning of lexical items in terms of dispositions.

So, the second motivation for intention-based semantic was to reject behaviourism — as mentioned when discussing Fregean metaphysics, it is obvious that we have intentions (of some sort). A strength of intention-based semantics, then, is similar to the rejection of Fregean thoughts — we know that we have intentions; linguistic agents *obviously* intend to convey something when they make an utterance, so we *should* try and incorporate intentions into a theory of communication and a theory of semantics.

2.2.3 Further Motivations

A further appeal of the Gricean project is that it closely knits the theory of meaning with the theory of communication. His paper ‘Logic and Conversation’ (Grice 1975, 121-133) provides a normative account of conversation, which relies on his account of the role of intentions. If, for instance, someone uses an utterance in a manner which disregards the rules of permissible language use, such as answering “yes” when asked if he has the time, then they are breaking the rules of communication by not being relevant.¹¹

¹¹It must be noted, however, is that it is possible to distinguish Grice’s semantic project from that undertaken in ‘Logic and Conversation’. The latter concerns the pragmatics of language use. As I am not dealing with pragmatics in this dissertation, I am not going to discuss the role of intentions in pragmatic determination of what is ‘said’, except when it is pertinent to the discussion. How much pragmatic interpretation is determined by pre-existing conventions (e.g. sarcasm seems to have particular conventional qualities, for instance tone of voice and so on) or on the interpretation by a hearer of their interlocutor’s intentions is without a doubt an interesting topic, but one which for our purposes is a digression.

We can clearly *do* different things with the same utterance. An utterance that is spoken as an assertion, e.g. ‘continents drift’, may be uttered with assertoric, imperatival, or interrogative force.¹² But, the utterance can be made in a sarcastic tone, with a tone which makes it clear that the speaker means what he says, or can be interpreted as an interrogative or as a demand to take the hat off during, say, Sunday mass. The context of the utterance helps to determine the *occasion meaning*.

According to the Gricean analysis, speakers have different intentions on the occasion of utterance and therefore they *mean* different things on those occasions. Since, as we shall see, literal meaning is made up of a disjunction of occasion meanings, it seems that each occasion of utterance does signal a distinct meaning: ‘nice hat’ *means* ‘nice hat (really)’ *or* ‘what an awful hat’ (utterance: ‘(pffft,) nice hat! (sarcastic)’ *or* ‘take that hat off’ (utterance: ‘nice hat (through gritted teeth)’).

I do not have much more to say in regards to whether the semantic content changes on occasions of utterance. I do not think that it does, but in a way this goes beyond the bounds of the chapter, which is to show that intention-based semantics (of the Gricean variety) is problematic and that allows me to conclude that conventionalism needs to be reconsidered.

One final point must be made, though. I mentioned that intention-based semantics seems to square nicely with many peoples’ intuitions about where meaning comes from. “Of course it’s up to the speaker,” they may say “it

¹²Speaking of ‘literal meaning’ of an utterance is, however, somewhat problematic, since it is not clear what this means for an intention-based theory of the Gricean sort. This is because, I shall argue, it leads to an inherent inconsistency — that of timeless meaning being constructed by a disjunctive account of occasion meaning (in turn dependent of speaker meaning), whilst accounting for the timeless meaning of certain sentences (and subsentential lexical items) by using the notion of the ‘conventions’ of the use of such sentences. This is an issue we will turn to in due course.

is, after all, the speaker who *intends* to communicate something.” Although this is contrary to my own intuitions, I can see why it is appealing.¹³

2.3 Intention-based Semantics

Thus far we have seen several reasons that got intention-based semantics off the ground: the rejection of the third realm and the rejection of behaviourism being the most important. But, we also considered whether occasion meaning is actually semantically relevant *and* why intention-based semantics seems to square with the intuitions of others. Now we need to consider how incorporating intentions into a theory of meaning leads to an account of timeless meanings of sentences — thus making it clearer why intention-based semantics is often thought to be a viable theory, whereas conventionalism falls by the wayside.

The reason that intentions are ineliminable for Grice is that, without them, we cease to have a way of developing a semantics for natural languages. Having intentions provide the foundation of linguistic meaning has the benefit that the agent can *decide* what an utterance means on a particular occasion— indeed, the agents *determine* what an utterance means — in a particular context or what they meant to convey to their interlocutors in a particular context. So, from individual intentions we can broaden out the meaning of an utterance, creating multiple occasion meanings.

I agree that the utterance of an individual speaker does affect the *purpose* of the utterance. However, I do not think that it changes the literal meaning. Moreover, timeless meaning of lexical items being dependent on the speaker meaning of individual agents seems problematic. Some of the

¹³Of course, this is not the whole story — for that we must consider the two-part reduction, which more fully explains how timeless meaning can be derived. We shall turn to that in due course.

problems encountered (discussed in section 2.4, following my interpretation of Grice’s theory) with intention-based semantics are:

- (a) the Humpty Dumpty problem: the Humpty Dumpty problem indicates why one must have a grasp of a language *prior* to grasping the intentions of the speaker.
- (b) that intention-based semantics is open to the charge that intentions are post hoc.
- (c) that intention-based semantics is open to the charge that intentions are indeterminate and confabulatory.
- (d) that intention-based semantics is open to the charge that intentions can either be too fine or coarse grained.

Before turning to the objections, however, it is necessary to gain a decent understanding of Grice’s theory.

2.3.1 Natural versus Non-Natural Meaning

Before turning to the discussion which concerns only meaning_{nn} and thus intention-based semantics, I think it is important to note Grice’s distinction between natural and non-natural meaning, since Grice starts his discussion (Grice 1957) with this distinction. We don’t need to spend too much time on this distinction — but it is useful for seeing why meaning_{nn} is that which concerns semantic content of lexical items. Grice mostly illustrates the distinction by giving examples and so shall we.

The basic idea is that there are some cases where we use the term ‘mean’ to indicate that one sort of phenomenon is a reliable indicator of another. “Spots mean measles,” or “Tree rings indicate the age of the tree,” are not semantically relevant, since such sentences merely indicate *reliable indicators*,

e.g., spots indicating the presence of measles. Consider a further example: “the smoke means fire.” What ‘means’ indicates, as in the previous examples, is that there is a *law-like* relationship between smoke and fire (*ceteris paribus*).

Grice distinguishes such examples of natural meaning from *sentences and words*, which have what he calls non-natural meaning (meaning_{nn}). If we say “‘Robin’ means a species of bird indigenous to North America,” Grice contends, we use ‘mean’ in a quite different sense, since this claim does not involve the same sort of law-like regularity. The idea in the background is that non-natural meanings are semantically important and involve communicative intentions of agents, whereas natural meaning does not involve communicative intentions.

But it is important to note that meaning_{nn} doesn’t only attach to linguistic items. Let us consider one more example to illustrate that communicative intentions, and so the class of utterances, according to Grice, can be conceived of as broader than we might initially think (Grice 1957, 106). Grice asks us to consider this case:

- (1) I show Mr *X* a photograph of Mr *Y* displaying undue familiarity with Mrs *X*.
- (2) I draw a picture of Mr *Y* behaving in this manner and show it to Mr *X*.

I find that I want to deny that in (1) the photograph (or my showing it to Mr *X*) meant $_{nn}$ anything at all; while I want to assert in (2) the picture (or the drawing) and showing it meant $_{nn}$ something (that Mr *Y* had been unduly familiar)... [this case shows that] I do not want to maintain that *all* our uses of “mean” fall easily, obviously, and tidily into one of the two groups I have distinguished; but I think that in most cases we should be at least fairly strongly inclined to assimilate

a use of “mean” to one group rather than to the other. (Grice 1989a, 218)

The photograph, then, carries natural meaning, whereas the drawing carries meaning_{nn} , because (2) is a particular sort of product of Grice’s communicative intention directed at Mr *X*. The photo, on the other hand, is, as in the case of the previous examples, merely a reliable indicator (of what happened in front of the camera).

To conclude, typically the things with natural meaning will be non-linguistic, but indicate a law-like relationship. Non-natural meaning typically attaches to linguistic utterances, but his example of the drawing is meant to show that non-natural meaning can attach to non-linguistic cases also, in which case they are taken to be a different *type* of utterance.

2.3.2 Reductive Definition of Meaning

The entire project that Grice pursues relies on the two-part reduction — the two part reduction concerns, as mentioned, meaning_{nn} . In later papers, (Grice 1989c) and (Grice 1989b), this reduction is expanded upon, but it still retains the fundamental features of the original reduction. Therefore, the 1957 version will be focussed upon unless otherwise indicated.

Now, let us look at the reduction of linguistic meaning (that is, meaning which has become fixed, i.e., standard meanings) to intentions. The reduction begins with a (partial) definition of timeless meaning. The second section of the reduction is vital to the development of timeless meaning. For intention-based semantics, timeless meaning must reduce to intentions. What one might say is that a sentence *x* means *y*, *z*, and *w* (given the context of utterance) and always will (disregarding for convenience the possibility of language evolution). This is not to say that there are no constraints on timeless meaning. Indeed, the relevant constraints will become vital to the discussion in due course. First, though, let us get the reduction Grice

advocates into a clear form:

“ x means_{nn} [(timeless)] that so-and-so” might at first shot be equated with some disjunction of what “people” (vague) intend (with qualifications about “recognition”) to effect by x . (Grice 1957, 108)

Part one of the reduction, then, reduces timeless meaning to the utterer’s occasion meaning — considering that timeless meaning is a construction of all the occasions that an agent meant (non-natural) something by his utterance. Second:

“ x meant something” is (roughly) equivalent to “Somebody meant_{nn} something by x .” Here again there will be cases where this will not quite work. I feel inclined to say that (as regards traffic lights) the change to red meant_{nn} that the traffic was to stop; but it would be very unnatural to say “Somebody... meant_{nn} by the red-light change that the traffic was to stop.” Nevertheless, there seems to be *some* sort of reference to somebody’s intentions. (Grice 1957, 108)

And, this in turn reduces to the intentions of the speakers — part two of the reduction:

“ A meant_{nn} something by “ x ” is (roughly) equivalent to “ A intended the utterance of x to produce some effect in an audience by means of the recognition of this intention”; and we may add that to ask what A meant is to ask for a specification of the intended effect (though, of course, it may not be possible to get a straight answer involving a “that” clause, for example, “a belief that...”). (Grice 1957, 108)

The reduction takes one from timeless meaning to utterer’s occasion meaning and from utterer’s occasion meaning to intentions. If, e.g., “He’s green” is uttered by persons a and b , a may mean by the utterance that he is a newbie. And person b might mean by the utterance that he is literally green. So, taken disjunctively, “He’s green” means “He’s a newbie” or “His skin has a green tinge” or... (whatever else the utterance is used to

convey).¹⁴ Thus the utterer’s occasion meaning and timeless meaning are inextricably bound up with intentions. Both concern the utterer’s intentions which determine the meanings of sentences.

The reason why intentions become central is because they concern the result in an explanation of occasion meaning, which shows what the speaker is *doing* with the utterance — how he intends to be understood — these occasions of use taken disjunctively to be the timeless meanings of lexical items means that everything must reduce to intentions (in one way or another). Note that in later works ((Grice 1989c), Grice (1989b)) the audience becomes more of a vital element of the theory of meaning. These later definitions incorporate the audience’s recognition of the intentions of the speaker in a similar fashion as the 1957 distinction. But, in *addition*, the propositional attitudes of the utterer and their interlocutor towards utterance x as well as the speaker’s desire for the audience to share the same propositional attitude towards x and so on are built into the extended analysis.

To conclude this section, the two part reduction presented in ‘Meaning’ (Grice 1957) is of vital importance to the Gricean project, as it is the only clear indication of how timeless meaning can be based on intentions.

2.4 Objections

There have been many attempts to undermine intention-based semantics. In all three papers discussed here — 1957, 1968, 1969 — Grice anticipates several objections and proceeds to provide what, at first glance, appear to be adequate responses. However, there are several objections which I think are (at least potentially) devastating for intention-based semantics, regardless of

¹⁴To explicate timeless meaning one must say something like: if the timeless meaning of a statement is $x \vee y$, then in order to determine if a meant _{nn} x or meant _{nn} y we need to know the intentions of the speaker.

the Gricean responses. The potential of providing serious objections shows that semantic conventionalism is not “obviously” in a worse position than intention-based semantics.

Even though several individual objections are aimed at the first part of the reduction (timeless meaning \rightarrow utterer’s occasion meaning) and some against the second part (utterer’s occasion meaning \rightarrow speaker meaning) there is a recurring theme applicable to all. The important objections focus on the fact that, surprisingly, Gricean theory of meaning depends on conventions, and this in turn implies a circularity.

2.4.1 Timeless Meaning — Humpty Dumpty

If it is truly intentions that determine speaker meanings, which in turn determine the disjunctive account of timeless meaning, then we are left with a problem of how to constrain the items included in the disjunction. One must wonder *how many* uses of sentences (or words) can be included in the disjunction.

As we shall see with the ‘Humpty Dumpty’ problem, a speaker cannot take an utterance to mean *anything* they like on pain of ruling out such a thing as timeless meaning — in short, a disjunctive account would cease to be a possibility.

The famous ‘Humpty Dumpty’ problem is that Humpty Dumpty says to Alice “There’s glory for you.” To which Alice responds with “I don’t know what you mean.” “Of course you don’t” says Humpty Dumpty “because I have not yet told you what it means. I use “there’s glory for you” to stand for “there’s a nice knock-down argument.”” (Carroll 1994) But if Humpty Dumpty can do this, then his alternative way of expressing “there’s a nice knock-down argument,” i.e., “there’s glory for you,” should be added to the disjunction of timeless meaning.

But Humpty Dumpty’s speaker meaning of “there’s glory for you” cannot be included in the disjunctive timeless meaning. If it *were* possible to add Humpty Dumpty’s intended meaning to the disjunction, then any number of equally spurious occasion meanings would need to be added to timeless meaning.¹⁵

Moreover, Humpty Dumpty’s utterance (if added to the disjunction) would imply that even *possible* occasion meanings would need to be incorporated into the disjunction, not just *actual* occasion meanings. If this is so, disaster looms for Gricean semantics. There would not, in short, be any way of constraining timeless meaning (Searle 1969).¹⁶

Grice does have a response to this problem: the disjunction which makes up timeless meaning(s) relies not only on (1) what the speaker intends by the utterance, but also (2) what the audience expects the utterance to mean, and (3) the fact that there are *conventions* of use which determine whether utterances are correct or incorrect.

But, we immediately see a circularity, given (3). Intentions are meant to *explain* conventional meaning. However, the Humpty Dumpty case shows that we must be aware of the conventions of the linguistic community in order to communicate successfully. In this case, at least, it is necessary to grasp the conventions prior to the intentions in order to explain or understand the utterance.

¹⁵See for instance Searle’s objection from *Speech Acts* (Searle 1969) which we will turn to later.

¹⁶Note that it is not only possible occasion meanings that need to be considered, but possible *intentions*. There is an example later in the discussion (concerning the possibility of (unaccessible) intentions which a school teacher may or may not have) of what one *might* mean by an utterance. Which is to say, since meaning hinges on intentions and *possible* intentions become important to timeless meaning, then Gricean semantics becomes far more complex than it appears in the first case, and, as with incorporating possible occasion meanings, may lead to serious consequences for constraining disjunctive meaning.

To be fair, Grice’s position aims to be somewhat more sophisticated than this. In (Grice 1989b) he goes to great lengths in order to describe how one might constrain the expressions permissible in the disjunction. Consider, for instance, an incomplete utterance type (which is the term he uses to refer to words) such as “grass.” This can stand (roughly) for “lawn material” or “marijuana”.

The fact that incomplete utterance types can have multiple meanings is similarly the case for complete utterance types, that is, sentences. Consider: “I shall be pushing up the daisies when the sun goes supernova” is used to mean “I shall be dead when the sun goes supernova.” “Pushing up the daisies” means being dead. This is because language contains this particular idiom and it is *conventionally* taken to imply being dead.

Sentences uttered that, on the other hand, do not have an ‘usual’ meaning,¹⁷ such as “My clock will have stopped when the sun goes supernova,” do not mean “I shall be dead when the sun goes supernova,” since they are not usual expressions of English — for “my clock will have stopped” to mean “I shall be dead” depends, I think, not on an idiomatic reading of the words, but on pragmatics. That is “my clock will have stopped,” can be interpreted by the interlocutor as “I shall be pushing up the daisies.” But the interlocutor would have to make the interpretative leap. The hearer will be able to understand the speaker on the basis of conversational implicature, rather than idiomatic usage.¹⁸

The speaker who utters the latter may make himself understood to his

¹⁷I would say that we must even go so far as to avoid the ambiguous ‘usual’ tag, but can rightly assume ‘usual’ meanings to be conventional and are those which are sanctioned by the community.

¹⁸Remember that I distinguish between Grice’s semantic and pragmatic projects. If I am right, then the meaning of “I will be pushing up the daisies...” is relevant to the semantic theory and “My clock will have stopped...” is part of the pragmatic theory.

interlocutor by explaining what he means, or indeed by the context in which the speaker and interlocutor find themselves. But the latter utterance is a deviant case, as is the case of Humpty Dumpty’s expression and so suffers from the same problem.

The point in the 1969 article is this: both words and sentences have a usual/common usage. If they are used in a way that is not idiomatic then that use, or rather the occasion meaning, is not incorporated into the disjunction. That is why Humpty Dumpty cannot mean “there’s a nice knock-down argument,” by “there’s glory for you.”

But, even in light of Grice’s response the circularity is clear. How can he justify resorting to standard usage to account for intentions and then, in turn, use intentions to account for the common use? If he does so then intentions are not primary. Circularity, then, becomes not merely a possibility, but an *inevitability*.

2.4.2 Speaker Meaning — Post hoc

We shall now turn from speaker intentions to their application to timeless meaning, discussed above.

It is not unusual for an agent to be asked why he did, or said, some particular thing. What often springs to mind is not the intention *at the time of utterance*. Rather, the speaker considers what he may (must or most likely) have intended in order for those intentions to have conferred a particular meaning upon their utterance.

But this picture seems to imply that intentions are post hoc — therefore, it would be the case that the post hoc intentions are what determine meaningfulness. But one cannot have intentions being retroactively meaning conferring. If intentions are post hoc, then it cannot be consistent with Grice’s theory that intentions form the basis of the meaning of an utterance

(at the time of utterance), since the intentions simply were not *present* at the right time.

Consider: I may pat down my jacket, look confused, and utter “we have to go back to get my keys.” If questioned about my intention (at the time of utterance) I may answer “well, I suddenly realized that I left my keys in the office, and we have to go back and get them if we want to get into the apartment.” The conscious sequence of events from the patting down to utterance, however, may simply have been the conscious component “keys!” accompanied by a mental image of a bunch of keys lying on my desk. Or perhaps “keys” need not be a part of this picture at all and the mental image suffices for the utterance. The utterance would not, therefore, carry the intentions (at least at the time of utterance) that would allow it to fit into Grice’s analysis of meaning. (Dummett 1991)

An immediate reply would be that the intentions are not (usually) part of conscious thought processes, but a rapid and subconscious process from which we later form a rational reconstruction — the importance of intentions is that they account for what we probably *would have* intended had we reflected upon them at the time of utterance. The lack of the necessity of conscious reflection at the time of utterance calls into question the derivation of occasion meaning from intentions. If that step of the reduction fails, then it follows that the reduction from lexical meaning to intentions would fail too.

But if one allows intentions to be part of a *hypothesized* subconscious process, then the Gricean is left with the fact that intentions can be attributed to the utterance *after* the utterance is made, and so, as previously noted, are post hoc.¹⁹ If certain intentions are not present at the time of

¹⁹Grice admits (Grice 1957) that often intentions are unconscious and posited post hoc. Shortly we shall consider whether intentions may actually be present at the time of utterance, but that they are processed so quickly that no conscious reflection at the time

utterance then they could not have been that which is meaningful. They do not, in short, make the utterance mean what it does. To this there is no Gricean answer it seems. Intentions *must* be present at the time of utterance or else something other than intentions must be meaning conferring. And, thus, intentions fail to contribute to occasion meaning or timeless meaning.

To restate the point somewhat more succinctly: if there are no intentions (of the right sort) attached to individual utterances *at the time of utterance*, then at least in, e.g, the case of the lost keys, we have a counterexample to speaker intentions imparting meaning.²⁰

Further, there is no indication that cases such as the lost keys are isolated and rare. To the contrary, it is much easier to think of examples where intentions are post hoc than examples where intentions are clearly present. This further makes it impossible to characterize the disjunction of timeless meaning as a disjunction of the intentions of the linguistic agents.

But let us, for the time being, grant Grice that intentions may be post hoc, and that this does not negatively affect his theory. In 1957 Grice states that we perform a rational reconstruction of our intentions and attribute the most likely intentional candidates to the utterance. If this is so, then occasion meaning (derived from intentions) still constitute the basis on which linguistic meaning is founded. Grice states that:

Explicitly formulated linguistic (or quasilinguistic) intentions are no doubt comparatively rare. In their absence we would seem to rely on very much the same kinds of criteria as we do in the case of nonlinguistic intentions where there is a general usage. An utterer is held of utterance is necessary or possible. But, I shall argue that this means that either (1) intentions are confabulatory or (2) explicit intentions of speakers as representative of what probably motivated an utterance provides skewed data, since explicit intentions are so rare. What they might signal is not the *usual* intentions behind an utterance, but *unusual* intentions. This is because the recognition of implicit intentions are so rare.

²⁰Or at least of underlying an occasion of utterance and hence not making it into the disjunctive timeless meaning.

to intend to convey what is normally conveyed (or normally intended to be conveyed), and we require a good reason for accepting that a particular usage diverges from general usage... (Grice 1957, 222)

The problem is that there is no reason to believe that any particular intentions were present at the time of utterance, except on rare occasions. How, then, can they be meaningful, or convey the timeless meaning of lexical items? How can one build a theory of meaning when the intentions which are supposed to form the basis of the theory are missing? To these questions there seem to be no answers for the intentionalist apart from, once again, appealing to conventional usage.

2.4.3 Speaker Meaning — Determining Intentions/Confabulation

The problem here is two-part in nature: (1) the determination of intentions on occasions of utterance may (as we have seen) not be determinate, and so (2) the speakers and their interlocutor may not be able to grasp the intentions. If intentions are indeterminate then the timeless meaning of an utterance equally becomes unclear.

Consider in contrast cases where intentions are determinate.²¹ Even in these cases²² the data may be skewed. The problem is this: if explicit intentions are so rare, then it is not clear how we can rely on them as representative of internal (implicit) ones. Determinate intentions may in fact be the exceptional cases, which are not representative of the majority of cases where intentions are indeterminate, even to the speaker.

Since we cannot be sure that cases where intentions are explicit are mir-

²¹An example could be: “I would like a coffee.” Unless the context indicates otherwise, it can be assumed that the intention of the utterer is merely to convey that they want a coffee.

²²That is, that both the speaker and hearer can grasp the intention(s) behind the utterance.

rored by cases where they are implicit, it may be the case that the explicitly determinate intentions are *abnormal*, and skew the data: the cases where intentions are explicit may differ radically in nature from those that are implicit. We simply do not and cannot draw an analogy between the two as there is no way to acquire empirical data.

The implication is that when the intentions of an utterer are explicit they may not have the same characteristics of implicit intentions. It follows that explicit intentions may *confound* the conclusions that one might want to draw on the basis thereof in regards to the status of intention-based semantics. Explicit intentions may be different enough in nature from internal ones that no connection can be made.

2.4.4 Occasion meaning — Fine Grained

The next problem to be considered involves giving two counterexamples to the part of the Gricean analysis which concerns occasion meanings. These counterexamples have been dubbed the problem of “grainedness” (Dummett 1993, 299-300). The counterexamples will show that, in some contexts, occasion meanings are too fine-grained. The next section will deal with occasion meanings being too coarse-grained.

In regards to fine-grainedness, the framework for an example is provided by Grice’s account of occasion meaning in ‘Utterer’s Meaning and Intentions’ (Grice 1989b):

“*U* meant something by uttering *x*” is true if and only if, for some audience *A*, *U* uttered *x* intending:

- (1) *A* to produce a particular response *r*
- (2) *A* to think (recognize) that *U* intends (1)
- (3) *A* to fulfill (1) on the basis [or at least in part] of his fulfillment of (2).

(Grice 1989b, 151)

Consider the example, briefly discussed in Chapter One: an American soldier, call him Ted, is captured by Italians during the Second World War. He wants the Italians to believe that he is a German officer and so utters the only German phrase that he knows: “Kennst du das Land wo die Zitronen blühen?” (Lycaen 2000, 107)²³

Now, Ted utters “Kennst du das Land wo die Zitronen blühen?” intending:

- (1) The Italian soldiers to produce a belief that Ted is German.
- (2) The Italian soldiers to recognize that Ted intends them to form this belief.
- (3) The Italian soldiers to come to believe that Ted is a German on the basis of [or at least in part because of] the recognition of Ted’s intention for them to form this belief.

Thus, it seems that Ted’s utterance, due to the occasion meaning of the utterance, means “I am a German soldier.” But, this is too *fine*-grained. Ted’s utterance cannot mean “I am a German soldier” anymore than it may convey to his interlocutors that he thinks that he is a pink giraffe, a unicorn, and invisible elephant, or that his officious tone of voice means “Release me now you incompetents” and so forth, since his interlocutors also do not understand German. His interlocutors simply cannot pick out what Ted means — his utterance is too fine-grained for it to mean what Ted wants it to.²⁴

²³This is a well known example. It originates in Searle’s book *Speech Acts* (Searle 1969).

²⁴The analysis of meaning offered by Grice expands between 1957 and 1969. The later amendments, however, essentially result in the same analysis: that it is by grasping the speaker’s intentions on the occasions of utterance that the speaker’s interlocutor grasps

One important addition to later Gricean analyses of meaning is the idea of ‘modes of correlation’. As far as I can tell, these modes of correlation relate to conventional meaning. The possibility of intending x relies to a degree on the ‘conventional’ usage of x . Grice is explicit that Ted can justifiably mean “I am a German officer” by “Kennst du das Land wo die Zitronen blühen” (Grice 1989b) — and so “I am a German officer” *should* be one of the usual usages of the sentence “Kennst du das Land wo die Zitronen blühen.” This is a conclusion that most Griceans reject, since the intentional meaning (that which Ted is *trying* to convey) would deviate too much from the standard linguistic meaning of the utterance.

Although most Griceans refuse to accept that “Kennst du das Land wo die Zitronen blühen?” means “I am a German officer,” Grice himself does not think that it is so ridiculous. As Grice says concerning this exact situation:

[T]o point out that the German line means not “I am a German officer” but “Knowest thou the land where. . .” is not relevant. If the American could be said to have meant that he was a German officer, he would have meant that by saying the line, or by saying the line in a particular way. . . (Grice 1989b, 162)

It is possible, and quite right I think, to view Ted’s utterance as another Humpty-Dumptyism. Part of the earlier discussion concerned Grice’s own attempt at curtailing the possible meanings conveyed by the intentions of the utterers by appealing to the usual usage of sentences. This seems consistent with rejecting what he says in the above quotation. If this is correct, and in addition, if most Griceans would not be willing to admit that Ted (or Humpty Dumpty) can use any words or sentences to mean anything they the meaning of the utterance. Features added to Grice’s analysis of meaning between 1957 and 1969 are, e.g., the role of propositional attitudes of utterers and the grasping thereof by their interlocutors, etc. But, the expanded elements of the analysis still rely on the utterer’s intention, which create appropriate thoughts in the audience so that they can grasp the intentions.

wish, then there must be some way to limit which speaker meanings can count as part of the disjunctive timeless meaning.

One way to draw a boundary to speaker meaning is, as mentioned, to require that the utterance is used in such a way as to conform to the conventional usage of it. But, again, we are left with the Humpty Dumpty problem of circularity. In the case of timeless meaning, it cannot be that anyone can mean absolutely anything by their utterance. This is especially clear when Grice himself appeals to the ‘usual’ meanings of words and sentences.

2.4.5 Occasion meaning — Coarse Grained

In addition to intentions being too fine-grained to constitute speaker meaning, they can also be too coarse-grained. Take the following example: a parent says to her young child “It’s ten thirty!” The broad intention is to get the child to go to sleep. The parent knows that the child knows that ten thirty is considerably past bedtime and that she must go to bed. Of course, the child, understanding the intentions, will most likely draw the correct inference that going to bed would be in her best interest.

In this case, though, we appear to get a multiplicity of intentions. Initial intentions which influence the meaning of the utterance would appear to be: (1) the parent intends for the child to go to bed; (2) the parent further intends for the child to understand this intention; and (3) the parent intends for her to go to bed because of understanding the parent’s intention. However, the literal meaning of the utterance merely reflects the time, not “Go to bed!” I admit that it does seem that, in combination (1)-(3) fit the Gricean analysis given above and so “It’s ten thirty” *would (questionably) mean* “Go to bed!”. But, the multiplicity does not cease here.

The parent may have (1’) work to get done (and so intends for the child

to from the belief that the mother has work to do);²⁵ (2') the parent further intends for the child to understand that this is her intention; and (3') the parent intends for her to go to bed because of understanding the parent's intention.

Another motivation might be that the parent has (1'') a need for some relaxation time, and wants the child to form a belief to this effect; (2'') the parent further intends for the child to understand that this is her intention; and (3'') the parent intends for her form the belief because of understanding the parent's intention.

Or (1''') a babysitter might be arriving who will expect the child to be asleep when he turns up and so the parent intends the child to think that the parent intends her to go to bed because of the sitter's looming arrival; (2''') the parent further intends for the child to understand that this is her intention; and (3''') the parent intends for her to go to bed because of understanding the parent's intention.

What then is the speaker meaning?

(1') to (3'') may (again, arguably) give the expression "It's ten thirty," the meaning "Go to bed." But all the intentions from (1') to (3'') may be in place at once, so the meaning of "It's 10:30" on a single occasion means all these things (and more): "It's 10:30 and I have work to do and the sitter is coming and I need some time alone." The meanings are all bundled together and undifferentiated. So there is no *one* speaker meaning.

Having multiple speaker intentions, in addition to them being unknown by the audience, fails the Gricean analysis of intention-based meaning. If

²⁵Notice that sometimes having intentions being determinate is *undesirable*. It may be the case, for instance, that the child is not sympathetic to her parent's need to work and sees that as taking attention away from her. Grasping the belief that the mother has work to do may have the opposite effect than desired and make it more difficult to get the child to bed.

a multiplicity of intentions are present, then, given the two part reduction, there are multiple occasion meanings. The single utterance would not have *a* meaning at the time of utterance, but multiple meanings at the same time. The question that arises is how the audience would be able to determine the meaning of the utterance — does it mean that “the parent is busy,” “the parent needs some time off,” etc? If there are intentions x, y, z (at the same time, then the audience cannot figure out *the* meaning of the utterance (not the timeless meaning, but the occasion meaning).²⁶ This is because there is more than one intention present at the same time and so there is no *one* occasion meaning (or at a minimum, not one that can be identified).

Apart from failing to conform to the Gricean analysis, it is not clear that multiplicities of intention are of any explanatory use. In this case, if all goes well, the child will comprehend the parent’s intention through contextual information (and so would actually be part of a pragmatic theory rather than semantics) that they should hurry off to bed, but with insufficient knowledge of the intentions of speaker to gather the multiplicity of meanings of the utterance. In this case, being unable to grasp the distinction between (1)-(3) and (1’)-(3’’), the meaning of the utterance surpasses the comprehension of the speaker’s interlocutor.

Grice does have a response to the possibility of having a bundle of intentions: in cases where there are multiple intentions *one* is (or should be) the primary intention which determines the occasion meaning. But this cannot be so. If speakers have multiple intentions on the occasion of an utterance then it follows that there is more than one communicative intention. But,

²⁶There may, of course, be exceptions to the rule. Consider, for instance puns and double entendres. It is precisely the multiplicity of meanings that make these work. But they certainly do not represent standard usage. Further, one might argue that it is pragmatic interpretation, not semantics, that determines whether an interlocutor can interpret the multiplicity of these cases.

consider again the analysis which was given in the example of Ted uttering a German phrase to an audience intending them to take it to mean “I am a German soldier.” What would happen on an occasion with multiple intentions? Let us run the analysis again to check: A mother utters “It’s ten thirty” intending:

- (1) her daughter to produce a belief that the mother has work to do and that the mother needs some relaxation time and that a babysitter is arriving.
- (2) her daughter to recognize that the mother intends the daughter to produce a belief that the mother has work to do and that the mother needs some relaxation time and that a babysitter is arriving.
- (3) her daughter comes to believe that she should go to bed on the basis of [or at least in part because of] the recognition of the mother’s intention for the daughter to produce a belief that the mother has work to do and that the mother needs some relaxation time and that a babysitter is arriving.

At this point the Gricean analysis is not clear and seems to go against the example where it is an *intention* not *intentions* that a hearer must grasp to understand the speaker meaning. The only way for the analysis to work is by saying that there was not *one* speaker meaning on the occasion of utterance. It cannot be said, then, that there is a speaker meaning at all, but that there are multiple speaker meanings on the occasion of utterance. What this seems to result in is not a single occasion meaning on the basis of the Gricean analysis, but multiple meanings. Consider the analysis again: “*U* meant something by uttering *x*” is true if and only if, for some audience *A*, *U* uttered *x* intending:

- (1) *A* to produce a particular response *r*

- (2) A to think (recognize) that U intends (1)
- (3) A to fulfill (1) on the basis [or at least in part] of his fulfillment of (2).
(Grice 1989b, 151)

If the above example does not convince the reader consider the following example. Bob utters “nice hat” to Bill as a colleague walks by. Intending to sound as if he were being sarcastic and that thinks the hat is ghastly, the analysis should be:

Bob utters “nice hat” intending:

- (1) Bill to produce a belief that the hat is ghastly.
- (2) Bill to recognize that Bob intends [for] Bill to form this belief.
- (3) Bill comes to believe that the hat is ghastly on the basis of [or at least in part because] the recognition of Bob’s intention for them to form this belief.

But, suppose that Bob’s sarcasm hides a desire to cover up the fact that he has romantic feelings for the hat-wearer. Whilst uttering “nice hat” he intends for Bill to *think* that he is being sarcastic, but really he thinks the hat looks rather dashing on the object of his desire. The analysis now goes as follows:

Bob utters “nice hat” intending:

- (1) Bill to produce a belief that the hat is dashing.
- (2) Bill to recognize that Bob intends him to form this belief.
- (3) Bill comes to believe that the hat is dashing on the basis of [or at least in part because] the recognition of Ted’s intention for them to form this belief.

But, Bob patently *does not* want Bill to think that the hat is dashing. To make this more explicit, let us suppose that Bill recognizes that sarcasm is often taken to be a hidden form of flirting or an indication of romantic feelings: this recognition makes it *even* clearer that Bob not only has multiple intentions and thus multiple speaker meanings on occasions of use, but that they must have different analyses — and so different meanings.

However, intention-based semantics does not seem consistent with having concurrent multiplicity of meanings. Grice's stance is that, even if the utterer has a multiplicity of intentions, one of them is the primary intention. And it is the primary intention that must be grasped by the audience in order for (1) the utterance to be meaningful and (2) the audience to grasp the meaning of the utterance. But this supposes that the primary intention is that which makes the utterance meaningful, not that the utterance can have multiple meanings at the same time. In the case of Bob especially, there does not seem to be one intention which can be picked as the primary one in a non ad-hoc fashion.

2.5 Conclusion

In this chapter several important objectives have been achieved.

First, we must remember, as was discussed, that Grice's theory was historically very important in rejecting the third realm and behaviourism (but so too does conventionalism) and that the other motivations are important to take into consideration. I have also explained why intention-based semantics seems to square well with the intuitions of many people.

Second, intention-based semantics is important for contemporary philosophy of language. Many projects focus on Gricean (or neo-Gricean) theories of meaning²⁷

²⁷I myself have an active interest in such projects, for instance it is interesting to inves-

Indeed, I think that Grice's theory is a strong one and must not be rejected from the canon. However, the point of this chapter was to show that conventionalism is overlooked in favour of alleged relative strengths of intention-based semantics. And, this has been achieved. By using this one exemplar — and a particularly strong one at that — I have shown that there are objections which can be raised against it. Whether or not they are insurmountable has not been considered in too much detail. The main purpose of considering the objections was to show that this strong theory is not an indication that conventionalism must be rejected out of hand.

Moreover, as we repeatedly saw in considering responses to objections, there is a seeming need to appeal to *conventional* usage to salvage the intentionalist's intuitions about "literal meaning." While I do not want to make too much out of it here, this does at least hint that it is worth considering a semantic theory that builds directly on these conventions — especially since, as noted, conventionalism shares the anti-platonist and anti-behaviourist virtues of intention-based semantics.

tigate what intention-based semantics can tell us about what is said, as opposed to what is implied (remember the link between Grice's theory of communicative practice and his semantic theory). And intention-based semantics has also been applied to trying to come to understand language deficits, which is most definitely worth pursuing.

Chapter 3

Luminosity

3.1 Introduction

In this chapter I consider the first supposed knock-down argument to conventionalism noted in the introduction. Here, I consider Williamson's charge that conventionalism implies manifestability and hence luminosity. The basic structure of this objection is that:

- (1) If conventionalism is correct, then manifestability follows and if manifestability is accepted, then there are non-trivial luminous mental states.
- (2) But, there are no non-trivial luminous mental states.
- (3) So, conventionalism is false.

I agree that luminosity must be rejected, whilst manifestability (of the sort I have in mind) must be accepted.

However, Williamson thinks that any semantic conventionalist is committed to luminosity. To show why conventionalism is wrong, he constructs what is known as the 'anti-luminosity' argument — so he does provide what is, *prima facie*, a knock-down argument. In agreeing with Williamson's anti-

luminosity argument, but retaining manifestability, I show that his argument is not knock-down — in effect by denying the second conjunct of (1).

So, my overall goal is to show that while luminosity is inconsistent with a strong semantic conventionalism, manifestability must be incorporated in semantic conventionalism.

To say that a mental state is *luminous* is to say that, whenever an agent is in it, the agent is *in a position to know* that he is in it. Consider specifically linguistic knowledge, luminosity (according to Williamson) translates to the *transparency* of meaning — if an agent grasps the meaning of a word, then reflection should enable the agent to know in detail all aspects of the meaning of the word.

Manifestability, as I present it, however, explicitly rejects the necessity of such knowledge being available to an agent. It needn't be the case that an agent knows all the aspects of the meaning of a lexical item of a word. If an agent can *use* the lexical item correctly, then he has sufficient, but not necessarily exhaustive, knowledge of the meaning of the term.

Since language is externalized, I argue that manifestability means that by appealing to (some subset) of the community we can gain knowledge of lexical items. So, this is in effect internalism versus externalism. The latter I accept, the former I reject.

However, any luminous mental states reduce to triviality as we shall see when discussing the anti-luminosity argument.¹ The 'anti-luminosity argu-

¹This is assuming that luminosity implies that knowledge of the meanings of terms would be a type of mental state — I do not take up the discussion as to whether or not knowledge is internal or external as it would take us too far afield, but note that it is not a settled matter as to which holds. It is sufficient here to note that Williamson's argument against luminosity is part of a broader project — that of rejecting the KK principle. That is to say, he rejects the idea that if one knows (that they are in a condition) x , then one need also know that one knows that (they are in condition) x . This broader project need not concern us, however, since we are only concerned with how it affects

ment' is presented in Williamson's *Knowledge and Its Limits*. (Williamson 2000) The exposition of which will follow shortly.

The sections of the chapter are:

- 1 Specifying what luminosity *is*.
- 2 Considering why it seems plausible that a conventionalist *should* be committed to luminosity.
- 3 An exposition of Williamson's 'anti-luminosity argument'. Using this argument, he shows that there are nontrivial luminous states — that is, at no (non arbitrary) place in the transition between a condition C and a condition $\neg C$ can the agent come to know that they have failed to be in C and are thereafter in $\neg C$.
- 4 My response to the objection.

Essentially my response, if one wants to call it that, is to *accept* Williamson's argument that luminosity is an undesirable feature of conventionalism. But, the knock-down objection fails in light of the fact that a conventionalist need not, and cannot, be committed to it — luminosity must be rejected.

I will show that the rejection of luminosity actually helps to clarify the conditions of a community-based semantic theory. Giving up luminosity means that the mental states of the linguistic agents fall out of the semantic theory, and quite rightly so. The way that I draw out an expanded picture of anti-luminous conventionalism is by providing a series of examples where a linguistic agent attaches particular meanings to words, but in such a way that they depend on the community in order to tell if they understand the meanings of the terms and under what conditions sentences containing those terms are warrantably assertible.

the conventionalist semantics. The broader project concerns many metaphysical questions which are not pertinent to this discussion.

- (5) So, after providing a series of examples, it will become apparent that certain criteria should be incorporated into a semantic conventionalist theory.

The conditions which come to light and which will be incorporated into semantic conventionalism are: manifestability, externality of language, and the fact that it is the community which determines ‘the rules of the game’, not the individual agents. Meaning is still ‘use-based’, but appropriate use is determined by the community — a linguistic agent may actually be blind, e.g., to the appropriateness of an assertion, but can in principle be put right by being exposed to appropriate conventions. It follows, as I will show, that there must be some sort of division of linguistic labour incorporated into the theory.

3.2 Luminosity

Luminosity is, *prima facie*, a likely condition upon linguistic agents in that, if an agent attaches meanings to lexical items, then he needs to know (upon sufficient reflection, as mentioned), whether the meanings are the same or if they diverge.²

There is, unfortunately, only a small literature concerning luminosity, especially concerning semantic theories, so my exposition will be a bit rough and ready — but I think that Williamson’s is also. This chapter, however, provides sufficient information to deal both with the objection *and* make it clear why I respond in the way that I do.

Before we can consider why a conventionalist might feasibly be committed to luminous mental states in regards to lexical items and evaluate Williamson’s argument, a bit needs to be said about what he means when he calls a mental state ‘luminous’.

²How would he, e.g., know when he is able to make correct assertions?

A luminous mental state is one where, if an agent is in it, he is always able to tell that he is experiencing that mental state (or at least, he can come to know that they are experiencing it). Consider, for instance, pain, which is often taken to be luminous (how easy is it, e.g., to avoid noticing a migraine?).

To introduce a further piece of Williamson's jargon, a mental state is *non-trivial* if an agent can be in it or not. Again, pain is a useful example. The usual assumption is that an agent can always know whether they are in condition C or $\neg C$ (migraine or no) — not that the agent always *does* know, since he might, e.g., not be paying attention, but if an agent *did* pay attention he would be able to tell. This, however, turns out not to be the case, as proven by the anti-luminosity argument, as we shall see in a moment.

3.3 Plausibility

Why, then, does it seem plausible that a *conventionalist* must hold that there are luminous mental states in regards to knowledge of meanings, as Williamson suggests a conventionalist is?

The suggestion is that *knowing a meaning* of a lexical items is a luminous state.³ To be a competent linguistic agent the agent must 'master the rules of the language game'. One might say that for every term employed in a sentence, the agent must know the meaning of the term if he is warranted in asserting it.

It does seem plausible, as mentioned, that given the maxim that agents must know the rules of the language game that they must individually know the assertability conditions of sentences, and hence the meanings of the terms

³Note, though, that this is granting that knowledge *is* a mental state — Williamson is an externalist about knowledge, so it should come as no surprise that he finds luminosity so implausible.

employed. But, if this is the case, then the speaker must equally know when the meaning of two terms is the same or divergent. For instance, if an agent were to assert a sentence such as “The pavement is wet, it must have rained,” or “The sidewalk is wet, it must have rained,” then to know that either can be asserted under the same conditions, he must understand that the terms are synonymous — since he must know that pavement is interchangeable with sidewalk (if one is familiar with both British and Canadian English, say). Alternatively, if lexical items with divergent meanings are used in a sentence, then the agent must also know that the assertability conditions change. Knowledge of assertability conditions and meanings then does seem to be required of the individual agents.

We turn now to Williamson’s anti-luminosity argument, in part to show that he is right that knowledge of lexical items cannot be luminous. But I will then show that his argument does not ring the death knell for semantic conventionalism.

3.4 Williamson’s Anti-Luminosity Argument

In order for conventionalism to work, the argument goes, access to the ‘rules of the language game’ must imply luminosity of the meanings of lexical items for *individual speakers*. But, according to Williamson, language users do not have such access to the language. In order to show this, he constructs what I shall call the ‘anti-luminosity’ argument (Williamson 2000).

In regards to the charge of luminosity, bear in mind Williamson only concerns himself with one conventionalist, namely Dummett, not all conventionalists. Dummett *does* seem to be committed to luminosity. There is no reason to suppose that other conventionalists agree to luminosity — indeed, I do not, and shall explain why in a moment.

Therefore, I shall not concern myself with (what Williamson perceives

to be) Dummettian luminosity. Perhaps the reason that Williamson takes Dummett as, what we might call, the ‘luminosity role-model’ is that, given his perceived dominant position in regards to conventionalism, all other conventionalists fall into line. This, I will show, is unwarranted.

The basic structure of Williamson’s argument is fairly straightforward: he *assumes* that there is a *non-trivial* luminous state — a state C , which a person can be in or fail to be in, but not both at once. He reduces the assumption to absurdity by showing that for any such condition it is possible for a person both to be in it (C) and fail to be in it ($\neg C$).

Before sketching the argument in slightly more detail, we need to note one further epistemological claim Williamson relies on. While not committing himself to *reliabilism*, he does insist that knowledge of P implies a sort of *reliability* for the belief that P . In particular, if an agent *knows* P under certain conditions, then P must be *true* under *closely related* conditions (though in some of these related conditions P might not be known). He sometimes refers to this as a requirement for *margins of error*, something that gets encoded in principle \mathcal{M} below.

The argument against luminosity can be paraphrased as:

- ★ First we assume that there is a non-trivial condition C to be luminous.
- ★ Second, since an agent can both be in C and fail to be in $\neg C$, we assume that there is a transition for the agent between condition C and $\neg C$. The stages of the transition are s_1, \dots, s_n , with C holding in s_1 and $\neg C$ holding in s_n . Since C is luminous, the condition \mathcal{L} applies.
- ★ \mathcal{L} : If an agent is in condition C , then he can know that he is in C .
- ★ \mathcal{M} : If KC at s_i , then an agent is in C at s_{i+1} .

- Since C (at s_1), then KC at s_n by applying \mathcal{L} — which means that an agent at s_1 knows that they are in C .
- Apply \mathcal{M} — If KC at s_1 , then the agent is in C at s_2 .
- By repeated application, for every s_i , if you are in condition C , then KC at s_{i+1} , and so in C at s_{i+1} .
- Strictly speaking what Williamson shows is that any luminous mental state has to be trivial, where by trivial he means that being in C and $\neg C$ at the same time is possible — the sorites style argument shows why luminosity results (in this type of) triviality. A non-trivial mental state would mean that an agent never finds themselves in conditions C and $\neg C$ at the same time — there are unquestionably non-trivial mental states, but a condition of such states is that they are non-luminous.

Rather than *responding* to Williamson’s argument, I grant that luminosity results in triviality. So, any semantic theory that requires the existence of non-trivial luminous states is in trouble. This, given what I said above, means that conventionalism can plausibly be thought to be in trouble.

Semantic theories aside, for the moment, note that, if Williamson is right, then we cannot even help ourselves to the idea that no physical or mental states are states that, when we are in them, we always have knowledge of being in them,⁴ nor even that when we are in them we *could* know it. Consider, for instance, having a headache — there may be more pressing things to do than lying down with a cold compress, such as making dinner for the family. Being engaged in such an activity may very well lead someone to cease to be aware that they are in pain — other considerations, in

⁴This means that anti-luminosity implies just what the argument above showed: that we cannot always have knowledge of the conditions we find ourselves in.

other words, override the knowledge that one is in pain. This does not, yet, however, show that pain is non-luminous — you could still be in a position to know (even whilst making dinner) that you have a headache. To show that pain is non-luminous, then, we cannot merely appeal to distraction by alternative activities. Having a headache is plausibly a non-trivial state. Williamson’s argument, if sound, implies that there are possible cases where an agent has a headache, but *is not in a position to know* that this is so. The anti-luminosity argument is therefore one with important and controversial consequences, and so deserves careful attention from philosophers. I set these issues aside here because they are beside my present point, not because I think his argument settles the matter.

To sum up: whilst the conventionalist is *seemingly* committed to luminosity, it must be rejected. Since I reject the ‘undesirable’ feature myself, luminosity does fail in the sense that there can be gaps in the knowledge of individual agents in regards to the knowledge of the meanings of lexical items. This shows that the objection can be side-stepped. It thus fails to be an objection at all — least of all knock-down. This is one way in which I am, at most, a neo-Dummettian — luminosity, something Dummett takes to be a criterion of semantic conventionalism, does not make it onto the ‘list’ of what I take to be the defining features of conventionalism.

However, I do insist that meaning is *public* and *manifestable*. This means that anything which deserves to be called conventionalism no aspect of meaning is “hidden.” This suggests “knowability in principle” for all aspects of the meaning of any particular term. What I suggest is that this needn’t be “knowability” for the particular agent, but knowability for the community.

But what, exactly is meant by the community *knowing* the meanings of lexical terms?⁵ Without an answer to this our argument is incomplete. A

⁵As we will see in a later section (and in Chapter Five) *knowing* in relation to a community may, at first glance, appear to be odd. However, given the fact that I advocate division

community is a conglomerate of individuals but neither the understanding of lexical items nor assertability conditions are fixed by the individuals. Knowledge is distributed across the community. First we must consider what the community looks like in general before moving on to specifics.

Conventionalism, then, is not threatened by Williamson’s argument, given the manifestability of meaning. As long as knowledge of the meanings of lexical items are (externalized and) manifestable by the community, then the condition is merely that an agent not in possession of the knowledge of the meaning of a term *could* come to possess the knowledge, say, by asking the right person.

3.5 Examples

What the examples to come show is that the meanings of lexical items need not be transparent. In externalizing accessibility of the meanings of lexical items I show that the knowledge of individual agents is not primary, the “knowledge” of the community is. The externality is based on the notion of manifestability.

Returning to the sketch of the supposed “knock-down” argument against conventionalism with which we began, I will show that the conjunction ((conventionalism \rightarrow manifestability) \wedge (manifestability \rightarrow luminosity)) fails, since the latter conjunct is false.

These examples show that the community is central in determining how one comes to know the meaning of lexical items. All three involve manifestability in one form or another (but not luminosity). After the examples, further discussion will highlight several criteria of semantic conventionalism. Notice that when I talk about the ‘community’, I do not simply mean the of linguistic labour (spelled out in terms of manifestability in due course), and develop a type of states of information semantics (in Chapter Five), this oddness disappears.

(individual) linguistic agents of the community, but other resources, e.g., books, the internet, observation on the part of the agent, subsets of the linguistic community (e.g., groups of experts), etc.

- (1) The first example concerns what seems to be one word of which I possess knowledge of the meaning. It just so happens that there is a divergence of meaning when the word is applied under different circumstances. So, I know the meanings of two terms and I count them as two lexical items because they carry distinct meanings — w_1 and w_2 and I further know that they do not have the same meaning.

Consider the robin, which is a different species in North America, Europe, and Australia. I did not know that there was a different species of robin in Australia — my ornithological knowledge is distinctly lacking. I stumbled upon the Australian distinction by trying to determine the (precise) difference between the North American definition and the European definition of robins. I do not, or at least did not, know what the distinguishing features are of the Australian robin, but I can either ask an ornithologist (who, by being a member of the ornithological community can be taken to be an authority on the matter), by consulting a book, or even (heaven forbid) Wikipedia. All three would be able to provide me with sufficient information concerning different species globally. So, I can come to know that $w_1 \neq w_2$ and that neither is identical to w_3 .

But, in the case I have in mind, I was a *competent* user of the word before discovering this further distinction. I was “in a position to know” that there are distinct meanings, but not merely by reflection.

- (2) The second example is one in which I, or some other linguistic agent, have deficient knowledge of two objects to be able to tell them apart. However, they *should* be identified by different words — but I do not

know this *until after* an authority has been consulted. The linguistic agent, in this case, is incompetent.

Consider the terms w_1 , take this to be ‘dalmation toadflax’, and w_2 , take this to be ‘yellow toadflax’. I cannot tell the difference between the two, not being a specialist in regards to flowers. Assume, looking at some flowers that I take w_1 to be identical with w_2 (given my lack of knowledge in regards to flowers). In fact they are different species of the snapdragon family. The only distinguishing feature, apparently, is that they have different types of leaves. I had no idea about these flowers at all until I consulted the internet. However, my mother and grandmother, having a large number of books on botany would be able to look the flowers up in the books.

There would be two types of authority in this example. The first is that it can be pointed out by my mother and grandmother that dalmation and yellow toadflax are the different species (or they may witness incorrect usages of the term(s) and correct me). They might not know *immediately* the difference between dalmation and yellow toadflax, but they do know that they are (somehow) different and so they can look up the facts. I am no authority at all; my mother and grandmother could be said to be partial authorities, in the sense that they know that there *is* a difference and they know how to acquire the requisite knowledge; the second authority, which ultimately determines between dalmation and yellow toadflax, then, are the books which are consulted.

In this situation, I am not a competent linguistic speaker at all, but can appeal to the other authorities to update my linguistic competence — so that I can grasp the meanings of the terms involved.

There is thus, as in the robin case, information available to me which would allow me to distinguish w_1 and w_2 . But this requires externality of language — it must be *manifestable in principle*, i.e., it must be manifestable

that the words carry different meanings and apply to distinct objects.

Notice, though, that this example is somewhat different from the robin example — a limited knowledge of the distinction between species of robin was in my stock of lexical items. Although I did not know about the Australian robin, I knew the difference between the North American and European species sufficiently to know under what circumstances the words could be used. So, in this case, I was a competent linguistic agent in regards to two types of robin, but not at all about the two different species of flower. In the second example, I am an incompetent linguistic agent (since I did not even know that dalmation and yellow toadflax existed). My grandmother, on the other hand, is in a position similar to mine with respect to the two sorts of robins.

- Note that examples (1) and (2) are very similar in nature. One can extrapolate from them and come up with any number of alternative examples. Although there is some divergence between (1) and (2), we see a greater difference when we consider (3).

(3) The third example is one in which I have an *approximate* knowledge of a lexical item.

Suppose that someone has been raised by a geophysicist and, on various family holidays, attendance of public lectures, and so forth, gathers some information concerning what a geophysicist is and does. Suppose it is me.

There are two ways in which I have some knowledge of what the lexical item means. As in philosophy, there are distinct disciplines in geophysics. Some geophysicists pursue fields that are very close in nature to geological ones. Others concentrate on, e.g., four-dimensional seismic surveys. Can I be said to be a competent linguistic agent in this case? Most likely not *entirely*. However, by consulting textbooks, geologists themselves, etc., I

can come to know what the general definition of ‘geophysicist’ is (call this w_1). But, moreover, I can come to know that a geophysicist in discipline₁ and discipline₂ are actually different types of scientists, in the same way that there are philosophers in different fields (which may be as opaque to an outsider as someone outside the field of geophysics).

The differences in meanings of the lexical items in question are *manifestable* in all three examples, because in all three, one can appeal to outside sources to gain knowledge of the meanings of lexical items. Internal reflection is obviously not sufficient in any of the cases, so manifestability does not imply luminosity (thus showing the falsehood of the conjunction which is used in the construction of the ‘knock-down’ argument — the second conjunct is shown to be false).

By discussing the conditions of conventionalism which result from the examples, it will become apparent not just that I do not *think* that knowledge of lexical items is external, but that the meanings *must be* external.

3.6 Conditions on the Linguistic Community

How, exactly, then do the examples relate to the linguistic community and the conditions of semantic conventionalism? Of course providing examples is not the only way to draw out these conditions, but they are useful.

The examples above show that ‘mastering the rules of the game’ and accepting that ‘meaning is use’ can taken to be very broad notions — on the basis of the community, and the ‘authorities’ that it contains, we can tell whether an assertion is warranted or not (once the conventions are understood). The only thing an agent need know is under what circumstances an assertion is deemed to be warranted or not. And on the basis of the rules of the community, the linguistic agent can either stick to their assertion or must take it back.

So, a competent linguistic agent need not (immediately) possess the knowledge of the meaning of lexical items, but can make assertions on the basis of ‘approximate’ knowledge.

What, then, have we learned from the examples?

It is reasonable to draw the following lessons from the examples:

- (1) the community determines what the rules of the game are and whether they are being followed correctly.
- (2) meaning must be externalized.
- (3) meaning is manifestable.
- (4) there are different types of authority
- (5) these conditions make it clear that there must be some sort of division of linguistic labour to make semantic conventionalism acceptable.

(1)-(5) are importantly related to each other, as shown in the examples. The examples provided show that one cannot achieve a grasp of the conventions of a language unless the rules of the game are employed. If a rule is flouted and an unwarranted assertion is made, then an agent can (or at least should be able to) be put right. That is, the authorities — members of (or objects of) the linguistic community — determine the rules (conventions) for using lexical items. Thus, knowing the conditions under which an assertion could be correctly made is precisely understanding and following the externalized rules. The rules then require abiding by conventions. Otherwise, what the ‘language game’ consists in is unclear, if one could make sense of what this means in the first place.

In order for a conventionalist to survive Williamson’s attack, he must revise his understanding of what “grasping” the rules of the game or the meanings of lexical items actually implies. The rules of the game are not

determined by the individual, but by the community. We've known that all along — this is simply part of what it is to be a conventionalist. So, in order for an agent to be a competent speaker, the rules of the game do not need to be entirely *transparent* to him, as the examples have shown. As the examples have further shown, grasping the meaning of lexical items or the rules is displayed by the agent's basic *competence* with the rules — even if, as noted, he does not know *all* aspects of the rules.

Having accepted Williamson's anti-luminosity argument, I have accepted that having knowledge of the meanings of lexical items must mean that an agent is in a non-luminous state. I have also accepted that manifestability implies "being in a position to know" all aspects of the meaning of a word in some sense. But since grasp of meaning must be non-luminous, something other than internal reflection must provide access to knowledge of the meanings of lexical items, this, as mentioned previously relies on the acceptance of conventions. Meaning is constituted by these conventions, which are plainly external and community-based — otherwise no agent would be competent, no preexisting meanings would be available to grant the agent access to the language, and no authorities could exist. The examples show that meaning must be not only manifestable, but also intersubjective, and that there must be a division of linguistic labour.

Knowing the difference between the European and North American robin, shows that I am a competent linguistic agent, since I know what assertions are warranted on the basis of the knowledge that I have. *But* in order to gain knowledge of the Australian species I would have to consult an appropriate authority or authorities — be they sublinguistic groups, books on a particular subject, internet sites, or whatnot — the meaning of the terms must, therefore, be manifestable; the meanings of the terms are determined by the community (intersubjective). The division of linguistic labour is made

clear when we considered the different types of authorities contained in the linguistic community. Similarly, these conditions hold if I am an incompetent linguistic agent in regards to toadflax, and a colloquial user of the term ‘geophysicist’. The three conditions discussed, then, hold whether I am an incompetent, partially competent, or competent language user.

Manifestability, intersubjectivity, and division of linguistic labour raise an important point — the examples clearly show that, e.g., manifestability does not relate to the psychology of the linguistic agents.⁶

Other cases of deficient knowledge and ways of appealing to some authority to gain access to the conventions of the linguistic cases are easy to come by, but the principles are the same. One need not have full knowledge, nor implicit knowledge, of a language to be a member of a linguistic community — but they must have *access* (in some way or another) to the conventions of that linguistic community.

3.7 Conclusion

To conclude, we have learned that Williamson is quite right — luminosity would indeed be problematic, if not devastating, to a semantic conventionalism that required it. However, having accepted Williamson’s anti-luminosity argument, we have also shown several conditions of a community-based semantic theory — when considering what a conventionalist ought to say to take the sting out of Williamson’s argument, we have uncovered a version

⁶Perhaps a further example would be the ever occurring red lines under words when I am writing quickly. Who determines the correct spelling? Perhaps a dictionary: *but* a dictionary relies equally on language use of the community — a standard is generally set such that a certain amount of occurrences of a word must appear in print before accepted as a standard/conventional meaning. But, I am assuming that if an author of a dictionary discovers that all the 500,000, or whatnot, applications originate from the same author they will not accept the use as conventional.

of conventionalism that has the appealing features of manifestability, intersubjectivity, and division of linguistic labour.

The argument presented by Williamson shows that by anti-luminosity, luminosity of the meanings of lexical items is incoherent. That is, triviality follows *if* the linguistic agents must be aware of the meanings of lexical items in the way Williamson assumes other authors take such knowledge to be (i.e., luminous). But then by sufficient (internal) reflection an agent can come to know the meanings of the lexical items he employs. Williamson has quite rightly shown that internal reflection is insufficient.

In this chapter, then, we have learned several important lessons in regards to semantic conventionalism: community-based language is not threatened by the anti-luminosity argument. By placing the conditions of externality, manifestability and the division of linguistic labour on the community⁷ — made clear by the examples and the subsequent discussion of community-based semantics — I have shown luminosity *does not* follow.

Hence we have shown that the objection we began with is not knock-down. And in fact, by accepting Williamson's anti-luminosity argument, we have shown that it ceases to be an objection *at all*.

The conditions which have become apparent by considering Williamson's argument, and the examples given above, help in establishing several criteria that are necessary for a workable semantic conventionalist theory. So, in fact, by considering this 'objection' we are well on our way to making more explicit what I take to be a viable semantic conventionalism.

To sum up: the basic structure of the argument does have an effect on conventionalism. The first premise of the argument:

If conventionalism is correct, then manifestability follows and if manifestability is accepted, then there are non-trivial luminous mental

⁷Note that I take all three as contributors to the condition of intersubjectivity.

states.

is rejected, since we have shown the second conjunct to be false. So, as I have shown, semantic conventionalism should not (and cannot) be a viable theory if luminosity is accepted.

Chapter 4

Molecularism

4.1 Introduction

A popular view in the philosophy of language literature, advocated notably by Fodor and Lepore,¹ is that there are only two viable sorts of semantic theories, namely holism² or atomism.³ It is thought that there is no semantic “third way” because of the Quinean rejection of the analytic/synthetic distinction. Fodor and Lepore further argue that holism is “crazy” (or at least that no tenable arguments have been put forward in favour of it), and so it seems that the only acceptable semantic theory is atomism. To be explicit, the argument presented by Fodor and Lepore is a disjunctive syllogism. They maintain that *either* holism is correct or that atomism is. Since holism is crazy, it must be the case that atomism is correct.

It is on the basis of reasoning of this sort that one might think a second knock-down argument against conventionalism can be constructed.

The supposed knock-down argument goes as follows:

¹See the introduction of *Holism: A Shopper's Guide* (Fodor and Lepore 1992).

²Holism is the view that the meaning of all lexical items in a language are related to, and affected by, all other lexical items.

³The view that each lexical item in a language has meaning independently of every other lexical item.

- (1) If conventionalism is correct, then there must be a principled analytic/synthetic distinction.
- (2) There is no principled analytic/synthetic distinction.
- (3) Therefore conventionalism is not true.

The reason for (2) is that it is accepted, for Quinean reasons, that there is no principled analytic/synthetic distinction. I will not concern myself with (2). My focus is on premise (1), the rejection of which will show that the objection is mistaken.

First I need to show precisely why it might be thought that the above *is* a knock-down argument — what makes the objection appear to be plausible?

Fodor and Lepore argue as follows: either semantic *atomism* is correct or meaning is determined by the role played by a word in relation to other words. *And not both* by definition of atomism.

Atomism is inconsistent with the type of conventionalism that I advocate. Consider language games — if we are going to accept that there are such things, then it cannot be the case that lexical items have meaning in isolation, which is what atomism means. So, we know that atomism cannot be the right view for conventionalism.

Now, once one has accepted that meaning is determined by relations to other parts of the language, some might ask “*which* parts?” In order to avoid holism (i.e., to avoid saying “all of them”), one needs a *principled way* to distinguish which relations confer meaning and which do not.

The analytic/synthetic distinction is the only candidate that seems to be on offer as a principled distinction to determine which relations are meaning conferring (the analytic ones) and which are not (the ‘others’). But, by (2), that is not a *principled* distinction. So, why not just accept holism? Because it is obviously false: it makes language unlearnable. In essence, the

conventionalist should reject holism for the same reasons Fodor and Lepore do.

Summing up: if you don't have a principled analytic/synthetic distinction and the meaning of words is determined by relationships between its occurrences of other parts of the language, then you have to say that all its relationships go to determining its meaning. Atomism, on the other hand, means that words have meaning in isolation, perhaps by their relations to things in the world. So, holism and conventionalism have a feature in common: the meaning of a word is its role in the sentences in which it occurs. Meaning is not, and cannot be, determined in isolation from the rest of the language. But, as mentioned, if holism is correct, there is no way of cordoning off sections of the language which are meaning-constituting. Since the analytic/synthetic distinction is the only candidate for doing *this* cordoning, conventionalism needs the analytic/synthetic distinction.

To put this point another way, conventionalism cannot be either atomistic *or* holistic. If conventionalism were atomistic, then it would seem that 'meaning is use' fails — and this maxim, in combination with the maxim that to be a competent linguistic agent is to know the rules of the game, is essentially what it means to be a conventionalist theory of meaning, since the community decides what appropriate use *is*. Atomism, relying on the acquisition of individual lexical items rather than learning pre-existing conventions, *from* which we grasp the meaning of words, is the wrong way round for a semantic conventionalist.

I argue that the supposed objection and its explication by Fodor and Lepore presents a false dichotomy — a semantic 'third way', molecularism, is possible, and plausible. What I have in mind is the development of a 'weaker' distinction than the analytic/synthetic distinction but one which is *principled*, which the analytic/synthetic distinction, at least if Quine is

right, is not — namely one which distinguishes meaning-constituting and nonmeaning-constituting inferences. This avoids both holism and atomism, but it does so in a way such that the conventionalist is not committed to the analytic/synthetic distinction.

In order to provide a principled distinction I: (1) establish a type of inferentialism which is consistent with conventionalism — one which does not fall into holism — as Fodor and Lepore argue is inevitable with any theory other than atomism; (2) show that such an inferentialist picture is molecularist. On the basis of a distinction between what I call materially correct and non-materially correct (but still permissible) inferences we *can* cordon off sections of the language. On the basis of such a type of inferentialism I show that it is correct to say that it is relations between lexical items that determine meaning, but only *some*, not all such relations. This molecularist inferentialism gives conventionalism a way of separating off the inferences that count as meaningful from the ones that aren't. Roughly, what I shall argue is that the inferences that count for meaning are what we shall call materially correct, but not logically correct.

There *is* a principled distinction to be made by using molecularist inferentialism: that is, between meaning-constituting and nonmeaning-constituting inference. What I call 'materially correct inferences' count as meaning-constituting and 'other inferences' are nonmeaning-constituting. The distinction is drawn on the basis of a notion of *particular types of inferential relations* which shall be developed — materially correct inferences determine meaning, and thus words gain content only on the basis of these, since these are the inferences licensed and determined by the community to be constitutive of the meanings of particular terms. The distinction rests on the differences between direct and indirect inferences. A direct inference is where, e.g., *B* is inferred from *A* (with no intermediate steps). An indirect

inference is one where the conclusion of an inference essentially involves more than one step, e.g., concluding C from A , but by way of a second step — A to B . So, the inference now is A to B to C .

To distinguish direct and indirect inferences I must establish a hierarchical account that does not allow one to jump immediately from A to C . This hierarchical account allows me to show why permissible inferences from A to B are materially correct, and thus meaning constitutive, while inferences from A to C (via B) are not meaning constitutive. To do this, I must provide a hierarchical account of inferences that does not allow transitivity of the relation of material correctness, i.e., in such a way that the inference between A and C is not materially correct. As a sort of corollary I will further show that there is an ‘asymmetry’ built into such a hierarchical account of permissible inferences.

It is also important to distinguish material from formally correct inferences. In formal inferences there is no *content*, so, whilst some are direct inferences they are not meaning constitutive. That is, in addition to the direct/indirect distinction, I will make clear that there is a second subset of inferences which, whilst permissible *and* (sometimes) direct, are not meaning-constituting — namely formally correct inferences. In contrast to holistic inferentialists, I argue that since logical vocabulary is *contentless*, logical inferences, even if inferentially direct, do not count as materially correct.

On the basis of licensed, meaning constitutive, inferences, the meaning of words (and so the existence of ‘concepts’)⁴ is determined by the licensed inferences endorsed by the intersubjective practices of the linguistic community — contrary to the traditional views that concepts are either platonic or psychological states of individual agents, both those theories have previously

⁴Note that, according to my view, contra other theories of semantic inferentialism, *conceptual content* falls out of the picture.

been rejected as unsuitable for semantic conventionalism. Concepts, on this view, are derivative from the meanings of words and sentences and so, being embedded in a semantic conventionalist theory, have to be determined by the community and are thus (at a minimum) shown to be intersubjective. But, the way that the community determines the meaning of words is by the licensing of materially correct inferences.

What this amounts to is the idea that *language games* are hierarchically arranged, and the relation of inferences being meaning-constituting inferences is not transitive. So, *only* once an agent has mastered inferences at one level is it possible to acquire concepts of a higher level. ‘Lower’ level inferences, whilst still meaning constitutive in relation to other lower level inferences, do not count as meaning constitutive in relation to higher level materially correct inferences, as we shall see.

The distinction that I draw is, as mentioned, *weaker* than the analytic/synthetic distinction — but I think this is not a problem. If it does the job of establishing that molecularism is viable, then it does not need to be as clear cut as the analytic/synthetic distinction was meant to be, which, after all, is rejected by many philosophers.

My story, then, is this: materially correct inferences determine meaning. Formally correct inferences are not materially correct, and so are not meaning-constituting inferences. The upshot of the discussion in this chapter is that conventionalism which is based on molecularist inferentialism is *at least* in no worse a position than other theories. To establish this much all I need to do is to establish the *possibility* that molecularist inferentialism can distinguish between meaning-constituting and nonmeaning-constituting inferences. This is not to say that I am presenting the *only* possible theory, nor that it is even the *right* theory — only that it is *one* way to make the distinction, i.e., a weaker distinction than the analytic/synthetic distinction

is not detrimental to semantic conventionalism and is principled in a way that the analytic/synthetic distinction is not. So, semantic conventionalism based on molecularist inferentialism will be shown to be a stronger position to take for conventionalism than the analytic/synthetic distinction, previously thought to be the only road to molecularism.

In what follows I borrow heavily from the inferential role semantics developed by Robert Brandom in *Making it Explicit* (Brandom 1994). The key feature of Brandom's view is that *some of the inferential roles of an expression constitutes its meaning*. But, since Brandom is an inferential holist, he accepts the argument from Fodor and Lepore that if *some* inferential roles constitute the meaning of a lexical item, then *all* of the inferential roles of an expression constitute its meaning.

I shall *modify* Brandom's view, showing that his work is a natural starting point for an inferentialism of a different character.

4.2 Holism and Atomism

For starters, let us get some definitions on the table — this will assist in establishing the need for a semantic middle ground.

The definition that I use for holism is that it is a view which implies that every lexical item of the language is relevant to determining the meaning of any individual term in the language. If inquiring into the meaning of a term or sentence, the entire language must be taken into account.

It must be noted that some authors suggest that there are different types of holism. There is the 'out and out' holistic account of meaning where, as mentioned, every lexical item is relevant to the meaning of other lexical items. If the meaning of one sentence (or 'node' in the language 'web') changes, then the meaning of all other lexical items are revised also. A somewhat different account of holism is that, if *relevant*, then a change in

meaning of a lexical item will only affect the meanings of others in the *vicinity*. For instance, if numbers are defined as abstract, platonic, entities and one shifts to a fictionalist account thereof, then the definition of *all* numbers would need to be altered. This does not imply that the meaning of other sentences changes, only those relevant to talk of number theory.

Talk of numbers is, what one might term, a *language game*. Being an independent language game from, e.g., talk of middle size objects (chairs, desks, people, etc.), means that one language game may not affect all language games at once. It could be argued that what changes is merely the language game in question — the others are left intact unless changes are made to these other language games. This latter form of holism strikes me as not really holistic at all. By distinguishing language games and the fact that changes in those games does not bleed through to others strikes me as molecularist. Indeed, it strikes me as close kin to the view I am defending. The challenge of avoiding collapse into ‘out and out’ holism by giving a principled account of *relevance* of the relations between parts of language is essentially the same project as I am tackling.

The definition for atomism is that the semantic atoms, namely words (or those items in the language which are concatenated, such as “dog-paddle,” “underhand,” “prehistoric,” etc.) are taken to have *a* meaning, and have their meaning in isolation from the rest of the language. Atomism is usually taken to be bound up with a causal theory of meaning. It follows that words are learnable and can, at least in principle, be added to a speaker’s linguistic competence, individually. Sentences, therefore, only have meaning in virtue of the meanings of the individual terms involved.

Notice that this turns the context principle on its head. The context principle is that the meaning of a word is determined by the roles it plays in sentences. The atomistic view, however, would have it that the meaning

of words are prior to sentential meaning. The latter is derived from word meaning.

4.3 Plausibility

As briefly suggested at the start of this chapter, the knock-down argument is closely related to one found in various forms in the philosophy of language literature. It will be worth our while to start with that argument, which gets a clear statement from Fodor and Lepore. Their argument is, using what I hope are sufficiently obvious abbreviations:

- (1) Atomism \vee \neg Atomism⁵
- (2) \neg Atomism \rightarrow Holism
- (3) \neg Holism
- (4) Therefore, Atomism.

This argument is obviously classically *valid*, so its plausibility depends on the plausibility of the premises. Since premise (1) looks pretty safe (even considering my advocacy of intuitionistic logic in Chapter Five), it is (2) and (3) that are of most concern. We will consider (3) in the next section. Let us now consider (2).

Fodor has been arguing for something like (2) for some time. Consider what he (Fodor 1987) and Devitt (Devitt 1993) call the “basic argument.”

- (A) Some of an expression’s inferential properties constitute its meaning.
- (B) If some of an expression’s inferential properties constitute its meaning then they all do.

⁵Fodor and Lepore call any theory that is not atomism *anatommism*.

(C) So, all of the inferential properties of an expression constitute its meaning. (Devitt 1993, 17)

If we accept that all non-atomistic semantic theories account for the meaning in terms of inferential properties, the basic argument becomes an argument for (2).

I have suggested that it is the lack of a principled analytic/synthetic distinction that grounds (B). Fodor and Lepore are quite explicit about this:

Premise 1 [Semantic properties of sentences] of a language are molecular.⁶

Lemma: If Smith has the belief that P, he must have other beliefs not identical to P.

Premise 2: There is no *principled* distinction between the propositions that Smith has to believe to believe that P and the propositions that Smith doesn't have to believe to believe that P.

Conclusion: The property of being-some-belief-or-other-of Smith's [and presumably being-some-formula-or-other] is holistic. (Fodor and Lepore 1992, 23-25)

Fodor and Lepore use examples to make premise (2) plausible. Consider:

- Theory₁: S₁: The trees are losing their blossoms.
- This is partially confirmed by S₂: There are high winds.
- The fact that there are high winds is partially confirmed by the observation report that S₃: The branches on the trees are all moving.
- S₄...S_n. Take S_n to be, e.g., High winds signal a coming storm.

⁶This is a paraphrase of the first premise of Fodor's argument. There is no need to provide the entire premise since, for our purposes, they amount to the same thing.

But, Fodor and Lepore assume that it is outlandish to suppose that one needs to believe that high winds signal a coming storm in order to believe that the trees are losing their blossoms. Moreover, we could come up with any number of examples: S_1 : the trees are losing their blossoms. S_2 : there is an elementary school in the area. S_3 : some children are vandals... S_n : if parents instructed their children in appropriate behaviour the trees would be left alone.

Extrapolate from these two examples and we see that Fodor and Lepore are right, to this extent — big problems ensue if meaning is determined by relations to other sections of the language *and* there is no way to cordon off sections of the language. Indeed, that holism follows from the denial of atomism if one rejects the analytic/synthetic distinction, as Fodor and Lepore do, is not implausible. The hidden premise, of course, is that the only candidate for blocking, in a non-arbitrary way, chains of inferences such as the above, is by allowing only analytically correct inferences.

4.4 Objections to Holism

Let us turn next to premise (3) of Fodor and Lepore's argument — that holism is false (indeed crazy). This, of course, is a claim with which the conventionalist can happily agree, if I am right.

The first reason to question holism is its sheer implausibility. It seems highly improbable that there are inferential relations between such concepts as ducks, bathtubs and, say, the number 2, such that the concepts are dependent on one another for their meaning.

What is the reason, one might wonder, to suppose that the semantic content of a sentence depends on the entirety of the language? It seems intuitive that there is, or at least *should be*, a way of demarcating sections of the language so that the meaning of a sentence is related to, gained by,

or changed by, only a small subset of the language, not the whole language. Even if one were to argue that all *concepts* can be inferentially related to each other, it is a leap from that, to the view that the *meaning* of a sentence is determined by the entire language.

Further, demarcating the boundaries of English is an important theoretical question for holism. This is because, if the language is constantly revisable, as English of course is, then every piece of the language must be revisable at the same time. But it seems doubtful that adding a new concept to a language, or having one that gradually changes over time, affects the rest of English all in one go. Such a change might affect the semantic value of *some* sentences that are in the semantic or conceptual neighbourhood, but it is highly doubtful that *all* sentences or concepts are, or even are possibly, affected.

Second, it is often thought that holism should be conceived as fundamentally related to beliefs. This means that the semantic contents of sentences depend on the psychology and individual understanding of linguistic agents. But this is highly questionable. If belief and semantic content are thus connected, then, for anything that changes in a person's web of beliefs, there is some corresponding change in the semantic content of everything in the web. Further, this would make language private, and would mean that there would never be the exact same content of a sentence for different members of a language community. Each change in the beliefs of an agent would signal changes in the meanings of sentences in that agent's idiolect — making a shared language highly unlikely.

4.5 Objections to Atomism

We have briefly considered Fodor and Lepore's case for atomism. Unfortunately for them, atomism faces equally serious problems which make it, like

holism, an untenable position.

There are broad theoretical problems with atomism. One example is that there is, as yet, no plausible account of the causal theory of meaning, which is proposed by many atomists as the theory needed to explain the semantic content of atomic expressions.

The causal theory, in a very crude form, is the view that a token of “cow” in ‘the head’ means COW because the tokening is caused by cows. It is the fact of having a type of token that is differentially sensitive to the presence of cows that *constitutes* the tokening of the concept “cow.” Semantic atoms supposedly gain their content from such causal relations.⁷

But, consider the disjunction problem. For instance, suppose that you are in your backyard at twilight, you see a shadowy object at the end of the yard which causes the tokening of “dog” in your head. However, it was not a dog, but a large cat, which caused the tokening. So tokenings of “dog” are caused by dogs and, sometimes, cats. But we can broaden this out: perhaps it was a huge rabbit at the end of the yard. So tokenings of “dog” are caused by dogs *or* cats *or* rabbits. Perhaps the sensible thing to say is that “dog” is caused by the presence of a suitably sized shadowy shape at the end of the yard. But then, by the causal theory, *a suitably sized shadowy object* should be (part of) the meaning of “dog”. One readily sees how such considerations are particularly problematic. One might ask: how then is it possible *ever* to misrepresent anything or for a concept to have a particular content? The atomist is neither able to account for causation as a suitably discriminating determinant for semantic content nor solve the problem of

⁷There is a multitude of issues that could be raised in context of atomistic language acquisition, such as the question of whether children acquire single words by ostensive definition, by some eye gaze detection mechanism, or some other cognitive ability. This question is an important one in the philosophy of language, but one which we are not able to pursue here.

misrepresentation. It is fair to say that most (non-causal) atomists (and even some causal atomists) agree that no satisfactory solution to the disjunction problem is yet on offer.

Dretske is perhaps the most famous advocate of the causal theory of meaning. His account variously (as his work progresses) relies on *reliable indicators* grasped during associative learning; the behaviour an agent exhibits in light of certain stimuli; *functions* from stimuli to behaviour; *dispositions* to behave in a certain way; and compositionality. However, Sturdee (Sturdee 1997) argues persuasively that none of the progress in Dretske's account of misrepresentation can adequately account for misrepresentation. In all cases the disjunction problem still looms large.⁸

Fodor's solution in *Psychosemantics* (Fodor 1987) is to reject what he calls the 'Crude Causal Theory' (see Chapter Four of *Psychosemantics*), which he attributes to Dretske's account of informational semantics in *Knowledge and the Flow of Information* (Dretske 1981). The way to deal with causally constrained semantic content is to provide a 'Slightly Less Crude Causal Theory of Content' (SLCCTC) (Fodor 1987, 126):

The Slightly Less Crude Causal Theory of Content offers the following two friendly amendments: for (2) [that is in the CCTC: "only instances of *A* cause tokens of 'A'.] read: 'If non-*A*'s cause 'A's', then their doing so is asymmetrically dependent upon *A*'s causing 'A's'. For (1) [which for the CCTC is: "every instance of *A* causes tokens of 'A'.] read: 'All instances of *A*'s cause 'A's' when (i) the *A*'s are causally responsible for psychological traces to which (ii) the organism stands in a psychologically optimal relation. (Fodor 1987, 126)

For an account of why the SLCCTC also fails to account for misrepresentation see Maloney's 'Mental Misrepresentation' (Maloney 1990). Fodor's

⁸Literature surveyed by Sturdee to make his case is: *Knowledge and the Flow of Information* (Dretske 1981), 'Misrepresentation' (Dretske 1986), *Explaining Behaviour* (Dretske 1988), and 'Conscious Experience' (Dretske 1993).

notion of asymmetrical dependence rests on counterfactual conditionals: ‘horse’ tokens HORSE if there is a nearby possible world where HORSE is not caused by a cow *or* a horse *or* a muddy zebra *or*... Maloney states that appealing to possible worlds where the token of horse is only caused by actual horses also does not work. He says:

Since cows form a natural kind and natural kinds essentially have their (intrinsic) causal powers (that is, the causal powers partially definitive of the kinds) cows in #’s worlds [the counterfactual worlds] have their characteristic causal powers [which I assume to mean that they have the same causal powers as they have in the actual world]. Yet in #’s worlds cows fail to cause ‘horse’ tokens. Why?... What prevents the cows in the worlds of # from causing ‘horse’ tokens in cognitive agents is that the relations that the cows bear to the agents in those worlds happen to preclude cows from exerting their powers to cause ‘horse’ tokens in the agents. But these are just the circumstances that are also germane to the causal connections between horses and ‘horse’ tokens. (Maloney 1990, 449)

Thus, on Maloney’s account, appealing to counterfactuals does not supply the ‘slightly less crude causal theory’ with any way which accounts for misrepresentation. Other objects must bear their usual causal powers in the counterfactual worlds, and the *correct* tokenings are still not immune to the disjunction problem. If horses have the same causal powers as they do in the actual world, then presumably, Maloney’s point is that it is ad hoc for other objects to be discounted as having the same causal powers.

4.6 Inferentialism

Consider again Fodor and Lepore’s argument. If I am right that *atomism* is not an acceptable semantic theory, something is clearly amiss. Of course, the present project is an investigation of the plausibility of conventional-

ism, which is incompatible with atomism anyway, but the preceding section suggests that conventionalists should have plenty of company in rejecting atomism.

What, then, of the prospect of rejecting (3), i.e., accepting holism? This too is incompatible with conventionalism, as well as being highly implausible on independent grounds.

Thus the conventionalist, and anyone else who agrees that both holism and atomism are deeply problematic, must turn to the job of explaining what is wrong with (2). As advertised, I propose to do so by offering an alternative to the analytic/synthetic distinction as a way to mark off those linguistic relations that are meaning-constituting from those that are not. I do so on the basis of inferentialism.

Robert Brandom, probably the most well-known contemporary advocate of inferentialism, presents a holistic inferentialist theory (Brandom 2000). I show that inferentialism *needn't* be holistic. The brand of inferentialism I offer is thus quite different from Brandom's. Nevertheless, it will be much more efficient to describe my view as a modification of Brandom's than to try to explain it from scratch.

The molecularist inferentialism rests on the distinction between *meaning-constituting* and *nonmeaning-constituting* inferences. I shall borrow, in modified form, a useful bit of jargon from Brandom's version of inferentialism — meaning-constituting inferences are 'materially correct' inferences — but more on this later.

We began the first half of this discussion with a schematic presentation of Fodor and Lepore's argument. Let us begin this half with a schematic presentation of my understanding of the conventionalist's argumentative predicament.

(1) Holism is incompatible with conventionalism, because it would make

language unlearnable.

- (2) Atomism is incompatible with conventionalism, because the community drops out of the picture — if the meaning of a lexical item is due to some suitable causal theory, then it is facts about the world that determine meaning, not conventions.
- (3) So, Fodor and Lepore have presented a false dichotomy.
- (4) And so, conventionalism is not shown to be unacceptable on the basis of a lack of a principled analytic/synthetic distinction.
- (5) Therefore, there must be a semantic third way.

One way (perhaps not the only one, but a viable one at least) of establishing a semantic third way is to develop molecularist inferentialism. Before going much further, let's set out characteristics of a couple of sorts of molecularism.

Note that not an awful lot of details have, as yet, been given in regards to a definition of molecularism. This does not matter for present purposes. However, others conceive of molecularism, the main point is that it can and will be developed further using inferentialism. We shall see how this is done later. But, note that the basic condition of molecularism remains the same: meaning is to be explicated in terms of (some) connections between distinct concepts seems to be the consensus:

Molecularism is a view of our contentful conceptual states in which the content or meaning of a concept is constituted, at least in part, by a small selection of its connections to other concepts...

Or:

Semantic molecularism, like semantic holism, holds that the meaning of a representation in a language \mathcal{L} is determined by its relationships to the meanings of other expressions in \mathcal{L} , but, unlike holism, not by its relationships to every other expression in \mathcal{L} . (Audi 1995)

So, molecularist inferentialism is our way out, as it were, for conventionalism — meaning, determined on the basis of material correctness and the corresponding distinction between meaning-constituting and nonmeaning-constituting inferences, allows the conventionalist to avoid the highly problematic atomistic and holistic theories. And further, such a distinction is principled and can take the place of the analytic/synthetic distinction.

4.6.1 Material Correctness

What can be said about materially correct inferences is not an easy matter. They are treated by Brandom as primitives in that they are “unexplained explainers” (Brandom 1994, 133). In this I agree with Brandom — until I get to the section regarding direct and indirect inferences I will have very little to say about the nature of materially correct inferences. Even at that stage, though, a *complete* account of what a materially correct inference actually is will not be forthcoming (because I think it cannot). However, my characterization *is* more robust than Brandom’s, since I can appeal to language games and the competency of linguistic speakers to cash out in more detail what materially correct inferences are.

Brandom advocates inferential holism: for example, he is and *must be* a logical expressivist (we shall expand on this in later discussion). He sees the role of logical vocabulary as being important in cashing out the ‘commitments’ of linguistic agents. So, any formally correct inference is meaningful because of the role it plays in the ‘scorekeeping’ practices of agents engaged in drawing materially correct inferences. Formal correctness, then, is not meaningful in the sense that formal assertions have semantic content, but only in virtue of the expressive nature which allow us to *understand*, or unpack if you will, the commitments of the speakers of a language.

This is just one example in which Brandom and I diverge in our views

— I do not think that logical inferences are materially correct because there is no content to the logical vocabulary.

But, since both Brandom and I do think that meaning is determined by materially correct inferences, and since my notion of materially correct can profitably be considered a modification, let us look at some examples to get some indication of what this means.

Consider:

- (1) It is raining.
- (2) Therefore, (*ceteris paribus*) the streets will be wet.

Of course, there could be counterexamples to such cases. Suppose that the street an agent lives on is lined with leafy trees. Then, perhaps, the canopy of the trees causes the streets to stay dry — hence the ‘*ceteris paribus*’ clause.

The materially correct inference above is not, Brandom insists, correct because of the employment of any ‘hidden’ logical rule, such as *modus ponens*. The contrast he has in mind with an argument with an embedded logical rule, such as (1) If it is raining, then the streets will be wet, (2) it is raining, (3) therefore the streets will be wet. It is important to Brandom that not all materially correct inferences are *logically correct*

Another example is:

- (1) Bob is a bachelor.
- (2) Therefore Bob is unmarried.

This is an interesting example — being, as it is, a stock example of an *analytic* inference — because while it, like the first is not *logically* valid (i.e., is not valid in virtue of *form*), it is unlike the first in not being *ampliative*. It is not possible that the first claim be true and the second not, so there is no need for a *ceteris paribus* clause.

Brandom's view, to reiterate, is that meaning has to be determined by materially correct inferences. Further, which inferences are materially correct is determined by their acceptance or rejection of the inference by the agents of a linguistic community. The 'scorekeepers' — that is, the deontic agents of a linguistic community — track the inferences and determine whether they are permissible or not.

In some ways this is similar to semantic conventionalism, since acceptability of an inference is determined by the community, and the competency of an agent is determined by their ability to assess and use sentences (or in this case, inferences) correctly on the basis of the conventions of the language game.

To turn this into an account useful for conventionalists, we accept that materially correct inferences are meaning conferring. But, we also distinguish *distinct* classes of correct inferences, and thus I avoid holism — some classes of inferences, i.e., the materially correct inferences are meaning conferring. Such inferences but must be *direct* in a sense to be clarified in a later section. And, further, logically permissible inferences, whether direct or not, *do not* count as materially correct. As we see from the above, if the two previous examples are taken to be materially correct, then some are ampliative — that is, one can draw conclusions which contain information that is contained in the premises. This is not permitted in my account of logical vocabulary. As I will show, logically permissible inferences fail to be ampliative, precisely because logical vocabulary has *no content*. But as the second example makes clear, some materially correct inferences are *not* ampliative, which is why I describe an alternative criterion for determining which vocabulary is logical.

4.7 Direct/Indirect Inferences

It is time for me to try to make somewhat clearer how I think the distinction between meaning-constituting inferences and other correct inferences can be drawn. I will use several examples to approach what is quite a complex matter, and I do not pretend that I will be able to give an airtight characterization here. The gist of this section is that: (1) meaning is determined by some special sort of correct inferential relation. As is typical of conventionalist views, whether an inference is correct or not is *ultimately* determined by the linguistic practices of the community. However, (2) failing to accept a *direct* inference shows an agent to be incompetent in a particular language game, and is grounds for judging that the agent has failed to understand the meanings of the lexical items which are at play in direct inferences.

Obviously, calling inferences ‘direct’ means that there must be a tight relation between the premise and conclusion. Now we see more clearly the nature of this tight relationship — the relation of the premise(s) and conclusion of an inference is said to be direct when the lexical items employed in the former are related to the latter in such a way that failure to recognize the correctness of the inferences is no mere error, but betrays a failure to comprehend the subject under discussion.

4.7.1 Non-logical Inferences

Consider the following examples, both of which are plausibly *direct* in the relevant sense, since someone who fails to accept them plausibly can be charged with failing to know what ‘two’ means, for instance:

There are two items on the table

There is more than one thing on the table

There are two items on the table

The number of things on the table is less than three.

When teaching a child how to count, it is common to use middle-size objects (e.g., oranges) to show the number of things in front of the child. Of course, they cannot immediately extrapolate from the number two to any particular number — perhaps the child only knows the numbers up to three. At that stage, we probably would not yet say the child has mastered ‘two’, and the class of meaning-constituting inferences for ‘two’ might not have sharp boundaries is facilitated with small numbers up to, say, one-hundred required, or is fifty enough? But it is clear that a grasp of the transfinite hierarchy is *not* required. Potential *vagueness* of the notion of *direct inference* doesn’t indicate that it is unprincipled, and there are clear cases on both sides of the direct/indirect divide.

If for example the child were to infer ‘there are two oranges on the table’, ‘therefore one of the oranges is mine’ — the inference might be correct, because of some feature of the context. But, if another child does not accept that inference, the second is *not* thereby shown to be linguistically incompetent with regards to ‘two’. That inference, if correct ‘mixes games’, and so is indirect.

Let’s now turn to a couple of further characteristics of this notion of directness that will enable us to show that it’s actually quite a powerful tool for explaining why holism needn’t follow from inferentialism. First, while *correctness* of inference is presumably transitive, *directness* is not. Perhaps Brandom’s example “It is raining, therefore the streets will be wet” is direct in the present sense, and so too, plausibly, is “If the streets are wet, slipping is likely.”; it is a good deal less reasonable to think we should convict someone

who fails to accept “slipping is likely” from “It is raining” of conceptual incompetence. Longer chains of inference could make the matter plainer, of course.

It is also clear that mastery of *some* language games is possible only after some others are acquired, and obviously this relation is asymmetric. One cannot master the complex numbers *before* one has mastered the small natural numbers, for instance. But most of humanity count as evidence that the other way around is quite possible. Thus language games will be *partially ordered*.

Finally, what counts as *mastery* of a concept is going to have a lot to do with what we *use a language game for*, and so the “hierarchy” of language games will reflect the concerns and interests of the linguistic community in important ways. For instance, *membership in the community* (of persons) plays a fundamental role in the functioning of our society. Thus whether or not an individual falls into some other class is a person or not is often part of the *content* of the word that attaches to that class. Someone who doubts that it follows from “That’s a dog.” that “It is not a person.” betrays either linguistic incompetence or careless reading of Peter Singer. The content of ‘dog’ in English on this account is not specified (atomistically) by the biological species.

Let’s look at an example that puts these ideas to work. It is *correct* to infer:

There is a duck in the bathtub

There is a bachelor in the bathtub

There are two things in the bathtub

The argument, though, is not *logically* correct, as the conclusion only

follows because ducks are not bachelors. One might think examples such as this will push us back towards holism — do we need to understand what a *duck* is to master the concept *bachelor*? I think not, and the hierarchy of language helps us see why

It is clear that we need to understand “unmarried” in order to understand “bachelor”. Knowing that being unmarried is something to do with people this links to the concept “person.” Further, to employ the concept “duck” properly we must know that it is not a person leading to the concept “person” (via “non-person”). Now the concept “duck” *reaches* “person,” as does “bachelor.” The various links explain *how we know* that a duck is not a bachelor, and so how we see the correctness of the inference. But the asymmetry of the meaning dependencies also allows us to see why this doesn’t create a link of *meaning* between ‘duck’ and ‘bachelor’.

Bachelor \Leftarrow Unmarried \Leftarrow Person/Non-person \Rightarrow Duck

One never acquires a concept in isolation to be sure. But, this does not mean that learning concepts in *relation* to others means that the *indirect* inferences are bound up in such a way that the inferences are materially correct. Simply having one concept being ‘reachable’ from another on the basis of inferential relations is not sufficient to establish a link of meaning. Multi-step links meaning-constituting inferences can be, and often are, *indirect*. This is just what it means to say that ‘directness’ is non-transitive. Indirect inferences are not materially correct and so are not meaning constitutive.

To be explicit, notice the necessary steps for linking bachelors and ducks. *Bachelor* goes to *unmarried* and then to *person*, while *duck* goes to *non-person* and *person*. So these concepts *both* depend on the concept *person*. But there is no immediate justification present that such an inferential relation means that “bachelor” and “duck” share meaning — the inference, in short, is not materially correct. This follows from the fact the ‘direct

inference' relation is not transitive.

Let's consider another example. Suppose, for instance, that a young child is taken to the park to watch the llamas — she, one would hope, would be aware that llamas are different in kind to people. But that does not require her to understand bachelorhood or being unmarried.

So, conceptual *linkage* is not sufficient for material correctness. The mere possibility of a *link* between disparate concepts, then does not show that the meaning of each depends on the other. So, conceptual linkage, in short, does not mean that the content of the concepts bachelor and duck are in any way connected semantically. *If* there is any shared content at all it conceivably stops at person. To understand and employ the concept “duck”, in other words, we do not need to understand the concepts “unmarried” or “bachelor”. Rather, they share a more basic concept, that of personhood, but one can master either concepts while having no acquaintance with the other.

The hierarchical account, under which some lexical items can only be mastered when you are in possession of other concepts, is both plausible and useful. Plausible because you can only acquire the concept “bachelor” if you have mastered the notion of “unmarried,” and this will be exhibited in the correct usage of the words. Similarly, perhaps you cannot master a notion of a duck if “non-person” is not mastered. But, the ability to grasp the notion of “duck” is independent of grasping the notion of “bachelor.” You *might* learn what bachelor means, and hence be able to make the indirect inferences to ducks, but this is only *in principle*.⁹ It is not *necessary* to know or employ the word bachelor appropriately to know how to do the same with the word duck. The *usefulness* of the hierarchical view is that the asymmetry allows us to explain why.

⁹As the llama case shows.

This phenomenon is persuasive, and it is a significant defect of holism that it cannot account for it. Once one begins to look, it is easy to produce more examples of this asymmetric dependence. Consider again the natural numbers, but in a different fashion than the first example. An agent might need to be able to understand the concept “natural number” to understand “fraction,” but it seems unlikely that he would have to have mastered fractions to understand natural numbers. Children, for instance, understand what numbers are very much earlier than they understand what fractions are.

So, it is not the case that every concept is connected in a meaning-constituting manner, even if there is an intermediate concept that is required for mastery of each of two words — the point is that knowing the meaning of an intermediate lexical item by knowing the licensed inferences in relation to one word, does not mean that the licensed inferences going from the *other* word need be understood.

Hierarchy of Concepts

Consider again the case of the child gaining access to the language game of numbers. Let us assume that he is competent in using the numbers one through ten. The inferences that he can draw on the basis of this knowledge can only concern *those* numbers. The language game which would extend his knowledge is not accessible to him yet (and so we might call him a *partially* competent language user in regards to the number game).

Similarly, one can understand what a natural number is, without understanding fractions, or without understanding (or even conceptualizing) the infinitude of the natural numbers.

Clearly, there is some sort of hierarchy of concepts on the go, such that one can only acquire one if an agent is a competent user of another language

game. Not only does this highlight the hierarchy, but it brings it together with the asymmetry and nontransitivity mentioned previously. Consider:

Natural numbers \rightarrow the infinity of natural numbers (\aleph_0) \rightarrow the powerset of the natural numbers (if the continuum hypothesis is true).

There is no reason, and it would be highly unreasonable, to think that a competent speaker of the *number* game would also have to understand what the powerset of the natural numbers is, let alone that he understands why it cannot be counted. So, \aleph_1 might be thought, in this case, to be the *highest* concept, then \aleph_0 would be a lower concept, and the series of natural numbers lower still.

Similarly as the bachelor example showed, one might need to know something about *personhood* to understand what ‘duck’ means and similarly for ‘bachelor’. But, although the two ‘meet in the middle’, the intermediate steps do not allow us to assume that both words have to be understood for someone to be a competent linguistic user of either lexical item.

4.7.2 Logical vocabulary

Earlier I mentioned that *logical* inferences are *not*, on the present view, *materially* correct, and so are not *meaning-constituting* (except, perhaps, in the attenuated sense of constituting the meaning of the *logical* vocabulary). I also noted, both that some logical inferences are *direct*, and that simple non-ampliativeness is not sufficient to pick out logical inferences. I therefore owe the reader an account of what distinguishes logical inferences from other sorts.

Although certain features of logical inferences are shared with materially correct inferences, the vocabulary is *content free*. So, while the rules of the logical language game are still determined by the community,¹⁰ logical

¹⁰In fact, with all the logical systems available these days ranging from classical to,

inferences do not count as materially correct.

Let us briefly consider again the Brandomian account of logicity. As mentioned, Brandom is a logical expressivist. So, the point of logical inferences is ‘making explicit’ the commitments of the members of a linguistic community — such inferences allow us to explicitly ‘keep score’ as deontic agents. (Brandom 2000). All the logical inferences do, then, is *encode* materially correct inferences in order to describe the behaviour of linguistic agents.

One reason that logical vocabulary and material correctness come apart is that logical inferences are non-ampliative in a way that materially correct inferences need not be — this simply means that one cannot derive from the premises anything that was not previously ‘contained’ in the premises. Notice that this is immediately different from the explication of the inference ‘It is raining,’ ‘Therefore, the streets will be wet’. While this is a materially correct inference, it is not logically correct. It is ampliative because it contains information in the conclusion, which it did not in the premises.

But this doesn’t yet distinguish logical inference from our second example: “Bob is a bachelor, therefore Bob is unmarried.” To do this one needs to consider again the ‘two-aspects of meaning’ in conventionalism, for what distinguishes *logical* vocabulary is that these aspects are in perfect harmony.

Two-aspect Theory

Since conventionalism is committed to a two-aspect theory of meaning, so too must be this sort of inferentialism. Thus the two-aspect theory is applicable to materially correct inferences. There must be correct circumstances and consequences of use for any particular linguistic item — i.e., materially e.g., paraconsistent and all the divergences between their rules of inference, say, one might make the case that they are determined by the community *to a higher degree* than other inferences. A justification of that claim will not be attempted in this dissertation.

correct inferences in which the item serves as a key term in a premise and where it features in the conclusion.

Much of our standard logical vocabulary has easily specified assertion conditions and consequences of use, i.e., easily specific classes of *direct* (or, as logicians prefer, *canonical*) inferences. These are given by the *introduction* and *elimination* rules for the bit of vocabulary in question. For instance, the meaning of ‘and’ is given completely by the rule that ‘ $A \wedge B$ ’ is assertible precisely when both A and B are assertible. While from the assertion of $A \wedge B$ commitment to the assertability of A and B follows.

But it is not easy specifiability that makes such concepts *logical*. Rather it is that the two aspect of meaning are in perfect harmony — what can be *extracted* from $A \wedge B$ via the elimination rules is *all* and *only* what is required to warrant an assertion of $A \wedge B$.

With this criterion of *logicality*, it becomes an open question whether, for instance, some of the concepts that play a starring role in so-called “Introduction to Logic” courses are really logical ones. Dummett, for instance, suggests in *The Logical Basis of Metaphysics* (Dummett 1993) that *classical negation* is not a logical concept on the grounds that its introduction and elimination rules are not in harmony.

For present purposes, though, notice the precise sense harmony provides for the claim that logical vocabulary is “content free.” Starting from a class of premises and manipulating them using jointly harmonious rules cannot add anything “new” to the premises with which we began.¹¹

That *ampliative* inferences are not logical is thus immediate. But does *bachelor* become a *logical* concept? Do we infer ‘ x is a bachelor’ from ‘ x is unmarried, adult, male’, and from the former can we get *only* the latter? I

¹¹Why it is so useful nevertheless is a fascinating question, but beside the present point. *The Logical Basis of Metaphysics* is a good place to start investigation of that question.

think not. The entanglements of the concept of bachelor are more complex. It may be stereotype alone that suggests that we can infer Bob's *loneliness* from his bachelorhood, but it is less clear that inferring that he's either *young* or a *sad-sack* or is not *gay* is correct. At least, arguably, the word does not apply indifferently to all unmarried adult men, but rather it involves some sort of derogatory context of 'spinster' as well.

Bringing this together with the two-aspect theory as discussed in connection to materially correct inferences, we have learned that all aspects of language have something *akin* to introduction and elimination rules — namely assertability conditions and warranted inferential assertions. So, the two-aspect theory applies in both cases of materially correct and non-materially correct inferences.

4.8 Materially Correct and Other Inferences

We have seen several important features of molecularist inferentialism during the discussion. However, before wrapping up, it is worth being explicit about what we have shown when taking these features together.

By discussing 'materially correct' inferences in different terms than 'logically correct' inferences, we have cordoned off those inferences from (at least two)¹² other sorts of correct inference.

What this shows is (at least the beginnings) of a molecularist inferentialism. We now have 'materially correct' inferences and 'other' inferences — the former are direct and meaning-constituting; the latter are either *indirect* or essentially content-free and so are not meaning-constituting.

The immediate point of producing such a theory is as a counterexample

¹²I say 'at least two' because it may be possible to cordon off other parts of a language. For our purposes, only showing how one might do this, by invoking materially correct and logically correct inferences is sufficient.

to Fodor and Lepore’s argument — we have established a position which is neither holistic nor atomistic.

4.9 The Linguistic Community and Knowledge of Language

Consider briefly conventionalism once again in order to show how it relates inferentialism: conventions indicate correct usage of (both materially correct and ‘other’) inferences. Using a concept competently in direct inferences (knowing when such inferences are licensed) means that a language user understands the meaning of a concept.

Moreover, we are able to use concepts even if we are not able to explicitly define them (so, again we avoid the analytic/synthetic distinction) — as long as we *use* a concept correctly.¹³ Notice, though, that we needn’t consider *which* conventions are particular to warranting an inference, be it meaning-constituting or nonmeaning-constituting. All that matters for the moment is the recognition that for a community-based theory of meaning conventions (in regards to the correctness of inferences) exist.

The notion of ‘directness’, which largely determines material correctness, likewise rests on community practice. Direct inferences are the ones such that anyone who fails to countenance such an inference (given opportunity for reflection) *is not linguistically competent* in that respect — they do not follow the rules of the language game being played in the community.

¹³And hence maxim that meaning is use, and the conventionalist criterion that correct use is determined by the community are shown to be key.

4.10 Conclusion

The argument presented by Fodor and Lepore is that without the analytic/synthetic distinction molecularism is in trouble — indeed it is unsustainable, and so conventionalism is also in hot water. However, as we have seen, there is a way of drawing a molecularist, non-analytic/synthetic, distinction on inferentialist grounds, that can underpin molecularism.

By using inferentialism we can provide a distinction that is a *weaker* one than the analytic/synthetic distinction, but one which is sufficient to prevent a slide into holism. Meaning now depends on the inferential roles an expression has, but only some of them. Therefore, we do not need to follow Fodor and Lepore in rejecting the first premise of the “basic argument” made explicit by Devitt: ‘*some* of an expression’s inferential properties do constitute its meaning.’ Instead, we do reject the second premise of the “basic argument”: “if some of an expression’s inferential properties constitute its meaning then they all do.” And so, holism is avoided.

“Material correctness,” then, determines meaning-constituting inferences, but there is still a class of ‘other’ (permissible) inferences, not related to meaning — however, they are still governed by the conventions of the community. Now we can see why it is not simply *desirable* for the conventionalist to establish inferential molecularism. Rather, if one is to abide by, e.g., the two-aspect theory, then it seems that there *must* be a distinction between meaning constitutive and nonmeaning constitutive inferences. Thus, the supposedly knock-down argument by Fodor and Lepore is refuted.

Chapter 5

The Paradox of Knowability

5.1 Introduction

The ‘obvious’ knock-down objection considered in this chapter runs as follows:

- (1) If conventionalism is correct, then all truths are known.
- (2) Not all truths are known.
- (3) Therefore, conventionalism fails.

Why would we think that (1) is acceptable? Because it is plausible to think that a conventionalist is committed to an epistemically constrained notion of truth. An epistemically constrained notion of truth is that:

anything which is true is something which could be known to be true.

(DeVidi and Solomon 2001, 319)

A constraint of this sort indicates acceptance of the anti-realist principle $p \rightarrow \Diamond Kp$. Such notions of truth are thought by many to give rise to the paradox of knowability, though — that is, for reasons to be described below, it is thought to follow from the knowability of all truths that all truths are known.

Now, why suppose that conventionalists are committed to knowability?

Knowability in principle relates to the maxim of ‘meaning is use’ — this relates to manifestability of meanings. Manifestability, put simply, is the requirement that one way or another, all aspects of the meanings of all lexical items are learnable. While the reasoning here may not be airtight, it is plausible to think as follows: consider some declarative sentence S . There is nothing about the meaning of S that is not, at least in principle, available to would-be learners of the language in question, which means that, at least in principle, all factors that determine the truth or falsity of S are similarly publicly available. Hence, it seems, S must be *knowable*, at least in principle — though of course for contingent, practical reasons this will often be knowability *only* in principle.

This bit of reasoning suggests that conventionalism about reasoning leads to (or at least tends towards) a certain sort of *anti-realism*. The intuition of many anti-realists is that it is in some sense correct to say that for any truth p , it is possible to know that p is true, which one can represent schematically as $(p \rightarrow \Diamond Kp)$, but they do not accept $(p \rightarrow Kp)$ for every p . $(p \rightarrow \Diamond Kp)$ is often called the verificationist principle, but is more commonly referred to as the anti-realist principle (Williamson 1987), (DeVidi and Solomon 2001). I shall refer to this as *epistemic anti-realism*.¹

The paradox of knowability (also referred to as ‘Fitch’s paradox’), which concerns an argument first presented in ‘A Logical Analysis of Some Value Concepts’ (Fitch 1963), has generated a great deal of philosophical discussion. In that paper Fitch presents an argument that some believe implies

¹Naming it ‘epistemic anti-realism’ is important, since it distinguishes other types of anti-realism, e.g. error theory. The first is important for a semantic theory, the latter is not. For the most part, unless necessary, I will refer to epistemically constrained anti-realism simply as anti-realism. This is the only type which we will be concerned with throughout the chapter.

that epistemic anti-realism reduces to absurdity.

What Fitch seems to have shown is that if one accepts the epistemic anti-realist principle, then one is forced to accept that for all p , if p is true then p is known: $(p \rightarrow Kp)$. Some have taken this to amount to a refutation of epistemically constrained anti-realism notions of truth, such as those that are central to what I have called epistemic anti-realism, since it is obviously false that all truths are known. Others, of course, claim that something is amiss — either with the proof, e.g., that it rests on unwarranted assumption of classical logic, or with its interpretation of the modality \diamond . The potential for multiple diagnoses for why this is not really a refutation of anti-realism, and responses to them, has given rise to a large literature concerning the paradox of knowability.

Thus, conventionalism seems to require a way out of this problem. My approach will be to grant, for present purposes, that conventionalism implies knowability, but *deny* that Fitch’s argument works. In effect, I shall show that there is *at least one* way to save conventionalism from this challenge.

I argue that we should consider so-called “logical revisionist” approaches to the paradox. In more detail, what I will show is that there are several logical revisionist proposals that can be used to block the paradox — but that only *one* — i.e., the application of intuitionistic logic — is the correct revision. The philosophical motivation which prevents the application of intuitionistic logic from being an ad-hoc logical revision is the development of states of information semantics.

While the argument sketched above deriving epistemically constrained truth from conventionalism can be questioned, I am quite happy to conclude that if conventionalism is correct then anti-realism follows (the type that on first glance appears to be susceptible to the paradox at least — that is epistemic anti-realism). Throughout the dissertation we have appealed time

and again to knowability ‘in principle’. The only real and motivated way to make sense of this notion is by appealing to states of information semantics (so a condition of semantic conventionalism is made more specific). This type of semantics leads to an epistemically constrained notion of truth — a state of information at which there is sufficient warrant to establish x is one where an agent *could* come to know that x . But it is not a requirement that *any* individual agent need reach this state (so, of course, the information can equally be inaccessible to the entire community).

Before explicating states of information semantics, I consider a relatively new logical revisionist suggestion for blocking the proof, namely dialetheism (the view that there are true contradictions — hence *reductio ad absurdum* is blocked). The dialethic approach is essentially that, on the basis of the knower paradox, it becomes clear that knowledge contains true epistemic contradictions.² (Beall 2000),³ I will argue, however, that the dialethic strategy is wrong. With the dialethic approach disposed of, I shall turn to what is in some ways the constructive core of this chapter (if the reader will forgive my pun). My first, uncontroversial, claim is that if intuitionistic logic is correct, then the paradox is blocked, so a conventionalist who argues that the correct *logic* is *intuitionistic logic* avoids the paradox. Such a “solution” to a paradox is, of course, open to the charge that it is *ad hoc*. My second claim is that a logically revisionist solution that advocates intuitionistic logic has a suitable philosophical motivation. In making the case for this, I consider several objections found in the literature to suggest that the knowability paradox can be blocked by advocating intuitionist logic, and show that they turn on a confusion of *intuitionistic logic* and *intuitionism*.

²This is not the same as the paradox of knowability. The knower paradox is meant to show that it is possible to know p and $\neg p$ at the same time.

³Explosion is standardly accepted as requirement for any strong system of logic — the rejection of RAA, on the other hand, prevents this rule of inference.

I then show that there are grounds other than intuitionism for thinking *intuitionistic logic* to be correct, in particular acceptance of *states of information semantics*. It is these semantics that allow one to provide a philosophical motivation, consonant with conventionalism, for the claim that intuitionistic logic is correct, and so correct for blocking the paradox.

Of course, this doesn't yet rule out there being any number of *other* logical revisions that might block the paradox and be motivated in some other way. I will try to show that my favoured logical revision is successful and *correct*, unlike others, by considering some rivals. Since paraconsistent logics are becoming increasingly favoured by some authors, I will show that several paraconsistent logic *can* provide counterexamples to the paradox of knowability, but they only provide a motivated solution if states of information are applied.

Naturally I cannot consider all systems of paraconsistent logics, as there are simply too many of them. But I do consider three interesting ones: minimal logic, a non-adjunctive system, and the Routley-Meyers paraconsistent logic.⁴ Paraconsistent logics seem *prima facie* somewhat odd. But, they are gaining momentum in the field of philosophical logic — and hence demand serious attention. The hand-waving of philosophers who are ingrained with the notion that contradictions necessarily lead to explosion must, at this time in philosophy of logic, be questioned. This is not a discussion that I will be able to take up in the dissertation. However, it does indicate *why* I picked paraconsistent systems for my investigation in the first place.

In regards to the three systems of paraconsistent logics I consider, I show that they can provide counterexamples to the paradox of knowability — and in interesting ways.

⁴As with Grice, considering the fact that there is neither time nor space to consider all alternatives, I will treat this as a case study.

Since Fitch’s argument, as we shall see, is essentially a formal proof, one can “solve” the paradox by advocating a logic that rules any particular step in it invalid. As mentioned, my preferred version of this is to advocate intuitionistic logic, but logical revisionism has been used on numerous occasions. A fairly new suggestion in the literature on the paradox is to apply dialetheism or paraconsistent logics — that is, systems which reject the rule of explosion.

The first argument which shall be presented, then, is that there are other systems of logic that, on the face of it, also seem to solve the puzzle of Fitch’s paradox. But, the application of these systems of logic are ad-hoc unless states of information semantic are applied — thus it is intuitionistic logic (and corresponding states of information semantics) are those things that provide a *principled* solution to the paradox. The point of discussing the other systems of logic is, not only to show that they are important supplements to the literature, but also the discussion shows that a conventionalist is *required* to accept my logical revisionism, since the others are ad-hoc, if a logical revisionist solution is to be accepted.

What we shall see, then, is a recurring pattern. If the paraconsistent systems block Fitch’s argument they can be massaged into providing “solutions” to the paradox. But they will be seen to do so not in virtue of some feature naturally associated with the logic in question, but because of an extra feature that is, so to speak, bolted on to the system. The solutions are thus ad hoc. The extra feature that is needed is, as mentioned, that in each case the solutions have a natural home in the states of information semantics of intuitionistic logic.

This, I think, provides evidence for the idea that the states of information semantics applied to intuitionistic logic is not just *a* solution to the paradox for the conventionalist, but *the* solution. Of course, I do not pretend that

this is a conclusive argument. For one thing, I do not intend to survey *all* possible alternative solutions. Nevertheless, I think that the argument has quite a bit of force.

Of course, that the states of information interpretation of intuitionistic logic provides an avenue of escape for the epistemic anti-realist⁵ is important to my project as — at least if my argument is correct — conventionalism is naturally viewed as involving commitment to anti-realism.

What follows in this chapter, then, is an exposition of the paradox of knowability; considerations of alternative logical revisionistic solutions; further, I shall make clear the case that there is a great amount of philosophical value to states of information semantics; on the basis of the latter, I reject several objections which arise in the work of Philip Percival. Thereafter, I will consider several paraconsistent systems of logic. On the basis of these I will show that (1) dialetheism has no force, and is unable to provide a suitable solution and (2) more importantly, those systems of logic which I consider only provide a principled solution precisely because the counterexamples rely on the ‘bolted on’ machinery of states of information semantics.

To sum up: conventionalism seems to imply an epistemically constrained notion of truth (the epistemic anti-realist principle). The knowability paradox seems to show such notions of truth to be absurd. The solution to the paradox is to apply intuitionistic logic, the philosophical motivations of which is the application of states of information.

5.2 The Paradox of Knowability

Fitch’s proof is really quite simple. The argument shows that from the anti-realist principle:

⁵From here on in, I shall use ‘anti-realist’, or ‘anti-realism’ without noting that it is epistemic — that is a notion attached to anti-realism throughout the chapter.

$$(1) p \rightarrow \diamond Kp$$

which means that for any p , if p is true, it is possible to know that p , it follows that:

$$(2) p \rightarrow Kp,$$

i.e that for any p , if p is true, p is known. (2) seems to be “obviously silly”, as Williamson (Williamson 1987, 256) puts it, and most would agree. Fitch’s argument uses a number of seemingly innocuous assumptions:

a) The distributivity of knowledge over conjunction: $K(p \wedge q) \vdash Kp \wedge Kq$

b) Factivity: $\vdash Kp \rightarrow p$

c) The necessitation rule: $\vdash p$, then $\vdash \Box p$

d) The following relationship between \Box and \diamond : $\Box \neg p \rightarrow \neg \diamond p$. In the literature it is often assumed that the correct logic for these modalities is either S4 or S5. But no such strong interpretation is required for Fitch’s proof to go through.⁶

e) Classical propositional logic.

One formulation of the proof is:

⁶Any modal logic which accepts the interchange rule $\Box \neg p \rightarrow \neg \diamond p$ is sufficient for Fitch’s proof. And any basic modal logic satisfies this condition.

- | | | |
|------|---|--|
| (1) | $K(p \wedge \neg Kp)$ | (Assumption for reductio) |
| (2) | $Kp \wedge K\neg Kp$ | (1 dist.) |
| (3) | Kp | (2 \wedge elim.) |
| (4) | $K\neg Kp$ | (2 \wedge elim.) |
| (5) | $\neg Kp$ | (4 factivity) |
| (6) | $\neg K(p \wedge \neg Kp)$ | (1-5 RAA) |
| (7) | $\Box\neg K(p \wedge \neg Kp)$ | (6 necessitation) |
| (8) | $\neg\Diamond K(p \wedge \neg Kp)$ | (from (d)) |
| (9) | $p \wedge \neg Kp \rightarrow \Diamond K(p \wedge \neg Kp)$ | (instance of principle 1) |
| (10) | $\neg(p \wedge \neg Kp)$ | (8, 9 MT) |
| (11) | $\neg p \vee \neg\neg Kp$ | (de Morgan's negation law) |
| (12) | $p \rightarrow \neg\neg Kp$ | (11 \equiv 12 given classical arrow) |
| (13) | $\neg\neg Kp \rightarrow Kp$ | (Instance of DNE) |
| (14) | $p \rightarrow Kp$ | (Hypothetical Syllogism/ Transitivity) |

5.3 The Paraconsistent Solution: Dialetheism

In ‘Fitch’s Proof, Verificationism, and the Knower Paradox’ (Beall 2000) J.C. Beall’s solution is to use *paraconsistent* logic to challenge Fitch’s proof. A logic is paraconsistent if and only if it is non-explosive (i.e., rejects the rule that from a contradiction anything follows). Beall is also a *dialetheist*, which is to say that he thinks that there are true contradictions. It is his contention that, if he can show that there is good evidence for true epistemic contradictions, then it is open to him to block Fitch’s proof at the step from line (5) to (6) — the application of reductio ad absurdum.⁷ So the initial dialetheic solution is to block the inference from a contradiction to

⁷However, it should be mentioned that Beall never *actually* recommends this solution, but just indicates that it is at least open to the anti-realist.

the negation of the assumption used to derive the contradiction, and so the general validity of *reductio ad absurdum* is denied.⁸

In Beall's words the motivation is as follows:

It is important to note, however, that appeal to a paraconsistent logic is not mere ad hocery. Fortunately, for the verificationist, there is independent reason for thinking that knowledge is inconsistent — that is, that a full description of human knowledge includes both Kp and $\neg Kp$, for some p . Such independent evidence of inconsistent knowledge is provided by the famous knower paradox. (Beall 2000, 243) [⁹]

Before discussing the details of Beall's suggested solution, it is worthwhile spending some time discussing what the motivation is for accepting dialetheism for this block to Fitch's proof, as opposed to using a non-dialetheic paraconsistent logic.

Advocates of paraconsistent logics, who are not dialetheists, can accept that contradictions do sometimes arise, for instance in the case of the legal domain, without accepting them as *true*. The paraconsistent point is merely that the domain is not trivialized, i.e. that everything follows, should a contradiction arise; and hence one is still able to make decisions and reason within the domain in question. This is not to say, however, that the inconsistency is, or can be, accepted as a true contradiction.

To use one of Priest's (Priest 1987, 153) examples to illustrate this point: suppose that the constitution of a country contains the following clauses (a) women do not have the right to vote and (b) all property holders have the right to vote. Suppose that at some time both (a) and (b) were the

⁸The question of whether it has some, more limited validity remains open.

⁹Note that the knower paradox is not the same as the paradox of knowability. The knower paradox allows one to derive Kp and $\neg Kp$ from the 'knower sentence': (k) k is not known, and so, like the liar sentence results in a contradiction. The paradox of knowability, on the other hand, allows one to conclude $p \rightarrow Kp$ from $p \rightarrow \Diamond Kp$, as we have already seen.

case. The law at this point is consistent. But suppose further that the country emancipates women to some degree and allows them to be property holders without granting the right to vote. Now the law is contradictory, i.e. inconsistent.¹⁰ Cases of this sort surely could, and presumably, do arise and yet legal reasoning continues nevertheless without *everything* being legal.

This is supposed to be a case where we have a discourse which is inconsistent, but not trivial. But the ‘friend of consistency’ (Priest 1987) will say that eventually the best thing to do would be to resolve the contradiction, i.e. determine which claim is true, since they cannot both be.

Of course, there is a degree of idealization involved here. It may be that the way to resolve a contradiction will not become apparent for quite some time. And it is at least a possibility that resolution will never be achieved. But just because a contradiction is never *in fact* resolved, does not imply that a contradiction is *in principle unresolvable*. Even if a contradiction remains unresolved it does not follow that the contradiction is *true*, but merely that no solution has (as yet) been found.

We can see why an advocate of Beall’s solution needs to be a dialetheist

¹⁰It is important here to point out that it is sometimes not clear what we mean when we talk about inconsistency/contradictions. Are they coextensive? (Restall 2000, 343) says:

Before continuing, we need to make clear that when we say “inconsistent state of affairs” we do not mean a state of affairs which has contradictory properties. . . [I]nconsistent states of affairs may exist (in the same sense as any non-actualised state of affairs might be said to exist). Their inconsistency means simply that they could not be actualised. (Restall 2000, 343)

All but the dialetheist will grant that the presence of a contradiction implies inconsistency in that sense. Even “consistentists” will often deny the converse, i.e. allow that there can be unactualizable states of affairs which do not contain contradictory pairs, i.e. both p and $\neg p$ for some p

and not merely a paraconsistentist. An advocate of paraconsistent logic who does not accept true contradictions would have little or no ground for rejecting the reductio in Fitch’s proof. After all, the solution to the contradiction is obvious in the case of a reductio, namely rejecting the assumption from which it was derived.

Let’s return to the assertion that the knower paradox provides independent evidence for the truth of epistemic contradictions. The knower paradox is in the family of self-referential semantic paradoxes. Consider the statement ‘(*k*) *k* is unknown’.

- 1 If $K(k)$ (where K means ‘is known by someone at some time’), then *k* is *unknown* since the claim must be *true* to be known.
- 2 If $\neg K(k \text{ is unknown})$, then it follows that Kk , since if *k* is unknown then (*k*) *is* known.
- 3 Hence $Kk \leftrightarrow \neg Kk$, which in most logics yields $Kk \wedge \neg Kk$.

Rather than saying that something is wrong with the knower paradox, i.e. that it is in need of clarification and/or resolution, the dialetheist claims that knowledge is fundamentally inconsistent, and the knower paradox is independent evidence for its inconsistency. It is not a paradox after all, the dialetheist will claim, rather it is an accurate representation of a person’s knowledge.

The strength of this argument depends, in part, on the assumed lack of sufficient solutions to the knower and other semantic paradoxes (the liar, the believer,¹¹ etc.). It is beyond the bounds of the chapter to discuss or

¹¹The believer paradox is that an agent can believe two contradictory statements. One of Da Costa’s examples is a wife who places an extra setting at the dinner table for a deceased family member. A contradictory statement might be: “I know that he is dead, but I just don’t believe it.” This is supposed to be an example of self-deception. The more likely explanation is not that the agent does not believe it to be the case that *x*, but that

evaluate all the possible solutions that have been offered, but it is worth pointing out that the burden of proof lies firmly with the dialetheist. As Greg Restall has pointed out, arguments of the following sort are obviously insufficient:

As there is no acceptable solution to the liar and like paradoxes, we ought to accept the argument to the paradoxical conclusion as not only valid, but sound. The liar is both true and not true. (Restall 2000, 342)

Admittedly, Restall is not a dialetheist, and so perhaps he is presenting an unsympathetic reading of the motivation for dialetheism. Consider an alternative explanation: Priest is the arch-advocate of dialetheism, concerning modern motivations for accepting dialethias he says:

Probably the major argument used by modern dialetheists invokes the paradoxes of self-reference, such as the liar paradox...and Russell's paradox^[12]...In the case of each paradox, there appears to be a perfectly sound argument ending in a contradiction; and if the arguments are sound, then dialetheism is true... There is, at any rate, no generally agreed upon solution to many of the paradoxes, particularly those of a semantic... nature. It is these facts that give dialetheism about the paradoxes of self-reference one of its major appeals. (Priest 1998)

But admitting inconsistent states of knowledge, belief, or whatnot, because there is, as yet, no solution is too hasty. The argument for dialetheism needs more than just saying that those hung up on consistency have not yet provided a good answer. What the positive thesis is that supports inconsistent knowledge is not clear, or at least not to me. But perhaps I am, as Priest and Beall would believe, too indoctrinated: perhaps I believe in consistency, not because it is an essential feature of logical systems, but because they wish it weren't so. (Costa 1990)

¹²Russell's paradox is that a set which is not a member of itself is not a member of itself if and only if it is a member of itself.

I have been taught that it is such.

However, this view ignores several *potential* solutions which, although not, as yet, successful, at least suggest a way out of the paradoxes with less cost than dialetheism. Some views which can be found in the literature are (1) appealing to groundedness,¹³ or (2) adapting the hierarchical approach

¹³This is Kripke's solution. He says: "Given a sentence A of \mathcal{L} , let us define A to be grounded if it has a truth value in the smallest fixed point $\mathcal{L}a$; otherwise ungrounded." (Kripke 1975, 71) Andreas Beck explains groundedness (in his Ph.D. thesis — 'The Liar Lies and Snow is White') as:

In Kripke's theory (Kripke 1975), the Liar sentence is neither true nor false because it is ungrounded. This solution is based on the idea that, before it is possible to assign a truth value to sentences of the form "S 'is true' or "S 'is false'," one has to assign a truth value to the sentence 'S'.

Kripke motivates this as follows. Suppose we have to explain the word 'true' to someone who does not yet understand it. According to Kripke, we would start with an object sentence like 'Snow is white' and explain that we are entitled to assert "Snow is white' is true' (or "Snow is white' is false') precisely under the circumstances when we can assert (or deny) the sentence 'Snow is white' itself. By understanding this rule, the person learns inductively to assign truth (or falsehood) to sentences which themselves contain a truth predicate. If a sentence can receive a truth value in this inductive process, then it is a grounded sentence. Otherwise it is an ungrounded sentence. The following example illustrates the difference between grounded and ungrounded sentences.

An example of an ungrounded sentence would be the liar sentence since:

One can assert L whenever one can deny L does not allow the truth value of L to be deduced from the truth value of any object sentence. The Liar is ungrounded, and therefore it is neither true nor false. (Beck n.d.) (Part I)

to knowledge.

Beall claims that such solutions appear to be ad hoc.¹⁴ But, even if these alternatives face objections, that is not to say that they are insurmountable. The burden of proof, then, lies with Beall to show that the responses to the objections against Kripkean groundedness or hierarchical approaches are insufficient.

However, there are bigger problems for Beall's proposal. The following objection is probably the nail in the coffin for dialetheism as a solution to the paradox of knowability. For *even if dialetheism is correct*, accepting self-referential paradoxes as dialetheas does not prevent Fitch's argument.

First, remember that the dialetheist does not necessarily advocate many cases of true contradictions. Most dialetheists only admit dialetheas (true contradictions) in special cases: candidates are (1) vague predicates which can lead to contradictions, e.g., this canvas is red and not red. It is possible to generate this contradiction because the colour continuum is not 'sliced up' into sections in such a way that there is a clear and distinct difference (at some point on the continuum) where 'red'/'not red' become obvious.¹⁵; (2) the liar sentence; (3) and the knower sentence. But these are supposed to be special cases, and in general inconsistency should be restricted to such cases.¹⁶ Even if one were to accept that the knower sentence is a genuine case of a true contradiction, and that it blocks Fitch's argument, the consequences

The charge of ungroundedness applies equally to (*k*), since it is effectively the same (type of) self-referential paradox.

¹⁴But, this is precisely a charge that many may want to accuse the dialetheist of in light of the above quotations.

¹⁵Of course, depending on one's view concerning vague predicates there can be more or fewer of them. However, the cases of interest here are the borderline ones, even for the paraconsistentist.

¹⁶For a more comprehensive account of theories that are inconsistent but not trivial see (Priest 1987, 151-153).

of applying dialetheism to the problem are too strong to make the dialetheic solution desirable, as we shall see.

Recall that the anti-realist principle is supposed to be general, i.e. for any p that is true, it is possible to know p . But Beall's solution is open to the charge that for *any* statement that is unknown there will be a corresponding true contradiction, or else it is not a solution to the paradox at all. For if the rejection of reductio is going to work as a rejection of Fitch's argument then it must do for all true but unknown statements, not just for isolated cases such as the knower sentence (which, recall, is *unknown* as well as being known). Beall is aware of this problem, it seems, as he says:

Grant, as has been argued throughout, that the knower proposition gives everyone reason to admit a world of inconsistent knowledge, a world in which $Kk \wedge \neg Kk$ obtains. The verificationist, with everyone else, admits as much, but the verificationist, unlike everyone else, must admit more — a lot more. Specifically Fitch's Proof shows that the verificationist must admit a world in which $Kq \wedge \neg Kq$ obtains for any actually unknown truth q . (Anonymous referee) (Beall 2000, 246)

To be clear that the objection above holds it is worth briefly running through an example to show that an epistemic contradiction follows from any unknown truth: suppose that you have an arbitrary unknown sentence, call it q . If q is both true and unknown, then it is a counterexample to the scheme $p \rightarrow Kp$. However, if we insert that particular q in place of p throughout Fitch's proof, then we seem to arrive at the conclusion that $q \rightarrow Kq$. Beall, therefore, seems committed to the claim that Kq and $\neg Kq$ are both true for every true but unknown q .

Beall's reponse to this is basically that the referee has presented an excellent argument, one which potentially rules out the dialetheist strategy. And he further says that the knower sentence does not give us reason, even *prima facie*, to suppose that every unknown truth leads to a true epistemic contradiction. In this Beall grants too much if he wants to present dialetheism as

a response to the paradox of knowability (or at least a possible response). Granting that for every unknown truth there is a true epistemic contradiction is doubly problematic: not only do we then have more true contradictions than dialetheist are usually happy with, but our defense against the worry that anti-realism implies that all truths are *known* is to say “yes, they are *and* they are also unknown.”

Let’s raise one more objection, one which has not been discussed in the literature or in Beall’s article. It begins with a little proof: suppose some p is true, but unknown. Given the above objection that for *any* unknown truth there is a corresponding true epistemic contradiction we can state that, for the unknown truth p , $Kp \wedge \neg Kp$. It appears that it is now possible to construct the following argument:

- (1) $Kp \wedge \neg Kp$ (true epistemic contradiction)
- (2) Kp (1 \wedge elim.)
- (3) p (2 factivity)
- (4) $p \vee \neg\neg Kp$ (3 \vee intro.)
- (5) $\neg p \rightarrow \neg\neg Kp$ (classical \rightarrow from (4))
- (6) $\neg\neg Kp \rightarrow Kp$ (DNE)
- (7) $\neg p \rightarrow Kp$ ((5), (6) HS/Transitivity)

The proof ends, then, with the conclusion if p is false, it is known — this is obviously even more ‘silly’ than the conclusion derived from the Fitch style argument.

The question now is: why would it be possible to reach this conclusion? The answer is that Beall makes no commitment to a *particular* system of logic. Simply rejecting reductio ad absurdum, then, is not enough. Given adequate classical machinery we end up with the ridiculous conclusion that there is a p which, if it is false, it is known. Since we cannot know falsehoods, this conclusion cannot be maintained under any circumstances — on the basis of this argument we can now say that dialetheism is entirely *wrong*,

rather than just an unlikely candidate for blocking the paradox. Rejecting reductio ad absurdum is insufficient as a block to the proof since it does not rule out other permissible rules of inference from classical logic.

It appears, then, not only that the dialetheist *should not* accept true epistemic contradictions as a block to the proof, but that they *cannot* simply on the grounds of rejecting RAA. A natural question, though, is whether an alternative solution to the paradox is available to Beall, one with a more substantial philosophical motivation. The brief answer is that there isn't.

In order to block the proof by using dialetheism a different rule of inference other than RAA must be rejected. Either \vee -introduction or disjunctive syllogism must be given up because, in combination, they lead to explosion. The choices, then, are either to give up the classical arrow or \vee -introduction. There seems little motivation to reject \vee -introduction. \vee -introduction merely gets us to line (4) of the proof which says that either p is true or that p is not not known.

The more easily motivated position is the reinterpretation of the arrow operator. In the present context, giving up disjunctive syllogism amounts to not accepting that \rightarrow is the classical conditional. This makes the move from (4) to (5) illegitimate. We will have an opportunity to consider the philosophical virtues of this move later when we consider relevance logic, which similarly rejects the classical arrow operator — this is a move that a dialetheist might make, but which Beall does not commit himself to, and one which has other (e.g., relativist) motives and which is arguably sufficient to block the proof without rejecting reductio.

To conclude, the rejection RAA is not only unmotivated, but insufficient, to provide a block the ridiculous consequences. The dialetheist finds himself on increasingly weak ground.

5.4 States of Information Semantics

The obvious interpretation of the diamond in the anti-realist principle is to explicate it in terms of possible worlds. The motivation for using possible world semantics is that one can say that p is true and that it is possible to know that p , but that, at *this* world, p is unknown. The statement that it is possible to know something, without it being known means that there is some world x , where p is true and, if an agent were to find himself in this world, they would know that p . Since p is unknown in the actual world some story would need to be told about what it means to talk of knowability in counterfactual worlds. Very briefly, in order to make sense of $p \rightarrow \Diamond Kp$, a possible world semanticist will say that: for any true p , there is some world where p is known. We shall see in due course why appealing to possible world semantics is problematic.

It was pointed out at the beginning of this chapter that the intuitionistic logician has a way of avoiding $p \rightarrow \Diamond Kp$ collapsing into $p \rightarrow Kp$. Ignoring the application of the intuitionistically unacceptable de Morgan's negation law,¹⁷ the way to block the proof is by the rejection of double negation elimination. Hence, while $p \rightarrow \neg\neg Kp$ follows in intuitionistic logic from $p \rightarrow \Diamond Kp$, $p \rightarrow Kp$ does not.

So now we are in the position where we can make the case that intuitionistic logic has the *machinery* to block the proof. But is this enough? Not really, since merely employing intuitionistic logic does not mean that the block fails to be ad hoc.¹⁸ Although intuitionistic logic is one natural

¹⁷Alternative versions of the proof can be run such that the de Morgan law can be avoided — hence rejecting the proof on that basis is insufficient.

¹⁸In the literature some authors skip this objection. Since Dummett argued that the correct logical basis for anti-realism is intuitionistic logic, I wonder if that is (at least one of) the reason that the application of intuitionistic logic is immediately taken to be a non ad hoc response to Fitch style arguments.

suggestion for blocking the paradox of knowability, then, it is still necessary to show how this is philosophically motivated.

While I shall be advocating intuitionistic logic as a solution to the paradox, appealing to constructivist semantics is problematic.

While some of the seeming oddities some authors have thought to attend the appeal to intuitionistic logic can be resolved (e.g., by being sure to consistently read \neg as intuitionistic rather than classical negation, so that the equivalence of $\neg p$ and $\neg Kp$ is not problematic) there are some issues that are not resolved. That is, an advocacy of intuitionistic logic runs into problems if it is conflated with constructivism, as I will show that it sometimes is. For example, it hard to explain on constructivist grounds why $p \rightarrow \neg\neg Kp$ is less problematic than $p \rightarrow Kp$.

Conveniently there is another semantic interpretation of intuitionistic logic which can provide a philosophically strong motivation for using this system, namely states of information semantics. In this semantics the connectives are interpreted in terms of states of information. It also gives us an important handle on ‘extensions’ of information. And, crucially, it allows us to explain naturally why an anti-realist should *welcome* the validity of schemes like $p \rightarrow \neg\neg Kp$ while rejecting $p \rightarrow Kp$. It will be convenient to take up these matters in reverse order

5.5 Explication of *S-O-I* Semantics

One of the benefits of states of information is that information is useful in *assessing* the cognitive states of agents and the community in general. Suppose that p is true, but agent a does not know it. Call the state of information where p is true but unknown, *s-o-i* x . Suppose further that there is an agent, agent b , who knows that p is true. This agent must have access to *more* information than a . In other words, b has accessed the

information at x , whereas a has not — therefore agent b has access to, what I call, a ‘richer’ state of information than a .

Further, an important concept for this chapter is that of *extensions* of states of information. We will see in due course that the counterexamples to the paradox rely importantly on extended states of information.

There is an obvious sense in which we can talk about one state of information extending another: suppose that p is true, but unknown. This means that in the described case there will never be any warrant to overturn p (once true, always true). Take a state of information x such that (using an, I hope, obvious bit of notation) $\vDash_x p$, $\vDash_x \Diamond Kp$ and $\not\vDash_x Kp$. What would it mean for $\Diamond Kp$ to be *true* at x ? A natural reading is that there is further information that is *compatible with x* that takes p from being an “open question,” to being *established*. That is, there must be a state of information y that *extends x* (we write $x \leq y$) which is such that $\vDash_y p$ and $\vDash_y Kp$. *S-o-i* y is said to be an extension of x , it is compatible with x which we idealize by supposing it includes x , but also (may) contain extra information. How do we actually make sense of this?

- (1) Suppose that $\vDash_x p$, then $\vDash_x \Diamond Kp$ (since the knowability principle is universal and therefore holds at every point in a model).
- (2) Suppose further that agent a does not know p , but agent b does.

If (1) and (2) are the case, then we can say that agent b has access to *more* information than a , and thus finds himself at a different state of information than a .

This is, of course, a highly idealized account of “extensions of information,” one that readily invites confusions. The following remarks may forestall some of them.

- (1) Different states of information may be concurrent.

Consider the case where a wakes up at seven a.m. on Monday March 26th, 2007 and b wakes up at four a.m. If there happened to be a shooting star at 4:30 and b was suitably placed to see the sky then he would know that “a shooting star was visible from Kitchener-Waterloo at 4:30 on Monday March 26th, 2007” (call this p). Agent a would not know this, however. Agent b , then, has access to *s-o-i* y (both at 4:30 a.m. and at 7 a.m.) and therefore $\models_y Kp$. Since agent a was still asleep at 4 a.m., then at 7 a.m. a does not know that p (and so does not have access to y). Therefore, there is some other state of information x , where a does not know p , so $\not\models_x Kp$. So, a has access to x and b has access to y , where both states of information can be ‘accessed’ at 7 a.m and are therefore distinct in the information that they carry, but are concurrent.¹⁹

- (2) A richer state of information may be *temporally prior* to one that it extends.

It is easy to become confused about the relationship between the information ordering and the temporal ordering. Talk of “acquiring new information” suggests that *richer* implies *later*, which we’ve seen isn’t right. One might still suppose, though, that a richer state of information must either be concurrent (as in example (1)) or that an agent simply missed the information, even though it would have been available if he had performed some action (say getting up at 4 a.m., rather than 7 (as in example (3))). In the latter case the richer state is not temporally prior, i.e., getting up at a time such that a state of information is ‘missed’ is not the same as a temporally earlier

¹⁹Note that in a case such as this we would not say $x \leq y$ unless we were willing to stipulate that *all* the information in x is *also* in y . However, we are using *information* as a sort of success term (*true* information), so even if b doesn’t know everything in x , there is *some* state of information *extending* x that includes what is in y but not in x (i.e., $x \vee y$). The talk of the information available to particular agents can thus be regarded as a helpful shorthand for a more precise story.

state of information, as the example below shows.

Suppose that an agent's grandfather had in 1950 some number n_1 of pairs of shoes in his closet. Call this state of information x . And suppose further, that no one bothered to count how many pairs of shoes there were²⁰ Over time, the shoes wore out and were not replaced with an equal number of pairs.

The information at x that there were n_1 pairs of shoes in the closet in 1950 *was* available (even if they were actually never counted. Now, suppose that the agent's grandfather dies. The information now is lost forever. Let us assume that in 2007 there is some state of information y . Since the information in regards to the agent's grandfather's number of shoes is lost forever, the information *is* not and *never will be* available to anyone. The assumption that richer states of information are either concurrent or temporally later has now been rejected.

So, what this shows is that states of information need not be concerned only with states of information that are temporally concurrent, nor that an agent may simply have 'missed' accessing a state of information due to happenstance. Rather, what it shows is that the temporal order and the information order are distinct.

The fact that x is *before* y and yet $x > y$ implies that there has been 'degradation' of information. The grandson will never find himself in a position to know what the information x contained. But, knowability *in principle* still holds. *Had* the grandson been alive in 1950, he would have been able to count the number of shoes. But, the degradation of information means that he never knows what n_1 was, and to reiterate, this means that the *richer* state of information is temporally earlier. Elsewhere I have summarized the point like this:

²⁰I, myself, for example, have no idea how many pairs of shoes I or my daughter possess at any one time.

A crucial question that confronts any attempt to show that some versions of anti-realism fit this states-of-information scheme is how one makes sense of the accessibility relation. In particular, the “in principle” cases... make clear that we cannot expect the temporal order to match nicely with the order of “extension of information”; at least, we cannot expect it if we hope to avoid the immediate refutation of the idea that truth is epistemically constrained... (Davies and DeVidi 2005, 5)

- (3) Another scenario, though, is this: suppose that no one was suitably placed at the right time to see the shooting star. It does not mean that the existence of the shooting star (is or) was unknowable.

This example does not show the failure of the universality of the anti-realist principle, under the interpretation we assume here. Given an appropriate interpretation of states of information, p is knowable. Anyone *could* have been awake at 4:30 and be looking towards the sky. Had they been so they would have witnessed the shooting star, it just so happened that no one did. There is still some *s-o-i* y , then, where $\vDash_y Kp$. It just so happens that no one gains access to this state of information. It is still the case that there is a state of information where p is not known and a state of information where p is known. Again y (where y could have been accessed, but never was) extends x : $x < y$.

(1)–(3) make clear how the states of information idea yields a natural account of the important notion of knowability *in principle*. Without building in knowability in principle into states of information semantics the universality of the knowability principle is jeopardized. For all p true at x it is or was knowable that p , so there is some state of information extending x at which Kp . But, as noted, this does not imply that for all p , p was or ever will be known — this is the notion of *in principle* knowability we have

available.

So, states of information allows one to maintain the universality of $p \rightarrow \Diamond Kp$ without thereby being committed to $p \rightarrow Kp$. The epistemic anti-realist principle, based on states of information semantics, allows one to say that, given the diamond operator, interpreted as possibly ranging over states of information, if $\models_x p$, then $\models_x \Diamond Kp$, and so there exists a state y such that $x \leq y$ and $\models_y Kp$. So, the universality of $p \rightarrow \Diamond Kp$ is maintained. But since $y \neq x$ is possible, $\not\models_x p \rightarrow Kp$, so $p \rightarrow Kp$, at least sometimes fails to be true.

Given that states of information semantics is the philosophical motivation for taking intuitionistic logic as the logical foundation for anti-realism negation cannot be interpreted classically. The natural interpretation of $x \models \neg q$ in states of information semantics is that x contains information that *rules out* q — that is, there is *no* information compatible with x that would establish q . Thus, in states of information semantics $\neg Kp$ does not mean that p is unknown, but that p is *unknowable*.²¹

As we shall presently see, the states of information semantics I have used in this sketch of a counterexample to Fitch's paradox yields, in a completely natural way, *intuitionistic propositional logic* as the correct logic. This has important consequences in the remainder of the chapter. The remaining counterexamples in paraconsistent logics rely on states of information semantics — it is in this loose sense that they rely on intuitionistic logic (since that is the logic for which states of information is developed). But, minimal logic, Jaśkowski's system, and relevance logic *do not* rely on intuitionistic logic *per se*. Rather, they rely on the extra machinery I add by using states of information semantics in order to develop the counterexamples.²² The

²¹This means, then, that $\neg Kp \rightarrow \neg p$ should be valid.

²²In other words, the counterexamples do not become *intuitionistic*, but require the semantics which provides the philosophical motivation for intuitionistic logic.

order of explanation is this: take, for instance, relevance logic — in order to develop a counterexample in relevance logic we need *more* machinery than relevance logic alone provides — extra machinery that brings states of information into the picture. Therefore states of information semantics is used in all counterexamples, but they happen to be *principled* for intuitionistic logic.

The relationship between states of information and intuitionistic logic is easiest to see by considering Kripke semantics for intuitionistic logic.

The relevant definitions of Kripke frames for propositional intuitionistic logic, then, are:

A Kripke frame $\mathcal{F} = \langle W, R \rangle$ consists of a *set* W of points and a relation R between those points. An *interpretation* \mathcal{I} is a map that assigns a set of $\mathcal{I}(x)$ of atomic sentences to each *s-o-i*/possible world. If xRy , then $\mathcal{I}(x) \subseteq \mathcal{I}(y)$. We define a sentence holding as a *s-o-i* as follows:

- If B is atomic, then B holds at x iff $B \in \mathcal{I}(x)$ (this reflects the idea that the interpretation tells us which atoms are true at each *s-o-i*).
- $P \wedge Q$ holds at x iff P holds at x and Q holds at x (this means that a conjunction holds exactly when each conjunct holds).
- $P \vee Q$ holds at x iff P holds at x or Q holds at x .
- $\neg P$ holds at some *s-o-i* iff there is no *s-o-i* extending²³ x at which P holds, i.e. $\neg \exists y(xRy \wedge p \text{ holds at } y)$
- $P \rightarrow Q$ holds at some *s-o-i* x iff for all y extending x if P holds at y then Q holds at y .(Davies and DeVidi 2005)
- We write $x \models P$ or $\models_x P$ for the claim that P holds at x .

²³'Extension of information' means that if state of information y is an extension of state of information x then it contains all the information contained in x (with whatever other information is added if it is a 'richer' state of information

This formal semantics is well known to validate intuitionistic propositional logic. Comparison of the clauses especially those for atoms, negations and conditionals, with the discussion above will make it no surprise that states of information semantics originates as a heuristic for discussion of these frames.

To conclude, this section has provided philosophic motivations for states of information semantics and therefore a reason for accepting intuitionistic logic as the underlying logic, and the machinery to do so.²⁴ Note that there is an immediate distinction between classical logic and intuitionistic logic that assists in employing states of information in discussion of knowability. When I say that $\neg Kp$ means that p is unknowable, it is because of states of information semantics. To say $x \models \neg Kp$ means, not that p is unknown, but unknowable, given the information in x — that is, there is no state of information $y \supseteq x$ where $y \models \neg Kp$.

Shortly we will be construct counterexamples in different systems of logic which will be shown to be non ad hoc on the basis of states of information semantics. However, as mentioned, we will first consider a more straightforward use of possible world semantics and constructivism as alternatives, and show them to be insufficient.

5.5.1 Possible World Semantics

It is not *immediately* clear why possible world semantics would or should be taken to be problematic. Indeed there do seem to be several benefits to this semantics:

(1) in possible world semantics the diamond in $p \rightarrow \Diamond Kp$ means that

²⁴Some work that has been done in relation to states of information semantics can be found in (Davies and DeVidi 2005), (DeVidi and Solomon 2001), and (DeVidi and Kenyon 2003).

there is some possible world in which it is known (by somebody) that p . Thus, it *appears* that $p \rightarrow \Diamond Kp$ doesn't collapse into $p \rightarrow Kp$.

- (2) one might imagine that one of the virtues of possible world semantics is that, since it is not generally required that if xRy and if $x \leq y$, then all atoms that are true at x are true at y (that is $q(x) \leq q(y)$), we need not assume that information is “persistent,” and so it allows us to take better account of the fact that information can degrade (i.e., things may be forgotten, not witnessed and so nobody — in this world — could know that event x happened). But, suppose that there are two agents, agent a , who lives in world x , and agent b , who lives in world y . a might possess different information than b and vice versa. But, one might say that neither needs to have access to a ‘richer’ state of information than the other. This is the case with *s-o-i* semantics. See note 22). Both a and b *could* gain access to the information state of the other. a gaining access to information from world y and b gaining access to information from world x may increase their knowledge base, but neither world contains more information than the other — just different information. The difference is that this new information may be *incompatible* with what is in x or y , so this needn't be an *extension* — accepting the new information might require rejecting something previously accepted.

A metaphysical reason for rejecting possible world semantics in favour of states of information semantics is the fact that a satisfactory answer concerning their nature has still not been provided: are they real, but nonactual? Are they real and actual (where actual would need to be given an appropriate reading other than the usual sense)? Are they neither real nor actual, but just a useful way of conceptualizing and explaining certain problems?

Possible world semantics *might* be employed to make sense of the paradox of knowability. Someone advocating this type of semantics would need to make clear how one could have two worlds x and y such that $\mathscr{W} \vDash_x p$, $\mathscr{W} \not\vDash_x Kp$, xRy , and $\mathscr{W} \vDash_y Kp$, for some x . How this is to be done is not clear and thus the required clarity is simply not available. States of information, on the other hand, can be made sense of without any heavy metaphysical disputes.

Second, and more importantly, a reason for shifting the discussion from possible world semantics to states of information semantics is that informational states concern *this* world. That is, although sometimes states of information are called ‘points’, this is merely for convenience. The information is in the actual world and the ‘points’ are just a metaphorical way of describing a collection of information. For instance, call a point x , at x there is information that there are mice in the house. There is some further point y , which extends x , where the information is available that there are mice in the house *and* there are bats in the yard. We can use information to capture what is meant by knowability for actual agents in a much more natural sense. The anti-realist principle would likewise be restricted, and it seems natural to interpret the knowability of truths as knowable by actual agents.

But, apart from these considerations there are more important reasons for rejecting possible world semantics.

Even if the above seemingly beneficial aspects of possible world semantics, noted at the beginning of this section could be fleshed out in detail, there is very important problem that is insurmountable for the anti-realist.

Why can the anti-realist who opts for possible world semantics not appeal to intuitionistic logic? Or, why is the semantics not sufficient in and of itself to block Fitch style arguments? It is, indeed, possible to interpret the

diamond in such a way that it necessitates a shift in worlds to go from $\Diamond Kp$ to Kp . *But*, underneath an appropriate interpretation of the diamond, the logic is *classical*. This means that the paradox still follows — the anti-realist principle would, in other words, collapse into $p \rightarrow Kp$. Take even a weak modal logic and a reasonably behaved knowledge operator, then anti-realist principle implies that all truths are known.

So, the paradox of knowability cannot be avoided and hence the anti-realist principle becomes absurd. An anti-realist simply cannot appeal to possible world semantics, because it does not have the machinery necessary to avoid the collapse from every truth being *in principle* knowable and every truth being known.

5.5.2 Kvanvig’s and Percival’s Objections

Intuitionistic logic provides an interpretation of negation and the conditional that diverges from that of classical logic, one under which you cannot move beyond line (11) in the proof given.²⁵ An argument was already presented in a previous section which showed that (1) intuitionistic logic provides a block to the paradox of knowability given the rejection of double negation elimination; (2) states of information semantics, which is particularly suited to intuitionistic logic, provides the philosophical motivation. Using intuitionistic logic is not, however, accepted as a solution by some authors,²⁶ as it is seen as ad hoc. Considering the problems these authors raise will clarify why it is not simply advocacy of intuitionistic logic that solves the paradox, but advocacy of a particular interpretation — an interpretation other than the usual constructivist one.

Line (11) of the proof is $p \rightarrow \neg\neg Kp$, so even if intuitionistic rather than classical logic is correct, $p \rightarrow \neg\neg Kp$ must be accepted by the anti-realist.

²⁵That is, you cannot conclude $p \rightarrow Kp$ from $p \rightarrow \neg\neg Kp$.

²⁶See amongst others (Kvanvig 1995) and (Percival 1991).

Kvanvig, (Kvanvig 1995) however, maintains that $p \rightarrow \neg\neg Kp$ is problematic in and of itself:

[I]t is hard to see what makes the second claim [$P \rightarrow \neg\neg Kp$] so much better than the first [$P \rightarrow Kp$]. At the very least, the intuitionist owes us an explanation about why to affirm the second claim (other than that the knowability paradox forces us to). (Kvanvig 1995, 481-500) from (DeVidi and Solomon 2001, 332-332)

However, DeVidi and Solomon (DeVidi and Solomon 2001) respond to such objections by claiming that not being able to describe $p \rightarrow Kp$ from $p \rightarrow \neg\neg Kp$ is entirely in keeping with anti-realism:

$P \rightarrow \neg\neg Kp$ says: given that P is true, then it's *possible* to come to know that P , in the sense that there is some *possible consistent extension of the current state of information which would establish that KP* . And that ought to be perfectly acceptable for an anti-realist, and indeed, seems to capture nicely what (AR) [the anti-realist principle] was originally formulated to capture. (DeVidi and Solomon 2001, 324)

This is the view that has thus far been developed in terms of states of information semantics.

The fact that it follows from the validity of $p \rightarrow \neg\neg Kp$ that, if $x \models p$, then there is *some* state of information where Kp is forced should be welcomed by the anti-realist, since all truths are in principle knowable. The response may be expressed as follows:

★ $p \rightarrow \neg\neg Kp$ holds at some *s-o-i* x iff for all y extending x if p holds at y then $\neg\neg Kp$ holds at y . In addition, given the interpretation of \neg : $\neg\neg Kp$ holds at some *s-o-i* z iff there is no *s-o-i* extending z at which $\neg Kp$ holds. So, for all $w \geq z \exists t$ such that $t \geq w$ and $t \models Kp$. More colloquially put: if p holds at some *s-o-i* x , then there cannot be a state which extends it at which $\neg Kp$ holds. But that means that for any extension of z , there is an extension where Kp holds. But

an extension of z is an extension of w . Putting things together, the validity of $p \rightarrow \neg\neg Kp$ amounts to: if $x \models p$, there is an extension of x at which Kp , i.e., if $x \models p$ there is some information compatible with x which allows us to know p — something an anti-realist should be happy to endorse. The states of information response shows why Kvanvig’s worry is not a worry at all.

Turning to the Percival objections: he equates constructivism with intuitionism, which in turn is equated with intuitionistic logic. His strategy is to argue that intuitionistic logic allows derivation of consequences from the anti-realist principle that are *not acceptable* on constructivist grounds. For instance:

Since $p \rightarrow Kp$ is plainly false — not every truth is known — Fitch’s proof appears to refute $p \rightarrow \Diamond Kp$. But against this Williamson ... and Rasmussen and Ravnkilde ... object that the propositional connectives involved should be interpreted constructively, that so read Fitch’s reasoning only proves $p \rightarrow \Diamond Kp \vdash p \rightarrow \neg\neg Kp$, and that under this interpretation the consequent of this entailment is true. However, this defence of the knowability principle doesn’t work. Constructive readings of ‘ \rightarrow ’ and ‘ \neg ’ (whatever they are, since p ranges over sentence expressing *empirical* contents), presumably warrant the intuitionistic move from $p \rightarrow \neg\neg Kp$ to $\neg Kp \rightarrow \neg p$. And the latter is false under a constructive reading: grounds for asserting that it is never known that p are not, and can’t be, transformed into grounds for asserting $\neg p$. (Percival 1991, 84)

In the background of this complaint is the assumption that an interpretation of intuitionistic logic must give a *constructive reading* to \rightarrow ; i.e., $p \rightarrow q$ is true when there is a *procedure* for turning an arbitrary proof of p , should one ever be found, into a proof of q . Hence both his reference to the difficulty of extending the interpretation to *empirical* claims (where the notion of *proof* does not apply) *and* his talk about transforming grounds for

one claim into grounds for another.

So, Percival contends that the logical revisionist solution using intuitionistic logic fails because of the unacceptability of:

$$1 \quad \neg Kp \rightarrow \neg p.$$

Consider also this sentence, which follows (1) and factivity ($Kp \rightarrow p$ implies $\neg p \rightarrow \neg Kp$ by an intuitionistically valid form of contraposition):

$$2 \quad \neg Kp \Leftrightarrow \neg p.$$

Percival also views this as highly problematic. An interpretation of the problem Percival puts forward is: suppose that p is a mathematical statement. If this is the case, then p being unknown ($\neg Kp$) is contingent, since whether something is known or not is always a contingent matter. p , on the other hand, is mathematical, and so either true or false *necessarily*. But since (2) *is* provable, we have a contingent and a necessary claim that are provably equivalent.

As noted, above, I claim that the principled intuitionistic solution to the paradox of knowability lies in the application of Kripke frames for propositional intuitionistic logic. A useful piece of evidence for this is that by using such frames the ‘problems’ of using intuitionistic logic quickly dissipate.

Turning to the solution for the first of Percival’s objections, then: The issue with $\neg Kp \rightarrow \neg p$ is, according to Percival, that grounds for not knowing that p cannot be transformed into grounds for asserting that p is false. But, there are several issues here. Percival’s statement assumes we are using constructivism (and the corresponding interpretation of \neg), which we are not forced to do simply because we are using intuitionistic logic.

Consider now the states of information reading: $\neg Kp \rightarrow \neg p$ means that at any state of information x such that $\vDash_x \neg Kp, \vDash_x \neg p$. This means that any x that rules out knowing p also rules out p . This is, in effect, merely to insist

on the anti-realist principle in other words.

So $\neg Kp \rightarrow \neg p$ is not counterintuitive on this reading and should be accepted by the anti-realist.

$\neg Kp \Leftrightarrow \neg p$, holds at some state of information x iff $\neg Kp \rightarrow \neg p$ and $\neg p \rightarrow \neg Kp$ holds at x . This means, first, that given the reading of $\neg Kp$ as “ p is unknowable” in states of information semantics, if p is unknowable, then p is “ruled out — there is no state of information which extends x such that p could be forced at that state. And, second, if $\neg p$ is the case at x , then there is no state of information that extends x at which p is forced — so, there is no state of information which extends x at which p is known. The worry about necessity and contingency being equivalent dissolves because both \neg and $\neg Kp$ work as peculiar sorts of modalities in this setting.

In short, given states of information semantics, this should be commonsensical and something the anti-realist should endorse.

Constructivism was originally intended to account solely for mathematical statements, which made the constructivist-theoretic semantics a natural choice. Giving up the constructivist-theoretic semantics (DeVidi and Solomon 2001, 325) in favour of Kripke semantics for propositional intuitionistic logic (based on states of information), allows one, as shown above, to resolve Percival’s objections.

5.6 Paraconsistent logics

As we have seen, dialetheism is an unsuitable ‘solution’ to the paradox of knowability for a variety of reasons. However, there are many paraconsistent systems that could seemingly block the paradox. The aim in solving a paradox is of course not just to show that one *can* solve it in a particular way but that it *should* be solved in a particular way, i.e. as previously noted, we need a principled solution.

Useful references to more systems of paraconsistent logics are to be found in (Priest 1987), (Priest 1998), (Brady 2003). Note, though, that the authors admit that their discussions are not exhaustive of all paraconsistent systems. It is not entirely clear *what* distinguishes one system from another, e.g., a system may have different axioms from another, yet the same principles are provable, for example. Other systems, by reinterpreting the operators, may have the same axioms but, given the alternate readings of the operators, lead to differences in what is provable in the systems. In seeking a *principled* solution, three paraconsistent systems in particular appear to be promising: Johansson’s minimal logic, Jaśkowski’s non-adjunctive system, and relevance logics. Of course, there are *many* relevance logics — too many to consider them individually, so I will focus on Routley-Meyer’s version of relevance logic.

5.6.1 Minimal — Intuitionistic — Classical Logic

Dag Prawitz, in ‘Ideas and Results in Proof Theory’ (Prawitz 1975) provides a Gentzen-style system of natural deduction²⁷ for Johansson’s Minimal logic. For present purposes, one important feature of minimal logic is that it makes clear that a non-explosive system need not reject reductio. The operators are defined in the familiar way in terms of introduction and elimination rules. Most salient for this discussion is how negation is introduced and eliminated. There is a negation introduction rule:²⁸

$$\frac{[A]}{\perp}$$

²⁷This just means that there are introduction, and corresponding elimination rules for each operator.

²⁸Notice that, as in intuitionistic logic, $\neg A$ iff $A \rightarrow \perp$ ($\neg A$ means that A reduces to absurdity). Therefore, \perp is suitable interpreted as ‘ad falsum’. The difference is that \perp in minimal logic does not imply everything, but merely the negation of the assumption for RAA.

—
 $\neg A$

And a negation elimination rule:

$A \quad \neg A$
—————
 \perp

Here, of course, \perp is a logical absurdity. These in combination with the standard rules governing the other operators $\wedge, \vee, \rightarrow$ result in propositional minimal logic. Notice that minimal logic is not explosive, but that *reductio ad absurdum* is a valid rule of inference — in fact, RAA is important in the definition of the negation operator — \neg -introduction is, essentially, RAA. To be explosive a further rule is added. To incorporate explosion one needs:

\perp
—
 A

Adding explosion to minimal logic rule results in intuitionistic logic. Notice that in neither minimal, nor intuitionistic logic, do any of the accepted rules of inference imply that *reductio ad absurdum* results in the permissibility of double negation elimination.²⁹

And adding classical negation elimination in the form of the rule:

$[\neg A]$
—
 \perp

²⁹It is possible to get confused on this point. It is helpful to distinguish *reductio* — the principle that allows one to conclude $\neg A$ if one derives an absurdity from A — from *indirect proof* — the principle that allows one to conclude A upon deriving an absurdity from $\neg A$. *Reductio* and DNE clearly make indirect proof valid; if $\neg A$ allows proof of an absurdity, then $\neg\neg A$ follows by *reductio* and A by DNE. Conversely, indirect proof validates DNE. Both minimal logic and intuitionistic logic validate *reductio*, but *not* indirect proof.

A

gives us classical logic.

It should be plausible, and it is not hard to prove, that explosion is not derivable from the other rules of minimal logic.³⁰ Minimal logic is, as a result, paraconsistent.

By employing minimal logic it is easy to block the knowability paradox. However, it does not block it by preventing reductio as Beall does. Instead, the key point is that minimal logic does not validate double-negation elimination. It is the rejection of DNE that blocks the proof. The proof is now blocked from lines (11) to (12): that is, we cannot move from $p \rightarrow \neg\neg Kp$ to $p \rightarrow Kp$.

As is well known, DNE *also* fails for intuitionistic logic. Indeed, there is substantial philosophical literature touting intuitionistic logic as the vehicle for a solution of the knowability paradox built around exactly this point. Notice that we are able to conclude that $p \rightarrow \neg\neg Kp$. This is *precisely* the point at which the intuitionist will and should be able to block the proof.³¹

The lesson to be learned from the discussion concerning minimal logic is not that a system of paraconsistent logic has been found that can block the proof, but that it does so in virtue of features it has in common with intuitionistic logic. Let us see how this squares with states of information semantics. Here I shall simply quote myself from above (the section concerning Percival's objections), since he objected to the derivation of $p \rightarrow \neg\neg Kp$

³⁰Minimal logic simply does not have the logical machinery for it to be explosive.

³¹Notice that both minimal and intuitionistic logic will actually block the proof from lines (10)–(11) in my formulation of the proof, since neither accept that the De Morgan negation rule. But since it is possible to reformulate the proof in such a way as to avoid using de Morgan negation laws, it is more appropriate to view the proof as blocked from lines (11)–(12).

in intuitionistic logic. Remember that the problem of DNE was solved in virtue of the states of information semantics as follows:

★ $p \rightarrow \neg\neg Kp$ holds at some *s-o-i* x iff for all y extending x if p holds at y then $\neg\neg Kp$ holds at y . In addition, given the interpretation of \neg : $\neg\neg Kp$ holds at some *s-o-i* z iff there is no *s-o-i* extending z at which $\neg Kp$ holds. So, for all $w \geq z \exists t$ such that $t \geq w$ and $t \models Kp$. More colloquially put: if p holds at some *s-o-i* x , then there cannot be a state which extends it at which $\neg Kp$ holds. But that means that for any extension of z , there is an extension where Kp holds. But an extension of z is an extension of w . Putting things together, the validity of $p \rightarrow \neg\neg Kp$ amounts to: if $x \models p$, there is an extension of x at which Kp , i.e., if $x \models p$ there is some information compatible with x which allows us to know p .

To sum up: minimal logic can block the proof, because it is not strong enough yet to permit double negation elimination. But what is the motivation? Perhaps someone will come up with a minimal-logic based philosophical motivation. For the time being, at least, it seems that states of information are the best way to motivate this. So states of information semantics and machinery shared with intuitionistic logic are what motivate this block.

5.6.2 Jaśkowski's System

One way to understand Jaśkowski's non-adjunctive system logical system (which is somewhat peculiar) is to imagine a discourse with more than one discussant. The contribution of each discussant is self-consistent, but may be inconsistent with the input of others. Anything that a discussant contributes to the discourse is classed as 'holding' (or 'true') in the discourse. And so, conjunction fails 'discursively'.

The reason Jaśkowski's system is paraconsistent is that A might hold in

the model and B might hold in the model, but A and B may be contributions from different discussants. If A and B are contributions from different discussants, then A and B hold in the model, but $A \wedge B$ does not (since, ‘non-adjunctive’ means that \wedge -I is not permitted. In particular, given the consistency requirement, it would need to be the case that, if B is $\neg A$, we *can’t* have (using an obvious notation) $\mathcal{M} \vDash_x A$ and $\mathcal{M} \vDash_x \neg A$, since it would mean that an agent asserts a contradiction. The system is classed as paraconsistent, since it is possible within the system for A to hold at some x , call it x_1 , and $\neg A$ holds at x_2 , $x_1 \neq x_2$, and so both A and $\neg A$ “hold” — but not for the same agent (or the same point) in the system.

Would such a logic provide a solution to the paradox of knowability? Perhaps: so far we have not said enough about the system to say with much confidence whether a counterexample can be developed. So let’s look at a few crucial features of Jaśkowski’s logic. First, notice that the arrow operator in Jaśkowski’s system is not equivalent to the material conditional (this will be made clear when the counterexample is developed).

If there is going to be a philosophical case for the claim that Jaśkowski’s system provides tools for the anti-realist to block the knowability paradox it must at least provide a counterexample to Fitch’s proof:

A case in Jaśkowski’s system which shows the invalidity of Fitch style arguments needs to make it the case that the anti-realist principle holds at all points³² in a Jaśkowski model, but that there is (at least) one case in which the conclusion fails. That is, the anti-realist principle holds universally, but $p \rightarrow Kp$ does not.

A crucial feature of the system is the definition of what it means for a discussant to *assert* something.

³²I interpret a ‘point’ to be that part of the model at which some agent a makes their assertions. So, if agent a asserts A , B , and C , then $\mathcal{M} \vDash_a A$, $\mathcal{M} \vDash_a B$, and $\mathcal{M} \vDash_a C$.

- $\mathcal{M} \vDash_d A$ iff for some w in \mathcal{M} , $\mathcal{M} \vDash_w A$.

We read this as follows: A holds discursively in a model \mathcal{M} iff at some world in the model, w , A is true at w . The worlds of the model are the discussants' discursive contributions. Further features of Jaśkowski's system that must be noted before proceeding to give a counterexample are:

- (a) In this system, the arrow operator is defined as 'discursive implication'. Discursive implication is defined as follows: $A \supset_d B$ iff $MA \supset B$, where M is interpreted as a possibility operator (read as something along the lines of: "if anyone states that A , then B " (Priest 1987) (174)). Note that I shall use \blacklozenge for the diamond in discursive implication in order to avoid confusion with the diamond in the anti-realist principle, while still avoiding the unfamiliar \mathcal{M} for possibility. Further we are here presuming that \blacklozenge is an S5 operator, by assuming that all discussants have "access" to all others, when "anyone" is used in this gloss. This is the way Priest (Priest 1987) interprets \blacklozenge .
- (b) The interpretation of \blacklozenge assumes that everyone has access to the assertions of everyone else.

The counterexample begins with a standard Kripke semantics for modal logic, involving a two-place accessibility relation R . Since there are two modalities at work in the system we need another accessibility relation R_{\blacksquare} . A model now consists of $\langle W, R_{\blacksquare}, R_{\square} \rangle$, where R_{\blacksquare} is the accessibility relation for \blacklozenge and R_{\square} is the accessibility relation for \lozenge . Given that both diamonds are types of possibility, they are both two-place relations.

What would a counterexample be like? As noted, it would have to be such that:

- (1) $\blacklozenge p \supset \lozenge Kp$ holds at every world x , $x \in W$

(2) $\blacklozenge p \supset Kp$ fails at (at least) one world x , $x \in Ws$

So,

(1) $\forall x \in \mathcal{M}, \mathcal{M} \models_x \blacklozenge p \supset \diamond Kp$

(2) $\exists x \in \mathcal{M}, \mathcal{M} \not\models_x \blacklozenge p \supset Kp$

From (1), we have

(3) $\forall x \in \mathcal{M}, \mathcal{M} \models_x \neg \blacklozenge p \vee \diamond Kp$

While (2) gives us

(4) $\exists x \in \mathcal{M}, \mathcal{M} \not\models_x \neg \blacklozenge p \vee Kp$

That is,

(5) $\exists x (\mathcal{M} \not\models_x \neg \blacklozenge p \text{ and } \mathcal{M} \not\models_x Kp)$

Let's refer to such an x as a ,

(6) From (5) $\mathcal{M} \not\models_a \neg \blacklozenge p$ and so from (3) $\mathcal{M} \models_a \diamond Kp$

So

(7) There must be some b such that $aR_{\square}b$ and $\mathcal{M} \models_b Kp$ where $a \neq b$, since we know that $\mathcal{M} \not\models_a Kp$. That is $\exists b (aR_{\square}b \text{ and } \mathcal{M} \models_b Kp)$

We can now see why it was necessary to include distinct accessibility relations for \diamond and \blacklozenge . For, since $\mathcal{M} \models_b Kp$, we have $\mathcal{M} \models_a \blacklozenge Kp$. But, as noted, $\mathcal{M} \not\models_a \diamond Kp$. So the accessibility relation between points for \diamond from the one for \blacklozenge must differ. In particular the accessibility relation for \diamond cannot be symmetrical.

So, we must *expand* the semantics by adding a new accessibility relation for \diamond . And it needs to be a relation that accounts for the fact that where there is an 'unknown truth', i.e. some p such that for some w , $\mathcal{M} \not\models_w p \rightarrow Kp$,

there is some *other* discussant 2 such that ${}_wR_2$ and $\mathcal{M} \models_2 Kp$. In any conversation, then, if p is ‘true’, but someone does not know it, *someone else* suitably related to discussant 1 *does* know that p . The question now is how to motivate *that* assumption.

This is where Jaśkowski’s system ceases to be of help to the anti-realist, as it does not provide a *principled* solution to the knowability paradox. In order to develop a philosophically motivated solution I will appeal again to *states of information*. I will argue that the only way to understand the difference in the cognitive states of the agents is to build in a notion of *extensions* of information. That is, it must be the case that discussant 2 has access to more information than discussant 1. So discussant 2 is in the position to know that p , but discussant 1 is not.

How, then, might a principled restriction be imposed? The question to start with is how one could motivate $\Diamond Kp$ being true at *one* point, with Kp holding at some other. A natural reading is to take the relation between the points as one of *extended information* (new discoveries and the like).³³ Such an interpretation of R_{\Box} seems to require that it be *reflexive* and *transitive*, but not *symmetric*. Thus it is a pre-ordering rather than an equivalence relation — otherwise Kp would be accessible to everyone.

If R_{\Box} is interpreted as a pre-ordering, then $p \rightarrow \Diamond Kp$ can hold at each world, but $p \rightarrow Kp$ fails at at least one world in the model.

The question now is whether we can make the counterexample compatible with the discursive semantics. It is, indeed, quite possible to construct a counterexample, but as noted earlier *not* merely by using the machinery of Jaśkowski’s system. Traditionally Jaśkowski’s system interprets the discursive

³³This is why states of information, or points, are here used interchangeably with, what Jaśkowski calls possible worlds. Extensions of information is a much more natural notion for this counterexample. For a discussion concerning this, see amongst others (DeVidi and Solomon 2005)

sive frames as ‘full-frames’, and so the accessibility relation relevant for the system must be symmetric, transitive, and reflexive.

But, by appealing to states of information semantics we see that a non-symmetric accessibility relation is necessary. Although all participants in the ‘discussion’ may have access to all of the assertions of the each other, it cannot be the case that they have access to equivalent *information*. This cannot be the case because, as made clear earlier, it is quite possible for $\mathcal{M} \models A$ and $\mathcal{M} \models \neg A$ to hold in the same frame, but then it must be the case that $\mathcal{M} \models_x A$ and $\mathcal{M} \models_y \neg A$, where $x \neq y$. The reason for this, to restate the point, is that otherwise a contradiction would be derivable at a single point in the system — something Jaśkowski’s system does not permit.

Privileging states of information means that the interpretation of ‘what holds’ in Jaśkowski’s system must account for an agent having more or less information available to them.

As with minimal logic, it seems that *this* solution is not peculiar to discursive logic. Once again, given that the notion of states of information semantics is explicated in terms of Kripke semantics for propositional intuitionistic logic, the solution is due to virtues of that semantics, not discursive logic in particular. $p \rightarrow Kp$ is *intuitionistically* true at w_1 if and only if for all w_x such that $w_1 R w_x$, if $\models w_1 p$, then $\models w_x Kp$, where R is a *pre-order* (i.e. reflexive and transitive), and so an agent at point w_1 does not have access to the information at w_x . Both p and $\Diamond Kp$ are now true at w_1 , Kp is true at some later state of information, call it w_2 , such that $w_1 \neq w_2$ and $w_1 R w_2$. So, now we have a counterexample. But we get it only by appending machinery we have seen to be naturally associated with intuitionistic logic; the philosophical motivation is similarly also not specifically Jaśkowskian, but due to states of information. Indeed, without the pre-ordering restriction of the relation between points, I do not see a way of getting a counterexample

to the proof at all.

So, in order to build a counterexample in this system, with suitable extra logical machinery ‘bolted on’ we *can* get a counterexample to the paradox. But, the extra machinery does not share motivations with Jaśkowski’s system. And it is precisely the machinery of intuitionistic logic which we saw provides the motivated solution.

5.6.3 Relevance Logic

In order to build a counterexample to Fitch’s proof in relevance logic we shall consider the standard account of the arrow operator given by Routley and Meyers (Priest 1987), which is:

$A \rightarrow B$ is true at a world $a \iff$ for all worlds b and c such that $Rabc$ either A is false at b or B is true at c .

The counterexample in relevance logic, shall be based, for similar reasons as the Jaśkowski counterexample was, on standard Kripke semantics for relevance logics. The relevant part of a Kripke frame consists of a set of worlds and a three-place accessibility relation among the worlds: $\langle W, R \rangle$. However, given that the anti-realist principle involves the modality \diamond , the definition of an appropriate frame expands to $\langle W, R, R_{\square} \rangle$. Here again, W is the set of worlds, R is the accessibility relation between worlds for the arrow operator, and R_{\square} is the binary accessibility relation for \diamond .

A counterexample will be an interpretation such that for all worlds x , $x \models p \rightarrow \diamond Kp$ and at the particular world a , $a \not\models p \rightarrow Kp$. So, it will effectively show that for all worlds the anti-realist principle holds, but that at some world a , $a \not\models p \rightarrow Kp$. Hence:

$$(1) \quad a \models p \rightarrow \diamond Kp$$

$$(2) \quad a \not\models p \rightarrow Kp$$

So from (1),

$$(3) \forall b, c \in W, Rabc \Rightarrow (b \not\models p \text{ or } c \models \Diamond Kp)$$

While from (2),

$$(4) \exists d, e (Rade \text{ and } (d \models p \text{ and } e \not\models Kp))$$

Let us look more closely at d and e . Since $Rade$ from (3) we have,

$$(5) d \not\models p \text{ or } e \models \Diamond Kp$$

Thus, since $d \models p$ by (4),

$$(6) \exists f (eR_{\Box}f \text{ and } f \models Kp)$$

But, $e \neq f$ because we already know that $e \not\models Kp$ from (4). This shows that it is possible to construct a counterexample to the knowability paradox, since the anti-realist principle holds at a (in fact at *all* worlds), but $p \rightarrow Kp$ does not hold at a .

Again, we must appeal to states of information semantics to make sense of the relevance logic counterexample: R_{\Box} is not an accessibility relation required by relevance logic. Appending it to the usual relevance logic semantics allows us to force it to be the case that Kp is false at e , but true at f and $eR_{\Box}f$.

What kind of relation would R_{\Box} be, if it's to be such that whenever p is true, but unknown at e , there is an R_{\Box} -related f at which p is known? Again, it seems to me, we must appeal to states of information semantics. e is an earlier (less rich) state of information than f , so $e \not\models Kp$ and $f \models Kp$. So, again, a solution is possible, but the motivations that block the proof do not come from relevance logic. Instead the motives are the same as the intuitionistic logic account.

So, $p \rightarrow \Diamond Kp$ holds at *every* s-o-i and Kp hold at some 'later' or 'richer' state of information. Whether or not an agent actually ever reaches this

state of information where they gain the information that p is true is not the issue, only that they *could*. This is what is essential to the counterexample.³⁴

To conclude, if the anti-realist principle is to make sense at all in terms of relevance logic, then it must be developed in terms of *s-o-i* semantics. The counterexample, then, is similar to the one developed for intuitionistic logic, minimal logic, and Jaśkowski's non-adjunctive system .

5.7 Conclusion

Returning to the investigation as to whether paraconsistency is a suitable method to block Fitch style arguments the following main lessons have been learned. First, according to J.C. Beall (Beall 2000) a dialetheist *can* use true epistemic contradictions as a solution to the paradox, but remains neutral as to whether they *should*. We have seen reasons to question even this limited optimism. The objections which I raised against dialetheism shows that a dialetheist, who does not adhere to a particular system of logic, but merely the rejection of RAA, *cannot* give a solution to the paradox at all — the result was shown to be wrong (not just questionable) on the basis of the last objection.

Second, the three subsequent systems of paraconsistent logics discussed can provide interesting blocks to the proof — all given in terms of states of information semantics for propositional intuitionistic logic.

Therefore, as shown above, there is nothing in paraconsistency *itself* that offers a motivated, non ad hoc, solution to the paradox of knowability.

³⁴What precisely accessibility *in principle* is is not fully discussed at this point. Although I have begun work in states of information semantics, my work is not complete and the story I have to tell thus far is based on the bare bones of a research project I intend to undertake in the future. Therefore a full discussion of the *in principle* notion of states of information semantics would take us too far afield of the current discussion.

Rather the motivation comes from the features that must be added to the system, that are part of the solution using intuitionistic logic. Those familiar with the literature, will find in this endorsement of intuitionistic logic as the logic for anti-realism an echo of much work done by anti-realists about particular domains of discourse.

The overall conclusion, then, is that intuitionistic logic is the system which should be used to block the proof³⁵ — and from the systems we have discussed is the only one which provides a non ad hoc solution. And, moreover, since intuitionistic logic is the logic that validates states of information semantics, it seems to be the correct system of conventionalism.

The Kripke semantics, laid out above, also shows that we can invoke a notion of knowability faithful to the externalist picture endorsed throughout the dissertation. It need not be the case that any one agent is in the position to know p , just that if p is true it is *possible* to come to know that p . Relating this to community-based semantics we see that this is highly desirable. If any one in the community has access to the state of information in which that agent knows p , and that agent is able to point out to others that p is or was the case.

But, it is also possible that no one in the community comes to know that p . Just because p is true, it need not be the case that p is known, even if it is *knowable*. This is consistent with the notion, raised in chapter one, that knowability for conventionalist semantics needs to be understood as knowability in *principle*, not actuality. Thus use-based semantics of conventionalism is a defensible and, indeed desirable, position, and one which is fundamentally tied to intuitionistic logic and states of information semantics.

³⁵Remember that intuitionistic logic provides the machinery necessary, and states of information the philosophical motivation.

Chapter 6

Conventionalism

6.1 Introduction

I have been calling the theory presented in this dissertation neo-Dummettian. In some respects I have remained close to the spirit of the original Dummettian stance, for instance by incorporating the overarching maxims. (Dummett 1973), (Dummett 1993). However, in the course of the investigations in the first five chapters we have seen various ways in which Dummett's views need clarification, supplementation, or modification. For instance, in Chapter Three we saw that de-fanging Williamson's objection required that we reject Dummett's commitment to luminosity, and that we could do so by giving closer scrutiny to what manifestability involves — a process that revealed the need to supplement Dummett's view with a commitment to a division of linguistic labour.

Division of linguistic labour is an essential feature of the externality of language. As I showed in Chapter Three one can be, with respect to a bit of language, a competent linguistic agent (the robin example); an incompetent linguistic agent (the toadflax example); or an agent can with a fairly colloquial competence, sufficient to know how to use the term appropriately

in *certain* circumstances, but with knowledge that is incomplete. But in all cases I can achieve some useful things using the words in question.

In the first case I can *use* the term appropriately (unless referring to Australian robins). In the second example, I cannot use the term at all, except via appeal to a suitable authority. The third example, regarding knowledge of the meaning of ‘geophysics’, is that I can apply limited knowledge in certain circumstance, but an appropriate authority can supplement the knowledge so that I can *use* the lexical item in more circumstances. The point is not that ‘geophysics’ has a limited meaning in my mouth and a fuller one in theirs. Rather, it is that they can teach me more about what the words means when *I* use it. Knowledge of meaning is inherently bound up with correct usages of lexical items, and hence with competence or incompetence as a linguistic agent. This is consistent with the Wittgensteinian maxims — which are basically what it means to be a conventionalist.

In sound bite form, as described in Chapter One, the major features of conventionalism are that:

- (1) meaning is intersubjective.
- (2) meaning is public (meaning does not reside in the head).
- (3) meaning is use-based (competent speakers must grasp the rules of the language game to be able to use lexical items appropriately). It is the patterns of behaviour of the linguistic community which determines whether the rules of correct use are being followed.
- (4) meaning is two-aspect in nature. That is, the meaning of, for instance, a declarative sentence, is determined *both* by the conditions under which an agent would be warranted in asserting it, *and* by the things an agent is warranted in inferring upon hearing it asserted. In short, the aspects are the assertion conditions and the (immediate)

consequences.

- (6) meaning is manifestable. There are no aspects of the meaning of a lexical item that are private, nor are there any that are beyond our ken — all aspects of meaning can be made “available” to any linguistic agent.

Before giving a more extensive explanation, note that, any decent or plausible theory of meaning must, according to Dummett, be closely tied to the notion of Fregean thoughts. But it’s important not to confuse Fregean thoughts with other things we might be tempted to call ‘thoughts’.

For Frege, thoughts — the contents of acts of thinking — are not constituents of the stream of consciousness. . . The reason is that thoughts are objective. . . On Frege’s view, thoughts and their constituent sense form a ‘third realm’ of timeless and immutable entities which do not depend for their existence on being grasped or expressed. (Dummett 1993, 22-23)

There is an important distinction between the role of thoughts in the philosophies of Frege and Dummett. Frege’s view was that thoughts are explanatorily primary: they are the meaning of sentences and understanding a sentence is a matter of grasping a thought — and so being in a special relationship with something in the platonic realm. Dummett reverses the Fregean view. Sentences have meaning, but there is no abstract third realm where thoughts reside. Thus *sentences* and the rules governing their use are explanatorily primary and *meanings* that are secondary. Dummett’s reasons for this inversion are, primarily, because one cannot gain access to the third realm. Some sort of bridge, as one might call it, is needed. Dummett turns to intersubjectivity, rather than objectivity.

6.2 Conventionalism — Motivations

It was mentioned in Chapter Two that Gricean semantics and Dummettian semantic conventionalism share (at least) two motivations. Both the third realm and behaviourism are rejected. However, whilst Grice opts for intention-based semantics, Dummett opts for conventionalism.

As noted in the introductory chapter, the main thing I set out to do, that is *my* motivation for investigating semantic conventionalism is that it seems to me a viable theory, even in light of semantic theories which are thought by some to be obviously stronger *and* in the face of potentially serious, knock-down, objections — a point of view not shared by many. Thus I needed to respond to two issues, namely: (1) the idea that semantic conventionalism is an obviously weaker theory than others and (2) that various knock-down arguments can be raised — thus showing conventionalism to be dead in the water.

Tackling (1), by using Gricean semantics as a case study, I have shown that semantic conventionalism is not obviously weaker than this very popular approach to semantics. Indeed, in light of objections raised in the chapter regarding intention-based semantics, it seems that, at least in typical cases, the conventions of linguistic meaning must exist prior to the intentions a linguistic agent has when making an utterance. One is only able to gauge the intentions of the speaker *on the basis of* the shared understanding of the language. Hence conventionalism is in a *better* position than intention-based theories, at least if these are supposed to be *semantic* theories.

Tackling (2) involved showing that the knock-down arguments were not knock-down after all. In these cases, though, the reasoning *not* involved in refuting the arguments has told us more about what conventionalism must be like if it is to be defensible. The main point of this concluding chapter is to sum up these lessons we have learned along the way.

Most of these lessons can be profitably described as clarifications of one sort or another of the six criteria that I take to be fundamental to Dummettian motivated conventionalism.

6.3 Conventionalist Conditions

6.3.1 Meaning is intersubjective.

Meaning is something that arises from the use of lexical items by the linguistic community. Note, though, that it is conceivable that at some time there might be *no person in the community* who has a grasp of the meaning of a term and yet the term has *a meaning*. That is, suppose that all theoretical physicists were to disappear, so everyone with a clear grasp of the use of the term ‘quark’ would be lost to the community. There are two options for dealing with such scenarios: (1) we could say that in this case the term ‘quark’ ceases to have meaning and is taken out of the linguistic repertoire. (2) we could say that the meaning of the term ‘quark’ can be taken to be *in principle knowable*. Supposing that those in know about the meaning of the term disappear completely is fine. One must not ignore the fact that this *could* have been otherwise, nor that there may come a time when, e.g., someone discovered the manuscripts of the physicists and relearns what ‘quark’ means. Whether (1) or (2) is correct would seem to depend on what information is left behind that would allow those left behind to learn the proper use of the term.

In Chapter Five states of information semantics was developed as a philosophical motivation for the application of intuitionistic logic. As that chapter made clear, the states of information framework is necessary and philosophically well motivated for anti-realism — which in turn, at least arguably, follows from conventionalism. We needn’t rehash the whole story here. The

present point is merely that the meanings of lexical items might be forgotten, even by the entire linguistic community. What is right to say in such a case depends on whether they are in a some state of information which contains only enough information that the word must be incorporated (again) into the stock of meaningful words (in which case *then* the community has access to the lexical item) or the information continues to exist *in the community* (e.g., in textbooks), even though not in any existing person’s mind and so “the community” is in a state of information that determines the meaning of ‘quark’ nonetheless. Basing meaning on community communicative practices allows us to make sense of the fact that meanings may be forgotten, regained, revised, or what not.

6.3.2 Meaning is use-based.

Conceiving of meaning as use-based implies that being a competent speaker of a language depends on the way in which one uses the language. A speaker must always follow the rules of the language to be a competent speaker. He must know how to *use* the language — and so conform to the conditions which apply to asserting, inferring, etc. The question that follows, of course, is how sentences (or words) are actually meaningful in virtue of use. Without such an account conventionalism can hardly be said to be a semantic theory.

An important aspect of conventionalism is that words have the meaning that they do in virtue of the role(s) that they play in sentences — accepts Frege’s ‘context principle’. But notice, that this is a feature shared by holism.

But, arguably, that meaning is determined by use follows from the fact that sentences have meaning in virtue of how they are employed by the linguistic community — only a use-based theory allows the community to determine meaning, and so allows for a theory that avoids the perils of psychologism and platonism.

Let us consider an example of a sentence that would *not* be accepted by the linguistic community — “Quadruplicity drinks procrastination,” (this is one of Russell’s examples) — the sentence has a perfectly regular form of subject, verb, object. But the sentence does not make sense in the English language because the word ‘quadruplicity’ has nothing to do with ‘procrastination’. ‘Quadruplicity’ means “fourfold nature; the condition of being fourfold, or of forming a set of four.” (*Oxford English Dictionary* n.d.) ‘Procrastination’ is used to refer to the common practice where important activities are put off, not, e.g., fourfold nature. There is, in short, no connection between the two words ‘quadruplicity’ and ‘procrastination’ and they are never used together when connected by the verb ‘to drink’.

The conventionalist is in a good position to offer a satisfying explanation for the sentence’s nonsense status — the sentence is *useless* in the sense that there are no conditions under which it is assertable, nor are there any sensible consequences of it, considered as a claim.

6.3.3 Two-aspect theory.

To call conventionalism a “two-aspect” theory simply means that the meaning of a sentence is determined both by its assertion conditions and from its consequences.

Concerning the conditions of use consider an assertion and the situation in which it is warrantably assertable. If I am in a position to warrantably assert “There is water in the bottle,” I should have some evidence for making the utterance (e.g., it is in a water bottle, it is a clear liquid, etc.). If I am correct then my interlocutor and I are put into position to draw particular conclusions, and/or act in a particular way, and draw further inferences (for instance going on to say “If I am thirsty, I shall take a drink from that bottle,” and then carrying through with the corresponding actions).

On the other hand, if it turns out that some person has switched the water for gin, and I find out that the bottle contains gin and not water, then I should retract my initial assertion. Although in such cases the utterer must retract his assertion, this need not mean that the circumstances of making the assertion were insufficient. But finding out that the assertion was wrong means that the consequences change. Different inferences are now licensed which previously weren't. And the previously licensed inferences fall by the wayside.

So much is common currency for conventionalist views. What we have learned about this in the thesis is two-fold: (1) if conventionalism is to avoid a disastrous collapse into holism, it must involve putting constraints on *which* assertion conditions and which consequences contribute to meaning and which do not.¹ (2) One way to try to make this out in terms of *direct* but *non-logical* inferences — what I call in Chapter Four, borrowing and modifying a useful term from Brandom, the *materially correct* inferences.

6.3.4 Conventions are arbitrary.

The arbitrariness of conventions in regards to the meanings of lexical items follows immediately from the fact that conventions are constituted by intersubjective rules, which are determined by the behaviour in place in a linguistic community. There is nothing about the word 'cup' for example that means that it designates cups. Drinking vessels of that particular type could easily have been termed 'schmuck'; or have had no word attaching particularly to them at all.

This does not mean, however, that conventions are particularly easy to update. Often the adoption of a new convention is a long and slow process,

¹There is an anticipation of just this point in Dummett's discussion in *The Logical Basis of Metaphysics* (Dummett 1991) of *canonical* verification conditions and consequences, though he raises the matter when discussing logical vocabulary.

since the entire linguistic community — or a relevant class of “experts” — must come to accept the change. Examples to the contrary are, of course, available, such as the adoption of the word ‘Google’ almost overnight — such an example shows that in the linguistic community change of conventions is possible (sometimes at a rapid pace), since once a new lexical item is accepted by the community a whole new set of assertions becomes possible. But generally speaking the conventional semantic theory is not open to objections that the conventions are unfixed/easy to change/changeable on a whim, or some such thing.

What I am currently interested in is a different form of arbitrariness of convention, which indicates that the conventions are stronger and more ingrained in the community than might be thought. Conventionalism, then, means more than accepting the arbitrary nature of word meaning. The possibility of change in the meaning of lexical items, or having had the lexical items be otherwise, is a form of conventionalism that (almost) everyone would agree to. For instance, no one would argue with the claim that the certain lexical items of a language is just so by accident. ‘Water’ could have been ‘shwhater’. It is basically a truism to say that sentences and words could have been expressed otherwise — we might all have spoken French, after all. It is worth pausing to consider whether language is *conventional* in some richer sense.

Poincaré

As pointed out in Chapter One, there are alternative forms of conventionalism other than those directly related to a semantic theory. Although they concern a different field, that is not to say that we cannot use them as a guide to the more ‘robust’ conventionalism mentioned above. After discussing geometric conventionalism, I will tie it in with semantic conventionalism.

One sort of conventionalism that has its (perhaps benighted) defenders is conventionalism about (e.g., scientific) truths, nicely summarized by Lawrence Sklar:

Let us agree that we can offer a variety of alternative hypotheses about the material world, all equally compatible with the possible totalities of sensory data. . . . To accept one of these hypotheses is as rational as to accept another. They are all “equally true.” What we need is a correct understanding of *truth*, realizing that it is a matter of a convention; but since conventional truth in these matters is the best obtainable, there is no reason to disparage it because of its conventionality. . . . (Sklar 1977, 123)

This is not quite the sort of conventionalism we want, for it is not *empirical truth* we want to have be conventional, but meanings. Let us turn, then, to a slightly different kind of conventionalism, also familiar to philosophers of science. Famously Poincaré was a conventionalist concerning geometry, as Carnap (Pinco n.d.), Murzi (Murzi 2006), Sklar (Sklar 1977), and I am sure others, have noted. As Sklar states:

Poincaré thought that we would stick to the Euclidean theory, come what empirical data may. But this wasn't because of an ineluctable psychological hold the theory had on us, but because of a self-conscious choice of the descriptively simpler alternative. . . . (Sklar 1977, 128)

Poincaré notes that there are several geometries and argues that each is compatible with all possible observations — we may need to postulate a different selection of basic forces, for instance, if we assume space is Euclidean from those needed to describe the phenomena if we assume a curved space, but anything that can be done with one assumption about space can be done with the other. There is, then, no fact of the matter as to which ‘truly’ describes space. Which geometry is assumed is a matter of convention — the point is that we must choose *some* geometry to do physics — geometry, then, becomes a *framework*. Which framework we use is not a forced choice,

but rather a pragmatic decision — the point is that geometry is not in the business of describing reality. It *is* just a framework. And, as Sklar notes, if Poincaré thought that our pragmatic grounds would always lead us to do physics in 3-D Euclidean space, a claim, ironically, made just a couple of years before Einsteinian physics moved to centre stage we should be conventionalists about geometry — we can decide which geometry to accept. Perhaps one might even choose one geometry for one purpose and a different geometry for another.

Poincaré’s geometric conventionalism is explicitly cited in *The Logical Syntax of Language* (Carnap 1934), as a source of his own conventionalism about linguistic frameworks, as formulated in his famous *Principle of Tolerance*.

The question now is what geometrical conventionalism squares can tell us about semantic conventionalism. What is “arbitrary” in the case of geometry is the choice of framework — this does not mean, however that everything is conventional — the empirical data which must be accounted for is framework independent.

Linguistic conventionalism assigns rules of language a status in many ways similar to that assigned to geometry by Poincaré. Poincaré did not think (notoriously) that there was any *real* prospect of the scientific community opting for non-Euclidean geometry; his point was just that, in principle, they could. Similarly, to think that our linguistic practices could be radically different, that the rules are *arbitrary*, does not mean that it would be easy or likely that we switch, nor that we were aware of making a choice when developing the rules in the first place. It means that there are alternative ways language might work that would let us get done the things we do with language.

Moreover Poincaré’s view is not simply that any old geometry will do.

It is that various geometries will all let us do something in particular — formulate adequate physical theories. The “choice” is constrained by the goal, and many possible geometries (e.g., one-dimensional geometries, to take a silly example) simply will not do. Similarly, I think, conventionalism about language needs to be understood as the claim that our linguistic practices could be organized in quite different ways and yet we would be able to accomplish the things we do accomplish with language.

6.3.5 Manifestability

For the conventionalist all meaning must be manifestable and externalized. Manifestability, however, does not (as was shown in Chapter Three) imply luminosity. There is little room in the semantic conventionalist account of meaning for *internal* psychological states to play a role. Dummett, for instance, allows the intentions of a speaker to play a role in determining meaning in a context:

When there is undeniable ambiguity, produced by there being two distinct conventional uses of the linguistic form, what determines the force attached to the utterance is how the speaker intends it to be understood: this intention selects between two existing linguistic practices but creates neither of them. (Dummett 1993, 119)

Note, though, that intentions have a role only *after* the possible meanings are fixed in another way.

Again, meaning can only be determined by looking at what happens ‘outside the head’ — in other words, the behaviour of the community. Dummettian, or neo-Dummettian, conventionalism, after all, requires thoughts to be extruded from the intentionality of agents — intentions, in short, cease to matter, except, as noted above, if disambiguation is needed.

It is the nature of the manifestability criterion about which we have learned the most in this thesis — but I have already described this once in

this concluding chapter, and so will not repeat the lessons here

6.4 Conclusion

To conclude, the six criteria of semantic conventionalism are that meaning is (1) intersubjective; (2) public; (3) use-based; (4) two-aspect in nature; (5) manifestable; (6) and the conventions are arbitrary.

These are all features that were drawn out in Chapters Two to Five, and some of them considerably clarified. It is interesting to consider these criteria together in the conclusion for a couple of reasons. The first is that it shows that conventionalism has not simply been protected from several potentially devastating objections and the charge that it is an obviously lesser theory — rather, in the process we have seen the emergence of a positive and viable theory.

Second, it has now become explicit that there is quite a bit of overlap between the criteria — this is only to be expected, however, since at the base lie the two maxims of meaning being determined by use and that to be a competent user of language one must master the rules of the game.

So, although the six criteria sometimes considered in isolation in Chapters Two through Five, they are inextricably bound up together. The interaction of the six criteria show that none of them have simply been ‘tagged’. They all need each other to *be suitable* for the list of a use-based theory of meaning. And hence, each criterion drawn out as suitable for semantic conventionalism cannot stand alone — giving us a stronger basis for conventionalism than others give it credit.

Bibliography

Audi, R. (ed.): 1995, *The Cambridge Dictionary of Philosophy*, Cambridge University Press, Cambridge.

Beall, J.: 2000, Fitch's proof, verificationism, and the knower paradox, *Australasian Journal of Philosophy* **78**, 241–47.

Beck, A.: n.d., *The Liar Lies and Snow is White, Part 1*, PhD thesis, University of Bremen, Munchen.

URL: www.informatik.uni-bremen.de/till/Beck/

Brady, R.: 2003, *Relevant Logics and Their Rivals, A continuation of the work of Richard Sylvan, Robert Meyer, Val Plumwood and Ross Brady*, Vol. II, Aldershot: Ashgate.

Brandom, R.: 1994, *Making it Explicit*, Harvard University Press, Cambridge, Massachusetts.

Brandom, R.: 2000, *Articulating Reasons*, Harvard University Press, Cambridge, Massachusetts.

Carnap, R.: 1934, *The Logical Syntax of Language*, Open Court.

Carnap, R.: 1963, *The Philosophy of Rudolph Carnap*, Open Court, La Salle, Illinois, chapter Replies and Systematic Expositions, pp. 859–867.

- Carroll, L.: 1994, *Through the Looking Glass and What Alice Found There*, Puffin Classic, Toronto, Canada.
- Costa, N. D.: 1990, Belief, contradiction, and the logic of self-deception, *American Philosophical Quarterly* pp. 179–197.
- Davies, N. and DeVidi, D.: 2005, Actually existing anti-realism, *Unpublished (work in progress)* .
- DeVidi, D. and Kenyon, T.: 2003, Analogues of knowability, *Australasian Journal of Philosophy* **81**(4), 481–495.
- DeVidi, D. and Solomon, G.: 2001, Knowability and intuitionistic logic, *Philosophia: Philosophical Quarterly of Israel* **28**, 319–334.
- DeVidi, D. and Solomon, G.: 2005, Empirical negation in intuitionistic logic, in D. DeVidi and T. Kenyon (eds), *The Western Ontario Series in Philosophy of Science: A Logical Approach to Philosophy: Essays in Honour of Graham Solomon*, Springer Netherlands, Dordrecht, The Netherlands, pp. 151–168.
- Devitt, M.: 1993, A critique of the case for semantic holism, in E. L. J. Fodor (ed.), *Holism: A Consumer Update*, Vol. 46, Grazer Philosophische Studien, pp. 17–60.
- Dretske, F.: 1981, *Knowledge and the Flow of Information*, MIT Press, Cambridge, Massachusetts.
- Dretske, F.: 1986, Misrepresentation, in R. Bogdon (ed.), *Belief, Form, Content and Function*, Clarendon Press, Oxford.
- Dretske, F.: 1988, *Explaining Behavior*, MIT Press, Cambridge, Massachusetts.
- Dretske, F.: 1993, Conscious experience, *Mind* **102**, 263–283.

- Dummett, M.: 1973, *Frege: Philosophy of Language*, Harvard University Press, Cambridge, Massachusetts.
- Dummett, M.: 1991, *The Logical Basis of Metaphysics*, Harvard University Press, Cambridge, Massachusetts.
- Dummett, M.: 1993, *Origins of Analytical Philosophy*, Harvard University Press, Cambridge, Massachusetts.
- Fitch, F. B.: 1963, A logical analysis of some value concepts, *Journal of Symbolic Logic* **28**, 135–142.
- Fodor, J.: 1987, *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*, MIT Press, Cambridge, Massachusetts.
- Fodor, J. and Lepore, E.: 1992, *Holism: A Shopper's Guide*, Basil Blackwell Inc., Cambridge, Massachusetts.
- Grice, H.: 1957, Meaning, in H. Geirsson and M. Losonsky (eds), *Readings in Language and Mind*, Blackwell Publishers, Oxford, pp. 103–109.
- Grice, H.: 1975, Logic and conversation, in H. Geirson and M. Losonsky (eds), *Readings in Language and Mind*, Blackwell Publishers, Oxford, pp. 121–133.
- Grice, H.: 1989a, *Studies in the Way of Words*, Harvard University Press, Massachusetts.
- Grice, H.: 1989b, Utterer's meaning and intentions, *Studies in the Way of Words*, Harvard University Press, pp. 86–116.
- Grice, H.: 1989c, Utterer's-meaning, sentence-meaning, and word-meaning, *Studies in the Way of Words*, Harvard University Press, Cambridge, Massachusetts, pp. 117–137.

- Kannetzsky, F.: 2003, Some problems of a conventionalist approach to communication, meaning and understanding, *in* C. P. G. Meggle (ed.), *Saying, Meaning, Implicating*, Universitätsverlag, Leipzig, pp. 30–62.
- Kennedy, J.: 2003, *Space, Time and Einstein*, McGill-Queen's University Press, Montreal.
- Kripke, S.: 1975, An outline of a theory of truth, *Journal of Philosophy* **72**, 690–716.
- Kvanvig, J.: 1995, The knowability paradox and the prospects for realism, *Noûs* **29**, 481–500.
- Lycan, W.: 2000, *Philosophy of Language: A Contemporary Introduction*, Routledge, New York.
- Maloney, C.: 1990, Mental misrepresentation, *Philosophy of Science* **57**(3), 445–458.
- Maund, B.: 1974, The conventionality of temporal relations in relativity theory, *Philosophy of Science* **41**(4), 394–407.
- Mundy, B.: 1985, Physics, philosophy, and psychoanalysis by R.S. Cohen; L. Laudén, *Philosophy of Science* **52**(2), 318–320.
- Murzi, M.: 2006, Jules Henri Poincaré (1854-1912).
URL: <http://www.iep.utm.edu/p/poincare.htm>
- Oxford English Dictionary*: n.d., *OED* .
- Pateman, T.: 1982, David Lewis's theory of convention and the social life of language, *Journal of Pragmatics* **6**, 135–157.
- Percival, P.: 1991, Knowability, actuality and the metaphysics of context-dependence, *Australasian Journal of Philosophy* **69**, 82–97.

- Pinco, C.: n.d., Carnap, Russell and the external world: To appear in M. Friedman and R. Creath (eds.), *The Cambridge Companion to Carnap*, Cambridge: Cambridge university press.
- URL:** <http://web.ics.purdue.edu/~cpincock>
- Prawitz, D.: 1975, Ideas and results in proof theory, *Proceedings of the Second Scandinavian Logic Symposium* **63**, 235–261.
- Priest, G.: 1987, *In Contradiction: A Study of the Transconsistent*, Martinus Nijhoff, Dordrecht.
- Priest, G.: 1998, Dialetheism, *The Stanford Encyclopedia of Philosophy* **Winter**.
- URL:** <http://plato.stanford.edu/archives/win98/entries/dialetheism/>
- Restall, G.: 2000, *An Introduction to Substructural Logics*, Routledge, London.
- Ryle, G.: 1949, *The Concept of Mind*, Hutchinson Publishing Company.
- Searle, J.: 1969, *Speech Acts: an essay in the philosophy of language*, Cambridge University Press.
- Skinner, B.: 1953, *Science and Human Behavior*, Macmillan, New York.
- Skinner, B.: 1957, *Verbal Behavior*, Appleton-Century-Crofts, New York.
- Skinner, B.: 1974, *About Behaviorism*, Vintage, New York.
- Sklar, L.: 1977, *Space, Time, and Spacetime*, University of California Press, Berkley.
- Sturdee, D.: 1997, The semantic shuffle: Shifting emphasis in Dretske's account of representational content, *Erkenntnis* **47**, 89–103.
- Williamson, T.: 1987, On the paradox of knowability, *Mind* **96**, 256–261.

Williamson, T.: 2000, *Knowledge and Its Limits*, Oxford University Press,
Oxford, UK.