

# NAMING WITHOUT NECESSITY

By

NIGEL SABBARTON-LEARY

A thesis submitted to The University of Birmingham

for the degree of DOCTOR OF PHILOSOPHY

Department of Philosophy

School of Philosophy, Theology and Religious Studies

College of Arts and Law

The University of Birmingham

March 2010

## ABSTRACT

In this thesis I argue that we should break with the dominant Kripkean tradition concerning natural kind terms and theoretical identity. I claim that there is just no interesting connection between the metaphysics and semantics of natural kind terms, and demonstrate this by constructing a version of descriptivism that is combined with the same metaphysics – that is, a nontrivial version of essentialism – found in Kripke, but which effectively avoids all of the standard criticisms.

With my version of descriptivism in place, I present what I take to be the most reasonable version of metaphysical essentialism, positing only what I call ‘thin’ essences. I claim that thin essences are perfectly adequate to underpin scientific realism, and moreover that they are sufficient to support the version of descriptivism developed here. In effect, what I offer here is an error theory of the Kripkean tradition: Kripke is right to think that there are interesting things to say about meaning and essence, but just wrong about what those things are. Thus whilst Kripke thinks that it is possible to make discoveries about the meanings of natural kind terms, I think, rather, that we make empirical discoveries that lead to revisions in meaning. Furthermore, whilst Kripke thinks there is a dichotomy between *de re* and *de dicto* necessity, and that theoretical identities are necessary *de re*, I think this distinction is both misleading and inaccurate, and that the necessity of theoretical identities is neither entirely *de re* nor entirely *de dicto*. By separating and insulating questions concerning meaning from questions concerning essence I show that whilst scientific discoveries are contingent and *a posteriori*, the definition of scientific terms are both necessary and *a priori*.

# TABLE OF CONTENTS

INTRODUCTION .....	7
CHAPTER 1 .....	13
1.1 Frege's Puzzle .....	14
1.2 Description theory of meaning and reference.....	16
1.3 Kripke's objections.....	18
<b>1.3.1 The Modal Argument</b> .....	18
<b>1.3.2 The Epistemological Argument</b> .....	21
<b>1.3.3 The Semantical Argument</b> .....	22
1.4 The Causal Theory of Direct Reference .....	23
1.5 A Solution to Frege's Puzzle? .....	29
1.6 Natural Kinds.....	32
1.7 Puzzles .....	37
CHAPTER 2 .....	44
2.1 Putnam's semantics of natural kind terms.....	44
2.2 Cross-world reference.....	50
2.3 <i>A posteriori</i> necessity .....	54
2.4 The Kripke-Putnam programme .....	57
2.5 Reference and Kuhnian incommensurability .....	60
2.6 LaPorte on Twin Earth .....	66
2.7 Later Putnam .....	68
CHAPTER 3 .....	75

3.1 Natural kind terms, predicates and identities.....	76
3.2 Revisiting Kripke.....	80
3.3 Soames' positive account I.....	88
<b>3.3.1 Essentialist Predicates</b> .....	90
<b>3.3.2 Correspondence rigidity</b> .....	96
3.4 Soames' positive account II.....	98
3.5 Simple theoretical identification sentences .....	107
3.6 Kinds of a type <i>T</i> ?.....	112
3.7 The Special Auxiliary Assumption.....	115
3.8 False theoretical identities .....	121
3.9 Complex natural kind terms and essentialism .....	126
CHAPTER 4.....	135
4.1 Salmon vs. Soames .....	136
4.2 Salmon's ( <i>CDR</i> ) .....	137
4.3 Objections to descriptivism about natural kind terms .....	143
4.4 Theoretical identity and <i>a posteriori</i> necessity.....	148
4.5 Kripke's essentialist premise .....	153
4.6 Essentialism, <i>a priori</i> or <i>a posteriori</i> ?.....	156
4.7 Salmon's account of theoretical terms.....	159
4.8 The trivialisation problem .....	162
4.9 The semantics of science .....	166
TAKING STOCK: THE STORY SO FAR.....	169

CHAPTER 5.....	173
5.1 Kripke vs. Description Theory .....	174
5.2 Rigidity and Descriptiveness .....	177
5.3 De Facto Rigidity, Essentialism and Descriptivism.....	181
5.4 Descriptivist Essentialism.....	186
5.5 Locating Meaning.....	189
5.5 <i>A priori</i> knowability .....	194
5.6 The metasemantics of natural kind terms .....	196
5.7 Intuitive Kinds vs. Theoretical Kinds.....	200
5.8 Conclusion .....	208
CHAPTER 6.....	211
6.1 Real Essence, Nominal Essence <i>and</i> Real Constitution.....	213
6.2 The Knowledge Objection.....	220
6.3 The Ontological Objection .....	226
6.4 The Better Candidate Objection .....	231
6.5 Locke's Contemporary Relevance.....	236
CHAPTER 7.....	242
7.1 Trivial Essentialism, Non-trivial Essentialism, and the Thick/Thin Distinction.....	243
7.2 Scientific Essentialism.....	246
7.3 Promiscuous Realism .....	249
7.4 Causal Intrinsicity .....	254
7.5 Are kinds impoverished without (thick) essences? .....	259

7.6 The erosion of the natural/artificial distinction, and the epistemology of essence.....	263
7.7 Challenge to necessity .....	266
7.8 A problem for descriptivism about reference? .....	267
CHAPTER 8.....	270
8.1 Distilling kinds .....	270
8.2 Distilling essences .....	275
<b>8.2.1 Thin vs. Thick Essences</b> .....	277
8.3 Grounding necessity .....	280
8.4 A final word on reference.....	285
BIBLIOGRAPHY .....	289

## INTRODUCTION

This thesis engages with two key topics in contemporary philosophy: the first is the semantics of natural kind terms, particularly those for substances ('tungsten', 'water', 'fosterite'), the second is natural kind essentialism, specifically those brands of natural kind essentialism which claim that essences are both *a posteriori* and metaphysically necessary.

Examples of so-called natural kinds are, for the most part, intuitive and obvious. Substances like gold, iron and water, species such as tigers, whales and chimpanzees, and fundamental particles like photons, muons and electrons are all traditionally supposed to be natural kinds. But what do all of these examples have in common? At work in each of the examples is an intuitive notion (of a natural kind) that is intended to capture the salient facts of our experience (including scientific experience) that inspire us to group particular individuals together, on the basis of similarity. For instance, if I drink two glasses of water, from a single glass, over the course of an hour, then I take it that although the two glasses, or rather the two bodies of liquid that fill the glass at different times, are not numerically identical, the substance I have drunk is one kind of thing. Similarly, if my dog has puppies I take it that each of the puppies is a member of the same kind, despite all of them being obviously distinct individuals.

This idea of a natural kind is supposed to be 'basic to thought and language' (Quine 1977: 157). Furthermore, it is closely connected to the notion of similarity or resemblance – according to Quine, they are ostensibly 'variations or adaptations of a single notion' (*ibid.*). Thus we might say that when someone identifies two objects as of the same kind, what they are claiming is that the two objects are similar in some relevant respect.

Of course, similarity alone does not get us to *natural* kinds. For instance, both sugar and salt are soluble in water, and thus have at least one property in common. However, they hardly seem to form a natural kind on that basis. Traditionally, what is thought to distinguish the natural kinds like gold from the artificial kinds (pencil, fridge) and the gerrymandered kinds (things-in-my-fridge) is their application in scientific inductions, and their independent existence (see Bird 1998; Collier 1996). Inductions about pencils and things-in-my-fridge are less likely to be reliable, since members of each kind can be radically different, than corresponding inductions about gold, where we take it that all instances of gold are similar in some (scientifically) important respect that serves to ground inductions, making our classifications of objects into kinds, and subsequent predictions on that basis, highly reliable.

The intuitive contrast, then, should be clear enough. On the one hand we have the natural kinds, which include natural substances that exist independently of us, and are susceptible to induction and scientific theorising. On the other hand we have the artificial and gerrymandered kinds, which exist *because* of us, are inductively unreliable, and not the subject of scientific theorising.

The rekindled interest in natural kind essentialism<sup>1</sup> – that view that natural kinds have essences, and objects get to be members of said kinds by possessing the relevant kind essence – and natural kind semantics<sup>2</sup>, stems from Saul Kripke's seminal work *Naming and Necessity* (1980), where he convinced the majority of the analytic philosophical community that there

---

<sup>1</sup> Natural kind essentialism is typically attributed to the scholastics, in particular Aristotle. Locke's discussion of essentialism in *An Essay Concerning Human Understanding*, for example, aims to criticise the Aristotelian view. More recently, David Oderberg (2007) has mounted a contemporary defence of Aristotelian essentialism.

<sup>2</sup> Although one might think that an interest in natural kind semantics stems solely from Kripke this is not strictly true. For example, John Stuart Mill, in *A System of Logic* (1843), explicitly discussed the semantics of natural kind terms. David Charles (2000) also argues that Aristotle had a particular semantics of (natural) kind terms.



could be certain necessary truths, those concerning identity and theoretical identity, that are nevertheless knowable only *a posteriori*. Put another way, Kripke argued that there are two brands of necessity: metaphysical and epistemic, and moreover they can come apart. So, whilst ‘all bachelors are unmarried men’ is both metaphysically and epistemologically necessary, identity sentences like ‘Hesperus is Phosphorus’ and theoretical identities like ‘water is H<sub>2</sub>O’ are metaphysically necessary but epistemically contingent.

These conclusions about the *a posteriori* necessity of theoretical identities have often been taken to motivate a return to a version of Aristotelian essentialism: the thesis that the world is objectively arranged into kinds, and that what it is to be a member of a particular kind is to instantiate the relevant kind essence. Indeed Hilary Putnam (1973), like Kripke, endorsed precisely this conclusion, albeit stemming from his independently developed, but clearly parallel, semantics for natural kind terms. In effect, both philosophers claimed to have derived metaphysically necessary truths from semantic and empirical premises, Kripke concerning individuals and natural kinds, Putnam concerning kinds. Whilst most philosophers were convinced by the Kripke’s semantic theses concerning proper names<sup>3</sup>, and the metaphysical necessities they generate, the parallel conclusions for natural kinds and theoretical identities – what we might call the Kripke-Putnam programme – continues to be more controversial. That controversy concerns both the semantic claims about the reference of natural kind terms, and in the derivation of *a posteriori* necessities concerning natural kinds.

However, the Kripke-Putnam programme was not *just* about the derivation of *a posteriori* necessities concerning names and natural kind terms. Both Kripke and Putnam

---

<sup>3</sup> There are, of course, some exceptions. For instance, Harold Noonan (1981) maintains that a Fregean account of names is still possible.

were also interested in undermining the then traditional theories of meaning and reference developed by Gottlob Frege (1948), which we can label (broadly) as *descriptivism*: the thesis, roughly, that name-like terms have descriptive meanings. Putnam, in addition, was also concerned with the problem of Kuhnian incommensurability (1996), the problem of referential stability across scientific theory change. Together they offered a host of objections to descriptivism, and their own semantic thesis – the causal theory of direct reference – is now the contemporary orthodoxy.

In chapter 1 of this thesis I will start by analysing Kripke's *Naming and Necessity*, presenting his semantic theses (first for proper names, and then extended to natural kind terms), and show how we are supposed to draw conclusions concerning natural kinds that are necessary yet *a posteriori*. I will then point out a number of puzzles that Kripke's initial account raises, many of which have already been identified in the literature.

In chapter 2, rather than move on directly to answer the puzzles raised at the end of chapter 1, I will present and analyse Putnam's semantic thesis, developed specifically for natural kind terms (rather than extended from names to natural kind terms, as Kripke's is). I will also take time to point out the dissimilarities between Kripke's project and Putnam's, since this is often neglected in the literature, and I will also discriminate between early and later Putnam, and present Putnam's later arguments *against* the conclusion that there are metaphysically necessary truths about natural kinds.

In chapter 3 I will discuss the first of the puzzles identified in chapter 1 – that natural kind terms and theoretical identities have some obvious dissimilarities with proper names and identity sentences. I will analyse and object to a recent strategy developed by Scott Soames (2002) to extend Kripke's less controversial conclusions about proper names to natural kind

terms, since it recognises the sorts of dissimilarities I identify, but nevertheless attempts to maintain the overall semantic thesis. I will conclude that this strategy is unsuccessful, and argue that the metaphysical necessities it purports to derive purely from linguistic premises (Soames 2002: 279) are, in fact, the product of controversial metaphysical presuppositions, which were identified by Nathan Salmon (originally published in 1982, all references to 2005 edition) to be implicit in Putnam's and, to a lesser extent, Kripke's original work.

In chapter 4, after identifying these implicit metaphysical assumptions in Soames, I turn to Nathan Salmon, and the second of the main puzzles raised at the end of chapter 1 and chapter 2: the implicit essentialism in Kripke and Putnam. As is well known, Salmon, in *Reference and Essence*, identified a key essentialist premise in Putnam's derivation of necessary *a posteriori* truths concerning natural kind terms. In this chapter I will present that analysis, and show (in more detail than Salmon) how it extends to Kripke. However, I will argue that Salmon's alternative thesis, which endorses the Kripke-Putnam semantics without the metaphysics, is unsuccessful for a number of reasons (both exegetical and philosophical). I will conclude that the most successful versions of the causal theory of direct reference, (*CDR*) for short, from the variations considered, are those that contain an essentialist premise.

In chapter 5 I will present and defend a novel version of descriptivism, based upon the conclusion of chapter 4, and a general objection to the sort of semantics that (*CDR*) offers for natural kind terms, which I take to be taxonomic terms. I contend that classificatory terms are *intended* to communicate descriptive information. My own thesis, descriptivist essentialism, I claim, captures the main desiderata of a taxonomy, in comparison to (*CDR*), where the notion of reference is simply assumed to be unproblematic. I will also demonstrate how this novel descriptivism avoids the standard objections identified in chapters 1 and 2. My main strategy is to use the same metaphysical assumptions, particularly essentialism, as proponents of

(*CDR*), to avoid both the modal and semantical arguments. I will then present a new metasemantic story about how taxonomic names, at least for chemical kinds, are introduced. By introducing a sharp division between *a posteriori* scientific investigation, and *a priori* naming events, I will claim that theoretical identities are plausibly *a priori*.

In chapter 6, having identified in chapter 4 the implicit natural kind essentialism of Kripke, Putnam and Soames (which does all the heavy lifting in the derivation of necessary *a posteriori* truths concerning natural kinds), and having rejected Salmon's version of (*CDR*) as implausible for taxonomic terms, and used this same essentialism myself to defend a version of descriptivism in chapter 5, I now identify the root of this *a posteriori* essentialism in John Locke. However, I will argue that, in fact, Locke was an anti-essentialist, and that the various appeals to essentialism, and the notion of a 'real essence', that have been made are speciously associated with Locke's philosophy of essence. In this chapter I show precisely why Locke is *not* an essentialist, and diagnose where the misunderstanding has occurred. I conclude that Lockean 'real essences' cannot do the work that either (*CDR*) or descriptivist essentialism requires.

In chapter 7, having dismissed Locke's notion of essence as the relevant (metaphysical essentialist) notion to feature in theoretical identifications, I turn to a recently developed brand of natural kind essentialism – Brian Ellis's scientific essentialism – that purports to be concerned with 'the scientific practice of theoretical identification' (Ellis 2001: 54), and further claims that truths about natural kinds are metaphysically necessary, yet knowable only *a posteriori*. I offer a variety of counterexamples to his essentialist theses, and argue that the empirical facts refute the theory. I conclude that the appeal to *a posteriori* essences to derive necessary *a posteriori* theoretical identities is unsuccessful.

In chapter 8 I conclude the thesis, mapping out a metaphysically thin notion of essence, where the notion of an ‘essence’ is treated as a species of definition, generated by the conjunction of intellectual fiat and the physical facts. I then take this notion and apply it to descriptivist essentialism, noting that when we treat the notion of an essence as a species of definition, the so-called essentialism loses its more metaphysically suspicious elements. In effect, I endorse Locke’s antiessentialism, albeit without its nominalist trappings, employ the notion of a ‘real essence’ as a philosophical term of art, intended to communicate the notion of a ‘definition’, and relativise definitions to particular theories (in this case, contemporary chemistry). The conclusion of my thesis is that there is just no interesting connection between the semantics and metaphysics of natural kinds.

Some of the work in this thesis has already appeared in print, in the form of both book chapters and articles. Chapters 5 and 8 both contain material that also appears in the paper ‘The Abuse of the Necessary A Posteriori’, co-authored with Helen Beebe, in *The Semantics and Metaphysics of Natural Kinds*, edited by Helen Beebe and Nigel Sabbarton-Leary, published by Routledge in 2010. The majority of chapter 6 appears in the article ‘How Essentialists Misunderstand Locke’ in *History of Philosophy Quarterly* Vol. 26, No. 3, pp. 273 – 292, published in 2009. The defence of promiscuous realism for chemical kinds that takes place in chapter 7, and the distinction between thick and thin essentialism sketched in chapter 8, had their first outing in my article ‘Natural Kinds: (Thick) Essentialism or Promiscuous Realism’ in *Philosophical Writings* Vol. 34 pp. 5 – 12, published in 2007 (which, incidentally, won the prize for best paper at the British Postgraduate Philosophy Association annual conference in 2007).

## CHAPTER 1

In this chapter I expound Saul Kripke's semantic thesis as found in *Naming and Necessity* (first published in 1972, but all references to 1980 edition). I present Kripke's thesis as a response to a descriptivist theory of meaning and reference which, in turn, was introduced in response to Frege's Puzzle. I demonstrate how Kripke's account is supposed to extend to natural kind terms, and use this extension to generate various questions that will set the topic for subsequent chapters.

The structure of the chapter is as follows: in section 1 I give a brief account of Frege's Puzzle in order to motivate the discussion of identity and co-referring names. In section 2 I present the Fregean response to the puzzle. In section 3 I introduce Kripke's response to Frege's theory of meaning and reference, and present each of Kripke's objections in turn. In section 4 I present Kripke's positive thesis, the *Causal Theory of Direct Reference (CDR)*. In section 5 I outline how Kripke's positive thesis is supposed to solve Frege's Puzzle. Finally, in section 6 I will extend the account to natural kind terms. In section 7 I will raise a number of puzzles, setting the agenda for the following chapters.

## **1.1 Frege's Puzzle**

We can start this investigation by asking what a name contributes to the meaning of a sentence containing it: does the name merely contribute its referent, or does it have some further meaning?

According to Frege, if 'a' is a name then an identity statement 'a = a' is knowable *a priori* and necessary. A (minimally competent) language user who has acquired the name 'a'

can come to know the truth of the identity statement 'a = a' merely by reflecting on the meaning of the name 'a'. But consider the identity statement 'a = b'. As Frege noted, this second identity statement 'contains a valuable extension of our knowledge and cannot always be established a priori' (Frege 1948: 209). That the second identity statement is an extension of our knowledge demonstrates that the names 'a' and 'b' have different cognitive values. However, if identity is understood as a relation between whatever names happen to refer to (i.e. what the name denotes), and 'a' and 'b' have the same referent, then the claim that 'a = a and a = b are obviously statements of differing cognitive value' (Frege 1948: 209) appears, false: if all there is to the meaning of a name is its referent, and 'a' and 'b' have the very same referent, then the identity statements 'a = a' and 'a = b' express the same proposition, and cannot differ in cognitive value. Consequently we are faced with what has traditionally been dubbed as *Frege's Puzzle*. Consider the following identity statements:

(1) Marie Curie is Marie Curie

(2) Marie Curie is Maria Sklodowska

If the meaning of a name is that to which it refers, and the names 'Marie Curie' and 'Maria Sklodowska' refer to the same person, then the meanings of sentence (1) and (2) are identical – they express the same proposition. But (2) *cannot* be a useful extension of our knowledge and yet express the same proposition as (1), since (1) is knowable *a priori*, and is *not* a useful extension of our knowledge. But (2) is knowable only *a posteriori*. Hence there is a cognitive asymmetry between (1) and (2): while the truth of the proposition expressed by (1) is knowable *a priori* the truth of the proposition expressed by (2) is knowable only *a*

*posteriori*. Were the two identity statements expressing the very same proposition there would be no such cognitive asymmetry. That there is an asymmetry indicates that the thesis that the meaning of a name is simply that to which it refers must be false.

The cognitive asymmetry can be made even more obvious by considering the following example:

(3) Marie Curie won the Nobel Prize in Physics.

(4) Maria Sklodowska won the Nobel Prize in Physics.

Both (3) and (4) express propositions that are potentially valuable extensions of our knowledge, but it is entirely possible to understand both (3) and (4) yet think that (3) is true and (4) is false without being accused of irrationality. But if 'Marie Curie' and 'Maria Sklodowska' meant the same thing, (3) and (4) would express the same proposition. If (3) and (4) express the same proposition then to believe that (3) is true, and yet maintain that (4) is false (or even merely to remain agnostic about whether it is true) would be inconsistent. That it is not inconsistent shows, again, that (3) and (4) do not express the same proposition.

## **1.2 Description theory of meaning and reference**

Frege's solution to the problem of cognitive asymmetry is to argue for something 'besides that to which the sign refers, which may be called the referent of the sign, wherein the mode of presentation is contained' (Frege 1948: 210). In the Marie Curie/Maria Sklodowska



example the solution would work as follows. While the *referent* of the names ‘Marie Curie’ and ‘Maria Skłodowska’ is the same, the *sense* of the name ‘Marie Curie’ is different to the *sense* of the name ‘Maria Skłodowska’. Consequently Frege can account for the cognitive asymmetry between co-referring names by arguing that singular terms do not solely denote objects in the world but all carry a *sense*, which ‘is the manner in which the term presents its denotation to the listener or reader’ (Salmon 2005: 9). Thus there is a distinction between the *sense* of singular terms such as the name ‘Marie Curie’, and that name’s referent. The sense and the referent of a singular term are related insofar as while the sense is conceptual, and it is possible for ‘the sense of ordinary proper names [to vary] from person to person’ (Burge 1977: 357), the referent of the singular term is ‘whatever uniquely fits the concept’ (Salmon 2005: 9) contained in the sense. The sense of a singular term should therefore be understood as playing *at least* a dual role within the Fregean thesis: it is both the ‘mode of presentation to the speaker’ (Burge 1977: 356) and that which is ‘determining the referent or denotation associated with the expression’ (*ibid.*).

The general Fregean picture, what we might call a description theory of meaning and reference (or just descriptivism for short), then, is that the sense of a term supplies ‘a set of conditions, or properties, and the denotation, if any, is what uniquely satisfies those conditions’ (Salmon 2005: 9). In the case of the name ‘Marie Curie’ the sense might be ‘the first woman to be awarded the Nobel Prize’. Thus, whosoever uniquely satisfies the description by *being* the first woman to be awarded the Nobel Prize *is* Marie Curie. Since the sense of the name ‘Maria Skłodowska’—perhaps ‘the woman who earned a physics degree in 1893 and a maths degree in 1894 at the Sorbonne’—is *not* the same as the sense of ‘Marie Curie’, we can explain why one can believe (3) to be true, whilst thinking that (4) is false.

Similarly, this explains why (2) is *a posteriori* and an extension of our knowledge, whilst (1) is merely *a priori*.

There are two possible ways to construe the Fregean thesis. It can be held as either (i) a theory of reference, or (ii) a theory of meaning (cf. Kripke 1980: 53-54). While the two options are not mutually exclusive – a proponent of (i) could also subscribe to (ii) – it is important to distinguish between them. If we were to subscribe to the Fregean thesis solely as a theory of *reference* this entails only that *the object satisfying the description is the referent of the name*. However, if we also subscribe to the thesis as a theory of *meaning* then this entails that *a name can be replaced, within a sentence, by the description which is its sense, without changing the meaning of that sentence*. As such if we substitute, within a sentence, some name for its synonymous description there will be no change of truth value since the description (which is the sense) is *synonymous* with the name. For the sake of this discussion we will be considering the Fregean thesis as one of reference *and* meaning, and considering the problems associated with the claim that names are synonymous with descriptions.

### **1.3 Kripke's objections**

The main objector to the Fregean account of semantics is Saul Kripke (1980). Following Nathan Salmon (2005: 23-31) I shall divide and label Kripke's arguments as follows: (a) the modal argument, (b) the epistemological argument, and (c) the semantical argument. In the following sections I will expound and discuss each in turn.

#### ***1.3.1 The Modal Argument***

The modal argument targets the Fregean's use of some simple set of properties to form the basis of a description which, when uniquely satisfied by some person (or object), (i) denotes the referent of a name, and (ii) is considered to be synonymous with some particular name. Consider, for example, the name 'John Locke', and assume that the sense of the name is 'the author of *An Essay Concerning Human Understanding*, and personal physician to the first Earl of Shaftsbury'. Thus according to (i) and (ii) we can say that the description 'the author of *An Essay Concerning Human Understanding*, and personal physician to the first Earl of Shaftsbury' both denotes the referent of the name, and is synonymous with it. As such, consider the following three sentences:

(5) John Locke, if he exists, is the author of *An Essay Concerning Human Understanding*, and personal physician to the first Earl of Shaftsbury.

(6) If anyone is the author of *An Essay Concerning Human Understanding*, and personal physician to the first Earl of Shaftsbury, then he is John Locke.

(7) Someone is the author of *An Essay Concerning Human Understanding*, and personal physician to the first Earl of Shaftsbury if and only if he is the author of *An Essay Concerning Human Understanding*, and personal physician to the first Earl of Shaftsbury.

Given (ii) we are able to derive (7) from (5) and (6) by substituting the description ‘the author of *An Essay Concerning Human Understanding*, and personal physician to the 1<sup>st</sup> Earl of Shaftsbury’ with its *apparently* synonymous name ‘John Locke’. However, (7) is evidently a *necessary* truth, and as the only difference between (7) and (6) or (5) is the substitution of ‘John Locke’ for its synonym ‘the author of *An Essay Concerning Human Understanding*, and personal physician to the 1<sup>st</sup> Earl of Shaftsbury’ then both (6) and (5), like (7), must be necessary.

But if we consider (5) then it seems obvious that we can imagine a possibility where John Locke was not the author of *An Essay Concerning Human Understanding* because, while at Oxford, he discovered a passion for literature, and never came to do philosophy. Furthermore we can also imagine that Locke, as a result of his all encompassing passion for literature, never studied medicine. As such he was of no use to the first Earl of Shaftsbury when he came to Oxford seeking medical treatment, and thus never became his personal physician.

The result of the argument, given the possible situation in which (5) turns out to be false, is that (5) *cannot* be necessarily true<sup>4</sup>; any proposition that can be false cannot be necessarily true! Therefore, since (5) comes out as necessary on the Fregean thesis (assuming, of course, that the Fregean would pick *that* description), that thesis must be false.

What motivates Kripke’s modal argument is the (Millian) intuition that names like ‘John Locke’ still refer to the same person notwithstanding any counterfactual scenarios in which, for example, Locke failed to become the author of *An Essay Concerning Human*

---

<sup>4</sup> An analogous argument can be run for sentence (6), where someone else does all the things cited in the description, which according to the Fregean are sufficient for picking out the referent, but is not John Locke (perhaps David Hume did, in some counterfactual scenario) do all the things that John Locke did in this world.

*Understanding*, and personal physician to the first Earl of Shaftsbury. In other words names such as ‘John Locke’ continue to denote their referents in counterfactual scenarios where the referent failed to do the things which we (contingently) associate with them in this world.

### ***1.3.2 The Epistemological Argument***

The epistemological argument is similar to the modal argument, but targets a language user’s ability to *know* the truth of a given proposition expressed by some sentence. If we again consider (7) and either (5) or (6), the Fregean claims that the name ‘John Locke’ is synonymous with the description ‘the author of *An Essay Concerning Human Understanding*, and personal physician to the first Earl of Shaftsbury’. Hence not only is (7) is knowable *a priori*, but since (5) and (6) express the same proposition as (7), the truths of (5) and/or (6) are *also* knowable *a priori*. In other words, to imagine John Locke without the attributes of being the author of *An Essay Concerning Human Understanding*, and personal physician to the first Earl of Shaftsbury ‘should be as difficult as imagining a “married bachelor” or an “unmarried husband”’ (Salmon 2005: 28). The commitment of the Fregean thesis is that the *meaning* of the name ‘John Locke’ precludes Locke *not* having the properties of being the author of *An Essay Concerning Human Understanding*, and being the personal physician to the 1<sup>st</sup> Earl of Shaftsbury in just the same way as it is precluded of anyone to whom the term ‘bachelor’ truly applies that they have the property of being a married man.

However, not only do we think that we *can* imagine a situation where John Locke failed to do the things we commonly attribute to him, and that this counterfactual scenario is possible, we also do not think that such reflections are precluded merely by reflection on the meaning of the terms in sentences like (5) and (6). The point is that sentences like (5) and (6)

may have the *prima facie* appearance of sentences like ‘a bachelor, if he (or one) exists, is an unmarried man’, but while the bachelor sentence is knowable *a priori*, (5) and (6) are knowable only *a posteriori*. We do not think that sentences such as ‘John Locke, if he exists, is the author of *An Essay Concerning Human Understanding*, and personal physician to the first Earl of Shaftsbury’ are true by definition (and hence knowable *a priori*). Therefore, *mutatis mutandis* for names in general, we should admit that the Fregean theory of semantics misrepresents what it is to be a singular term such as a name, as illustrated by the Locke example.

### ***1.3.3 The Semantical Argument***

The semantical argument differs from both the modal and epistemological arguments in that it targets the notion of reference directly and is, according to Salmon, ‘the strongest and most persuasive of the three kinds of argument from the primary thesis of the direct reference theory’ (Salmon 2005: 29).

Consider the following example: take the name ‘Charles Darwin’ and consider the properties that a language user might associate with the name that make up the Fregean notion of *sense*. Imagine, for the sake of argument, that the description was something like ‘the British naturalist who theorised that evolution was the effect of natural selection on advantageous traits’. According to the Fregean thesis of reference, the name ‘Charles Darwin’ denotes whomsoever satisfies the description ‘the British naturalist who theorised that evolution was the effect of natural selection on advantageous traits’. But suppose that Charles Darwin did not theorise that evolution was the effect of natural selection, although we mistakenly think that he did. Suppose further that some other British naturalist did so theorise

about evolution, say, Alfred Russell Wallace. Finally suppose that we are totally unaware of Alfred Russell Wallace's existence, and *a fortiori* totally unaware that he had any theories on evolution.

The problem facing the Fregean is that they are committed to saying that the name 'Charles Darwin' refers to Alfred Russell Wallace if he satisfies the description 'the British naturalist who theorised that evolution was the effect of natural selection on advantageous traits' because the Fregean holds that 'precisely whom a name denotes depends entirely on whoever happens to have certain properties uniquely' (Salmon 2005: 37). In the scenario outline above it is Alfred Russell Wallace who uniquely satisfies the description associated with the name 'Charles Darwin'. As a result the name 'Charles Darwin' denotes Alfred Russell Wallace – despite us having no knowledge that such a man even exists.

Kripke's point is that the fact that Alfred Russell Wallace happens to have the properties referred to in the description 'the British naturalist who theorised that evolution was the effect of natural selection on advantageous traits' should have no bearing on who the referent of the name 'Charles Darwin' is. Even if Alfred Russell Wallace did have Charles-Darwin-like properties, this should be irrelevant to the denotation of the name 'Charles Darwin'. That the particular properties do play such a prominent role in the Fregean thesis is sufficient to show, given the previous example, and *mutatis mutandis* other examples involving names, that the theory is false. Thus we require an alternative thesis. It is to this – Kripke's positive thesis – I now turn.

#### **1.4 The Causal Theory of Direct Reference**

Kripke takes the preceding three arguments to be a refutation of the Fregean view of semantics, and uses the problems left by the three arguments to motivate his own alternative account, first of proper names, and then natural kind terms. In what follows I will present Kripke's alternative picture of semantics, show how it purports to avoid the modal, epistemological and semantical arguments, and in so doing start to discriminate between what I take, in fact, to be two separate concerns within the Kripkean thesis.

As stated at the beginning of this chapter, let us label Kripke's semantic thesis as the *Causal Theory of Direct Reference (CDR)*. The thesis provides an account of both the *meaning* – claiming that the meaning is the referent – and the *reference* of names, and presents us with an idealised picture of how reference to the bearer of a name is achieved. The theory is aptly characterised as causal since the meaning of a name is passed *causally* from speaker to speaker through a community of language users. The claim is that speakers along the causal chain intend to refer to exactly the thing that the initial speaker referred to when they first deployed the name. In other words, 'there is a causal chain of *reference-preserving* links leading back from' some speaker's use to the first use of the name as deployed by the introducer of that name (Evans 2008: 316). The 'first use' alludes to what Evans terms a 'name-acquiring transaction' (*ibid.*) where some object, *x*, is picked out by some language user, *Lu*, in what Kripke calls an 'initial "baptism"' (Kripke 1980: 96) where some object is 'named by ostension, or the reference of a name may be fixed by a description' (*ibid.*). However, even in cases where reference *is* fixed via a description, on Kripke's view 'the description used is not synonymous with the name' (Kripke 1980: 96 n42), and nor is it *uniquely* identifying of just that object. The fundamental thought is that in most cases a language user 'is acquainted in some sense with the object he names and is able to name it ostensively' (*ibid.*) without the mediation of any description – hence the *direct reference*.



A rough statement of the theory might be the following: An initial ‘baptism’ takes place. Here the object may be named by ostension, or the reference of the name may be fixed by a description. When the name is ‘passed from link to link’, the receiver of the name must, I think, intend when he learns it to use it with the same reference as the man from whom he heard it. (Kripke 1980, p. 96)

The first thing to note about Kripke’s positive account is that it is an ‘as-if’ account. Kripke is *not* claiming that these dubbing ceremonies *literally* take place, conceding rather that there are ‘artificialities in this whole account’ (Kripke 1980: 139 n70). The point is simply that names (and, in due course, natural kind terms) ‘behave semantically *as if* they were introduced in Kripke style definitions’ (Brown 1998: 279 *emphasis added*).

The second, and more important, thing to note is the apparent force of the Kripkean picture in relation to the modal, epistemological and semantical arguments used against the Fregean. Kripke ‘*endorses* Mill’s view of *singular* terms’ (Kripke 1980: 135) where ‘names do not refer to their referent by specifying a condition which their referent uniquely satisfies’ (Hughes 2004: 1) but rather refer directly. As such ‘there is no reason why we cannot stipulate that, in talking about what would have happened to [a person *P*] in a certain counterfactual situation, we are talking about what would have happened to *him*’ (Kripke 1980: 44). Consider, again, the name ‘John Locke’. If meaning and reference accords with the Kripkean semantic picture then the name ‘John Locke’ was fixed *to* a particular person during an initial baptism, and refers directly to him. Language users who are causally downwind of this event

intend, when deploying the name, to refer to whomever the initial dubber – presumably John Locke’s parents – referred to when they introduced the name.

By telling this type of meta-semantic story about how the reference of a term is fixed, Kripke can avoid the semantical argument. The semantical argument, which targets reference *simpliciter*, demonstrated a particular problem with a semantic thesis committed to the view that the referent of a name is whatever uniquely satisfies some description. The problem was that *if* singular terms such as names have a *sense* – which is a description picking out some set of properties – then should that sense, through ignorance or error, be uniquely satisfied by some *other* person or object (as in the Charles Darwin/Alfred Russell Wallace case), then that *other* person is the referent of the name. On the Kripkean view *there is no sense*. Names refer directly, and are non-descriptive (their reference is not mediated by description). Therefore irrespectively of our abilities to discern any necessary (and perhaps sufficient) conditions, which the referent of the term actually does uniquely satisfy, we can refer simply in virtue of the meta-semantic story about how the reference of the name was fixed.

However, there is a further corollary of the non-descriptiveness of names (cf. Salmon 2005, Soames 2002) such that we can ask counterfactual questions about *this* or *that* person without worrying whether we can identify *them* in any given counterfactual scenario. So, as Kripke says:

[A]lthough the man (Nixon) might not have been the President, it is not the case that he might not have been Nixon (though he might not have been called ‘Nixon’)... [and] it is because we can refer (rigidly) to Nixon, and stipulate that we are speaking of what

might have happened to him (under certain circumstances), that ‘transworld identifications’ are unproblematic in such cases. (Kripke 1980, p. 49)

Thus Kripke introduces his notion of referring *rigidly* such that reference across possible worlds is unproblematic. The notion of rigid reference, or of being a *rigid designator*, is just the idea that a term is rigid if ‘in every possible world it designates the same object’ (Kripke 1980: 48). Since ‘names are always rigid designators’ (Kripke 1980, p. 58), any counterfactual scenario concerning a person, be it John Locke, Richard Nixon or Marie Curie, does not raise problems of identity, because in asking questions about what might have happened to *them*, we have stipulated that it is exactly *them* we are talking about.

The introduction of the notion of rigidity to Kripke’s semantic thesis (*CDR*) serves to explain why the modal and epistemological arguments have no force against this view. The epistemological argument, recall, targeted a language user’s ability to *know* the truth of a proposition expressed by some sentence, concluding that the Fregean was committed to saying that (5), (6) and (7) all express the same proposition because a name is synonymous with its sense, and can be substituted within a sentence with no loss of truth value. As such given that the truth of (7) was obviously knowable *a priori*, and expressed the same proposition as (5) and (6), their truth was also knowable *a priori*. The modal argument, on the other hand, revealed the Fregean’s commitment to the necessary truth of certain propositions which, under normal circumstances, we consider to be contingent. Thus, again, the synonymy of a name with a description, and the substitutions from (5) and (6) to (7) entailed that since (7) was necessary so too were (5) and (6).

For Kripke although the referent of a singular term can, in some cases, be fixed *via* description, the meaning of the name just is the referent, and thus is *never* synonymous with some description. Names, ‘as a corollary of the more central thesis that they are nondescriptonal, together with an account of how their reference is fixed in the actual world’ (Soames 2002: 264) are always *rigid designators*. The point can be neatly illustrated using the previous example.

In the example, sentence (5) states that John Locke is the author of *An Essay Concerning Human Understanding*, and personal physician to the first Earl of Shaftsbury. It might be that the description ‘is the author of *An Essay Concerning Human Understanding*, and personal physician to the first Earl of Shaftsbury’ is the convention we used to affix the reference of the name ‘John Locke’. Does it follow from the adoption of this convention that what it is to be John Locke consists in being the author of *An Essay Concerning Human Understanding*, and personal physician to the first Earl of Shaftsbury? Intuitively, the answer is no. The reference of the name ‘John Locke’ is being fixed by an accidental (contingent) property of him, that he authored a particular book and was the physician of a particular person. But what we are trying to denote is that man, rather than whomsoever has the properties we associate with him. Hence there is an intuitive and obvious difference between the name ‘John Locke’ and the phrase ‘is the author of *An Essay Concerning Human Understanding*, and personal physician to the first Earl of Shaftsbury’. While ‘John Locke’ qualifies as a rigid designator, picking out the same man in all possible worlds, the phrase ‘is the author of *An Essay Concerning Human Understanding*, and personal physician to the first Earl of Shaftsbury’ does not designate anyone rigidly, since authoring *An Essay Concerning Human Understanding* and being the physician to the first Earl of Shaftsbury is a contingent property of John Locke, and hence not one he possesses in all possible worlds.

Thus we can clearly demonstrate why Kripkean semantics do not fall foul of the three objections to the descriptivist. On Kripke's story although the name 'John Locke' may have been fixed using the description 'is the author of *An Essay Concerning Human Understanding*, and personal physician to the first Earl of Shaftsbury', they are not synonyms. As such (5) does *not* express an equivalent proposition to (7), and thus (5) is neither knowable *a priori* (the conclusion of the epistemological argument), nor is it necessary (the conclusion of the modal argument). Although John Locke may be individuated via the property of being the author of *An Essay Concerning Human Understanding*, and personal physician to the first Earl of Shaftsbury, it would be a mistake to *identify* Locke with this property.

But this still leaves the original problem of cognitive asymmetry that troubled Frege: by endorsing Mill's view of names, and arguing that the *meaning* of a name is simply the referent, it would seem, *prima facie* at least, that (CDR) will have a problem explaining why knowing the truth 'a = b' is a useful extension of our knowledge if 'a' and 'b' are co-referring names, and thereby contribute the same content (the referent) to a proposition.

### **1.5 A Solution to Frege's Puzzle?**

Frege's Puzzle, as discussed in the opening section, is a puzzle about the apparent cognitive asymmetry of co-referring names. The puzzle was aimed specifically at an account of the meaning of names which subscribed to the view that the meaning of a name is just that to which it refers. As Kripkean semantics endorses such a view it seems to invite the problem all over again: if two names 'a' and 'b' denote the same referent, then the identity statement 'a = b' should be no more an extension of our knowledge than 'a = a', because if they have the same denotation, and all there is to the meaning of a name is its referent, then they also

express the same proposition. However, 'a = b' does seem to extend our knowledge, so there is some puzzle as to why it does.

Kripke's solution to the problem, and his view on identity, is clear from his discussion of Hesperus and Phosphorus. The statement 'Hesperus is Hesperus' is an identity statement that is analytic and its truth is knowable *a priori*, whereas the identity statement 'Hesperus is Phosphorus' is, if true, knowable only *a posteriori*, as no amount of reflection on either of the terms reveals the truth of the identity claim. The point is that while 'Hesperus' names the morning star, and 'Phosphorus' the evening star, both the morning and evening stars respectively turn out to be a single heavenly body, the planet Venus. That the morning star and the evening star are Venus remains hidden from view until we engage in empirical investigation. However, once we have found out that there are not, in fact, two heavenly bodies here, but one which occupies different positions at different times of year, we also discover that the terms we used to pick out these two heavenly bodies are in fact co-referential. If all names are rigid designators, then it follows that identity statements containing two names which are co-referential, such as 'Hesperus is Phosphorus', are, if true in this world, true in all possible worlds. In other words, the truth of the identity statement 'Hesperus is Phosphorus', the rigidity of the names 'Hesperus' and 'Phosphorus', and the necessity of identity entail that *necessarily* Hesperus is Phosphorus. Furthermore, since knowing the truth of the identity statement takes some empirical investigation, it turns out that there are some necessary truths that are knowable only *a posteriori*.

Kripke's answer to the problem of cognitive asymmetry, then, stems from the wedge he drives between the notion of identity and the semantic fact that, say, 'Hesperus' and 'Phosphorus' are co-referential. So, while the identity statement 'Hesperus is Phosphorus' is *necessary* because Hesperus is Phosphorus – there is only a single referent, and necessarily

everything is self identical – that the two names are in fact co-referential is not immediately obvious from reflection on the terms. In order to find out that the two names are co-referential we need to engage in some empirical investigation, tracking back two distinct causal chains of reference to two unique baptismal events. That this investigation is a prerequisite of judging whether the identity statement ‘Hesperus is Phosphorus’ is true demonstrates that the identity statement lacks epistemic necessity: it is not knowable *a priori*.

The important distinction, then, is between two types of necessity: metaphysical and epistemic. The identity statement ‘Hesperus is Phosphorus’ is metaphysically necessary – true in all possible worlds – given that it is true in this world, that ‘Hesperus’ and ‘Phosphorus’ are both rigid designators, and that everything is necessarily self identical. However, it is not epistemically necessary since mere reflection on the terms does not reveal the truth of the identity statement. By contrast the identity statement ‘a bachelor is an unmarried man’ is both metaphysically and epistemically necessary: it is true in all possible worlds, and its truth is knowable merely by reflecting on the terms ‘bachelor’ and ‘unmarried man’ respectively. The latter case, then, is an example of the necessary *a priori*, whereas the former is an example of the necessary *a posteriori*.

The crux of the distinction is that in the case of some *a posteriori* necessities ‘we can say that under appropriate qualitatively identical evidential situations, an appropriate corresponding qualitative statement might have been false’ (Kripke 1980: 142). Kripke illustrates this distinction using the Hesperus/Phosphorus case as follows: If ‘H’ and ‘P’ are rigid designators then if ‘H = P’ is true it is also necessary. However, imagine further that the references ‘H’ and ‘P’ had been fixed using the *non-rigid* descriptions ‘D<sub>H</sub>’ and ‘D<sub>P</sub>’ respectively i.e. ‘the heavenly body in such-and-such a position in the sky in the evening

(morning)' (Kripke 1980: 143). Although 'H = P' is necessary, 'D<sub>H</sub> = D<sub>P</sub>' is contingent, and thus Kripke says:

The inaccurate statement that Hesperus might have turned out not to be Phosphorus should be replaced by the true contingency mentioned earlier in these lectures: two distinct bodies might have occupied, in the morning and the evening, respectively, the very positions actually occupied by Hesperus-Phosphorus-Venus. (Kripke 1980: 143)

Thus although the fact that Hesperus *is* Phosphorus guarantees the metaphysical necessity of the identity statement 'Hesperus is Phosphorus' it does not guarantee epistemic necessity, and this is sufficient to explain the cognitive asymmetry of the identity statement 'Hesperus is Phosphorus' in comparison to the identity statement 'Hesperus is Hesperus', given that both express the very same proposition.

## 1.6 Natural Kinds

[M]y argument implicitly concludes that certain general terms, those for natural kinds, have a greater kinship with proper names than is generally realized. This conclusion holds for certain for various species names, whether they are count nouns, such as 'cat', 'tiger', 'chunk of gold', or mass terms such as 'gold', 'water', 'iron pyrites'. (Kripke 1980, p. 134)



According to Kripke his thesis about proper names is ‘supposed to extend to theoretical identity statement like “Gold = the element with atomic number 79” and “Water = H<sub>2</sub>O”’ (LaPorte 2004: 36). The basic line of argument is that ‘terms for natural kinds are much closer to proper names than is ordinarily supposed’ (Kripke 1980: 127), and that ‘[t]he old term ‘common name’ is thus quite appropriate for predicates marking out species or natural kinds, such as “cow” or “tiger”...[and]... mass terms for natural kinds, such as “gold”, “water” and the like’ (*ibid.*). If natural kind terms are a species of name then given that ‘names are always rigid designators’ (Kripke 1980: 58) we should be able to derive an analogue of Kripke’s conclusion concerning proper names – that identity statements can be necessary but, nevertheless, knowable only *a posteriori* – for natural kind terms.

Thus Kripke introduces a sub-species of identity statement: the *theoretical* identity statement. A theoretical identity statement is a statement ‘representing [a] scientific discovery about what this stuff *is*’ (Kripke 1980: 125), where *this stuff* is some natural kind. These new identities are ‘identities between terms for substances, and also the properties of substances and of natural kinds’ (Kripke 1980: 166), and are ‘generally identities involving two rigid designators and therefore are examples of the necessary *a posteriori*’ (Kripke 1980: 140). Kripke presents various examples of these ‘types of identity statement expressing scientific discoveries’ (Kripke 1980: 128) including ‘light is a stream of photons’, ‘water is H<sub>2</sub>O’, ‘lighting is electrical discharge’ and ‘gold is the element with atomic number 79’. The basic point is the same in all cases: if terms for natural kinds are analogous to names, and all names are rigid designators, then theoretical identifications will be, if true, necessary. As such, at least according to Kripke, ‘statements representing scientific discoveries about what this stuff *is* are not contingent truths but necessary truths in the strictest possible sense’ (Kripke 1980: 125).

Consider, by way of illustration, the theoretical identification sentence ‘tungsten is the element with atomic number 74’. If we apply Kripke’s story about proper names to this case we might say something like the following: the theoretical identity ‘tungsten is the element with atomic number 74’ is a species of identity containing two natural kind terms ‘tungsten’ and ‘the element with atomic number 74’ respectively. Each natural kind term, being analogous to a proper name, is also a rigid designator, which is to say that it designates that same thing in all possible worlds. After some empirical (scientific) investigation it has been discovered that the identity sentence ‘tungsten is the element with atomic number 74’ is true. The conjunction of the true identity sentence, both ‘tungsten’ and ‘the element with atomic number 74’ being rigid designators, and the necessity of identity, is sufficient to guarantee that tungsten *is* the element with atomic number 74 in all possible worlds. In other words, it turns out to be necessary that tungsten is the element with atomic number 74, but nevertheless *a posteriori*.

As in the proper names case Kripke distinguishes cases where the reference of the rigid designator is fixed using a *non*-rigid designator which gives an illusion of contingency. The contrast in Kripke’s discussion is between water, and so called fool’s water, i.e., polywater<sup>5</sup>. According to Kripke’s (intuitive) account, when we originally identified the natural kind water we did so via its characteristic properties, its ‘feel, appearance and perhaps taste’ (Kripke 1980: 128). Much later we made the discovery that water is H<sub>2</sub>O. But now consider two distinct scenarios: Firstly, imagine an actual substance with a ‘completely different atomic structure from that of water’, but which resembles water in all characteristic

---

<sup>5</sup> Polywater was hypothesised during the 1960’s by Nikolai Fedyaikin, and later by Boris Derjaguin (who called it ‘anomalous water’). However, this polymerised water was later found, by Denis Rousseau, *not* to be polymerised water at all, but just water containing certain impurities.

respects (*ibid.*); secondly, imagine a substance with the same atomic structure as water, but with ‘very different identifying marks from that of what we now call water’ (*ibid.*: 129). Kripke’s view is that in the first scenario, the substance with the identifying marks of water, but a different atomic constitution, is *not* water. Rather, ‘just as there is a fool’s gold there could be a fool’s water’ (*ibid.*: 128). Once we have discovered that water *is* H<sub>2</sub>O, nothing lacking that atomic structure could be water. In the second scenario the situation is the same, regardless of what identifying marks the substance happens to have, if it is H<sub>2</sub>O then it is water. The description of the appearance of water, by which we might have originally identified it, may well have helped us fix the reference of the natural kind term ‘water’, but it was a *non-rigid* designator. These considerations are, in effect, the analogues of those found in Kripke’s three arguments against the Fregean.

Consider the following three sentences:

(8) Tungsten, if it exists, is the silvery-white, lustrous metal found in China.

(9) If anything is the silvery-white, lustrous metal found in China then it is tungsten.

(10) Something is the silvery-white, lustrous metal found in China if and only if it is the silvery-white, lustrous metal found in China.

If, when we *identify* tungsten, we treat the identifying description as the *meaning* of ‘tungsten’ (in Frege’s terminology, as the *sense*), then from (8) and (9) we can derive (10). As in the proper names case this valid substitution has three counterintuitive consequences. Firstly, since (10) is *necessary*, and both (8) and (9) express the same proposition – given that ‘tungsten’ is synonymous with ‘the silvery-white, lustrous metal found in China’ – then (8) and (9) are *also* necessary (the modal argument). Secondly, since the truth of (10) is knowable *a priori*, and both (8) and (9) express the same proposition, they too are knowable *a priori* (the epistemic argument). Finally, as whatever satisfies the description ‘the silvery-white, lustrous metal found in China’ *is* tungsten then were there, in the actual world, to be some *other* substance which uniquely satisfied this description then *it* would be the referent of the natural kind term ‘tungsten’. Thus if gold satisfied the description ‘the silvery-white, lustrous metal found in China’ then it would be the referent of the natural kind term ‘tungsten’ (the semantical argument).

Kripke’s response is evident from the discussion of water earlier. It is *not* a metaphysically necessary property of tungsten that it be the silvery-white, lustrous metal found in China. While it might be true that the properties of being silvery-white, lustrous and found in China were used to fix the reference of ‘tungsten’ the description ‘the silvery-white, lustrous metal found in China’ is a *non-rigid* designator, and as such even if (10) happens to be true it is *not* necessarily so. Secondly, nothing about the meaning of ‘tungsten’ appears to preclude something to which the natural kind term truly applies *lacking* any one (or even all) of the properties listed in (10). The meaning of ‘tungsten’ does not tell us what properties tungsten has; finding tungsten’s properties is a matter of empirical investigation. In effect, any theory of meaning and reference that treats ‘tungsten’ and ‘the silvery-white, lustrous metal found in China’ as synonyms is committed to the necessity of obviously contingent properties

(the modal argument), and the *a priori* of identity statements that are *a posteriori* (the epistemological argument).

The *a posteriori* of theoretical identifications such as ‘tungsten is the element with atomic number 74’ plays a positive role in Kripke’s alternative to a description theory of meaning and reference in that it explains why there is a cognitive asymmetry between theoretical identities like (10) and ‘tungsten is tungsten’, despite both expressing the same proposition. Finally, an application of Kripke’s story about how name-acquiring transactions takes place, despite being an ‘as-if’ story, gives us a clear picture of why the properties an object contingently has play either (a) no role, or (b) a minimal role as reference fixers, in the naming of natural kinds.

### **1.7 Puzzles**

There are a number of puzzles that arise from Kripke’s application of his semantics of proper names to natural kinds. The first and most obvious question begins with the observation that many of Kripke’s examples of theoretical identification sentences seem, quite obviously, to contain descriptions. For example the theoretical identification sentence ‘tungsten is the element with atomic number 74’ (and any other theoretical identification for an element on the periodic table) looks to contain an *attributive* description i.e. ‘the element with atomic number 74’. If this observation is correct then we require some explanation of why a description like ‘the element with atomic number 74’ can be a rigid designator whereas ‘the silvery-white, lustrous metal found in China’ cannot, given that there is no significant grammatical difference between them.

This observation, however, is part of a broader question concerned with precisely how we should view the semantics of natural kind terms, and in that respect it is not new. There are two popular opposing views on the market.

The first, defended by Scott Soames (cf. Soames 2002), argues that Kripke did *not* give us an adequate account of the rigidity of natural kind terms. Soames argues that natural kind terms are most naturally treated as predicates, rather than general names. Furthermore, the ‘notion of rigidity to characterize the semantic similarities between natural kind predicates and proper names was unsuccessful’ (Soames 2002: 264) leaving a significant lacuna in Kripke’s thesis. That is not to say that natural kind predicates are not rigid – on Soames’ view they are indeed rigid. But the rigidity of predicates, and the semantic similarity that grounds an analogy between names and natural kind terms, stems from the more central thesis that names and natural kind terms are non-descriptonal. As Soames says, when setting out his investigation, ‘the important parallel between names and natural kind predicates may be their nondescriptonality, and the way their reference is fixed’ (Soames 2002: 265)<sup>6</sup>. *Prima facie* the claim that natural kind predicates are nondescriptonal is odd – after all, don’t predicates ascribe properties to objects, and thus receive a fairly natural treatment as descriptonal?

Soames’ view is that natural kind predicates are introduced in much the same way as proper names – that is, ‘either ostensively or via a description’ (Soames 2002: 267). When natural kind predicates are introduced, in contrast to proper names, typically a speaker will ostend a sample of individuals that are presumed ‘to be instances of a single natural kind’ (*ibid.*). With proper names, only a single individual is ostended. However, this does not

---

<sup>6</sup> Although as we shall see in chapter 3, Soames (2002) is not consistent in his treatment of natural kind predicates as nondescriptonal, claiming explicitly that certain natural kind terms are synonymous with descriptions.

constitute a metasemantic difference, merely a practical one. Moreover, when natural kind terms are introduced, although speakers might associate descriptive properties with the term, the extension of that natural kind term is determined independently of any description that a speaker happens to associate with it (*ibid.*: 266). Hence, according to Soames at least, ‘a speaker can successfully use a natural kind predicate to say something about members of a kind, even if he lacks the ability to accurately describe either the kind itself or its instances’ (*ibid.*).<sup>7</sup>

The second view, defended by Nathan Salmon (originally published in 1982, all references to 2005 edition), treats natural kind terms as straightforwardly analogous to proper names, and argues that Kripke ‘intended his notion of rigidity to apply to general as well as singular terms’ (Salmon 2005: 384). Salmon agrees with Soames that ‘Kripke’s thesis that proper names and certain general names alike, including natural kind terms, are rigid designators is secondary to a more fundamental thesis: that these names are *nondescriptional*’ (Salmon 2005: 386). However, he thinks that theoretical identifications such as ‘water = H<sub>2</sub>O’ are analogous to identities like ‘Hesperus = Phosphorus’ in that both contain names, albeit of different types.

There is, then, a decision to be made as to whether natural kind terms are (a) predicates (Soames), or (b) general names (Salmon); and this question of the semantics of natural kind terms will have important ramifications for the necessity of theoretical identifications. If we decide that natural kind terms *are* predicates, then we require an explanation of the necessity of theoretical identities such as ‘tungsten is the element with

---

<sup>7</sup> It should be noted, however, that Soames, initially at least, is not particularly clear about what the denotation of a natural kind predicate is; is it the kind or is it the extension? I discuss this issue in detail in chapter 3.

atomic number 74'. If we decide that natural kind terms are general names *rather* than predicates we need an argument as to why, given that general terms are standardly thought of as predicates. I will discuss (a) and (b) in turn in chapters 3 and 4 respectively.

There is also a question as to whether or not rigidity really is a corollary of non-descriptiveness. *Prima facie*, or so I will argue, this seems *not* to be the case. When Kripke gives his story about gold he says that 'we discover that certain properties were true of gold in addition to the initial identifying marks by which we identified it' (Kripke 1980: 119) and further that we have discovered that 'it is part of the nature of gold as we have it to be an element with atomic number 79' (Kripke 1980: 125). He goes on to say that:

[A] material object is (pure) gold if and only if the only element contained therein is that with atomic number 79. Here the 'if and only if' can be taken to be strict (necessary). In general, science attempts, by investigating basic structural traits, to find the nature, and thus the essence (in the philosophical sense) of the kind. (Kripke 1980: 138)

This suggests that the phrase 'the element with atomic number 79' is being used as a description. After all, Kripke claims that *having* atomic number 79 is the *essence* of gold, and that science discovered this property by investigating basic (presumably atomic) structural traits. Indeed he even claims that 'science can discover empirically that certain properties are necessary' (Kripke 1980: 128). If 'the element with atomic number 79' is in fact a description, and, following Kripke, a rigid designator, then it looks like some descriptions can be rigid, but as a corollary of the properties they pick out being necessary. The contrast is with the more



obviously name-like natural kind terms, for instance ‘gold’ or ‘tungsten’, which are supposed to be rigid as a product of how they were introduced (their metasemantics) and their nondescriptiveness.

This will create problems for both Salmon’s and Soames’ account, albeit for different reasons. Salmon’s account treats ‘the element with atomic number 79’ as a ‘logically proper general name’ (Salmon 2005: 396/7), whereas Soames argues that ‘the element with atomic number 79’ is a description of sorts. *Prima facie*, then, Soames’ account looks better than Salmon’s, given Kripke’s reference to ‘the element with atomic number 79’ as the *essence* of a kind. However, as both Soames and Salmon agree that rigidity is a corollary of non-descriptiveness (which, in turn, raises a question for Soames concerning how natural kind terms can be descriptive *and* rigid) if phrases like ‘the element with atomic number 79’ are descriptions and are rigid, then it looks like both accounts have serious faults.

This last issue also has some significant metaphysical implications. If the rigidity of certain terms is the product of the *essential* properties to which they refer, then there is a question as to precisely what role *essentialism* is playing in Kripke’s semantic thesis.<sup>8</sup> Nathan Salmon (2005) has argued forcefully that essentialism cannot be *derived* from semantics – we will look at this in more detail in Chapter 4 – but Kripke claims on various occasions that ‘the *origin* of an object is essential to it’ (Kripke 1980: 114 n57), that a belief in ‘essentialism, [is] the belief in modality *de re*’ (Kripke 1980: 39) and that statements representing scientific discoveries are ‘necessary in the non-epistemological sense’ (Kripke 1980: 123) i.e. *metaphysically* necessary. If theoretical identifications ‘resemble “Cicero is a product of the

---

<sup>8</sup> I am, until chapter 6, when I discuss the notion of ‘real essence’, and chapter 7, where I distinguish between brands of essence and essentialism, simply going to appeal to the same essentialism implicit (and sometimes explicit) in Kripke.

egg and sperm that generated him” rather than “Cicero = Tully” in respect that they aim to expose essence’ (LaPorte 2004, p. 37) then it looks like (i) essences are what make theoretical identifications metaphysically necessary (which, when combined with the empirical investigation required to ascertain their truth, makes them part of the category of the necessary *a posteriori*), and (ii) that the proper names case and the natural kinds case are actually rather disanalogous.

Finally, it is worth noting (briefly) some of the issues I’m *not* going to talk about. Unlike some authors (Besson 2010; López de Sa 2007; Martí 2004; Martí and Martínez-Fernández 2010; Schwartz 2002; Wikforss 2010) I am not explicitly interested in whether or not there is a *special* category of natural kind terms, as compared to other general terms. Indeed, for the most part (other than a criticism of Salmon’s view in Chapter 4, which appears to entail that there is *not* a special category) I simply assume that there is. Similarly, although I will discuss the notion of natural kind terms as predicates in Chapter 3, I am not going to engage with the recent debate concerning *essentialist* predicates, and whether or not they capture Kripkean intuitions (cf. Gómez Torrente 2006; Ahmed 2009).

I am also side stepping the debate concerning proper names, and whether or not descriptivism about proper names is tenable in light of Kripke’s criticisms (cf. Hughes 2004 for the case that it is untenable, and Dummett 1981 for the case that it is). Although my interest here is in part semantic, I am only concerned with the semantics of natural kind terms, and do not take anything I say here to automatically apply to names.

Likewise, I am also avoiding at least one way that the debate over the tenability of descriptivism has developed, namely two-dimensional semantics. Although I recognise that some issues in that debate reflect on some of what I say here – for instance, that one way to

get a broadly Fregean distinction between sense and reference again is to adopt the two-dimensional framework proposed by, say, David Chalmers (1996) – my overarching interest is with demonstrating that there is no interesting connection between metaphysical essentialism and the semantics of natural kind terms. Hence, when I defend descriptivism in Chapter 5, and later modify it in Chapter 8, I am, to all intents and purposes, ignoring contemporary developments in two-dimensional semantics (cf. Chalmers 1996, and Soames 2006 for a response to the project). However, in part my decision to ignore this contemporary development is motivated by my confidence that with the metaphysics appropriately isolated from the semantics, a much simpler version of descriptivism is available (more on this in Chapter 5). For now, though, I move on to a discussion of Hilary Putnam, and his thesis concerning the meaning and reference of natural kind terms.

## CHAPTER 2

Before unravelling the puzzles raised at the end of chapter 1, I will first present Hilary Putnam's account of the semantics of natural kind terms, developed independently but concurrently with Kripke's. Putnam, like Kripke, had an axe to grind with the descriptivist concerning the meanings of names and natural kind terms. However the majority of Putnam's arguments are addressed towards natural kind terms directly, rather than to proper names and then, by extension, to natural kind terms. As such, Putnam's thesis is of particular relevance to this investigation. In this chapter I begin, in section 1, by sketching Putnam's account of the semantics of natural kind terms, and show, through sections 2, 3 and 4, how his conclusions concerning the necessity of theoretical identities are the same as Kripke's. However, in section 5 of this chapter I argue that Putnam's account is markedly different to Kripke's in its remit. In section 6 I present a recent challenge from LaPorte concerning the modal intuitions we apparently have in Twin Earth style thought experiments. Finally, in section 7, I distinguish between early and later Putnam, and point out that while Putnam's early account of natural kind terms endorses the necessity of theoretical identity sentences, later Putnam rejects *de re* metaphysically necessary theoretical identities.

### **2.1 Putnam's semantics of natural kind terms**

Putnam's discussion of the semantics of natural kind terms begins with the claim that there is a (traditional) ambiguity in the notion of 'meaning', namely that it is the conjunction of two distinct concepts: 'intensions' on the one hand, and 'extensions' on the other. Consider, by

way of illustration, the term 'nickel'. The extension of 'nickel' is precisely the set of things that the term is true of, while the intension of the term is what we more commonly refer to as the meaning i.e. a concept associated with the term, grasped by an individual psychological act.

Putnam notes that while there are 'timeworn examples' (Putnam 1975: 218) where two terms have the same extension (and so the same meaning in one sense of 'meaning'), but have obviously different intensions (and so a different meaning in another sense of 'meaning'), there were no examples (at the time) where two terms had the same intension, but nevertheless possessed different extensions. The standard example of the former, which Putnam refers to, is the renate/cordate case. The general term 'cordate' (defined as a 'creature with a heart') has the same extension as the general term 'renate' ('creature with a kidney') since every creature with a heart is also a creature with a kidney. But clearly the intension of 'cordate' is different from the intension of 'renate' since the concept of having a heart is different from the concept of having a kidney. This, says Putnam, highlights an important assumption implicit in traditional theories of meaning and reference, namely that the intension of a term 'must *always* provide a necessary and sufficient condition for falling into the extension of the term' (Putnam 1973: 700), and that while two terms with different intensions can have the same extension, no two terms with different extensions can have the same intension.

Reflecting on this type of example, Putnam identifies two unchallenged presuppositions that traditional theories of meaning had come to rest upon:

- (i) to know the meaning of a term is to be in a particular psychological state,

and,

(ii) the intension of a term determines the extension of that term.

Putnam's contention is that these two conditions cannot be satisfied by any notion, and *a fortiori* cannot be satisfied by the notion of meaning. According to the dominant tradition in theories of meaning before the Kripke-Putnam revolution, to understand the meaning of a term like 'tungsten' consisted in the grasping, via a psychological state, of some public meaning (or Fregean 'sense', or intension). However, these traditional theories of meaning also assumed that 'no psychological state, properly so called, presupposes the existence of any individual other than the subject to whom that state is ascribed' (Putnam 1975: 220).

Psychological states, then, were construed as fairly solipsistic sorts of states, having only *narrow content*. That is, the content of the psychological state does not depend upon the environment within which the individual having it is located.

The contrast is with 'psychological state[s] in the wide sense' (*ibid.*). That is, mental states with *broad content* – where the content of the mental state depends both upon the individual having the mental state, and on the environment within which the individual happens to be situated. The wide/narrow content distinction is important because, according to Putnam, assumption (i) amounts to a claim about narrow content, namely that knowing the meaning of some term *A* is to grasp the intension of *A*, and to know that the concept one has grasped *is* the intension of *A*. More simply, assumption (i) is the claim that meanings are in the head, and that to know the meaning of a term is to be in the narrow content mental state 'knowing that *I* is the meaning of *A*, where *I* is an 'intension' and *A* is a term' (Putnam 1975: 221). If meanings are in the head, then an identical set of necessary and sufficient conditions

for falling within the extension of a term are supplied whenever (and in whatever possible world) there is an individual in that (narrow content) psychological state. The narrow content of the mental state determines the intension  $I$  of the term  $A$ , and the intension  $I$  determines the extension, which is the set of individuals to which the term  $A$  truly applies.

Putnam claims, however, that the intension of a term does not determine the extension, and that we can construct examples where two typical speakers,  $s_1$  and  $s_2$ , are in the same narrow psychological state, but where the extension of the term  $A$  in  $s_1$ 's idiolect is different from the extension of  $A$  in  $s_2$ 's idiolect. Putnam's (in)famous illustration of this claim – that extensions are not determined by the psychological states of particular speakers – is the well discussed Twin Earth thought experiment. There have been a few variations, but the central theme is as follows: imagine a planet that is a molecule by molecule duplicate of Earth, except that on this distant planet, call it Twin Earth, what speakers refer to as 'water' is not  $H_2O$ , but a liquid with a long and complicated chemical formula, abbreviated (for simplicity) to XYZ. So, on Twin Earth it is XYZ, and not  $H_2O$ , that falls from the sky, and flows in the lakes and rivers.

Now consider a typical language user on Earth, call her  $Sally_1$ , who refers to stuff with the chemical composition  $H_2O$  when she uses the term 'water', and consider her Twin Earth counterpart,  $Sally_2$ , who uses the syntactically identical term 'water', but picks out some other stuff, with the chemical formula XYZ. Since water on Earth is  $H_2O$ , and water on Twin Earth is XYZ we can say that on Earth 'water' refers to  $H_2O$ , and that on Twin Earth 'water' refers to XYZ. However, although the extension of water on Earth is different to that on Twin Earth, the example stipulates that in all other ways Twin Earth and Earth are precise duplicates. This means that  $Sally_1$  and  $Sally_2$  are intrinsically exactly alike, and as such there is no belief that  $Sally_1$  has about water that  $Sally_2$  does not have about what she calls 'water' (as typical

language users, neither Sally<sub>1</sub> nor Sally<sub>2</sub> know anything about chemical composition). But the extension of 'water' on earth is H<sub>2</sub>O, not XYZ (whether anyone knows it or not). But if this is correct then (ii) must be false, since both Sally<sub>1</sub> and Sally<sub>2</sub> are in identical psychological states, but nevertheless the extension of 'water' on Earth is not the same as that on Twin Earth. To put it another way, the identical mental states of Sally<sub>1</sub> and Sally<sub>2</sub> respectively, underdetermine the reference of 'water', since the extension of 'water' on Earth is a different *kind* of stuff to the extension of 'water' on Twin Earth.

That water is H<sub>2</sub>O, however, is (as Putnam notes) a fairly recent discovery made the empirical sciences. Thus consider a scenario in which we 'roll the time back to about 1750' (Putnam 1973: 701), to where nobody knows what water is made of, and so no one associates water with H<sub>2</sub>O (or, if you're on Twin Earth, XYZ). If, as in the previous example, 1750-Sally<sub>1</sub> and 1750-Sally<sub>2</sub> are in identical psychological states, being, as they are, exact duplicates, then is the extension of 'water' on Earth really distinct from its extension on Twin Earth? According to Putnam the answer is yes: 'the extension of the term "water" was just as much H<sub>2</sub>O on Earth in 1750 as in 1950' (*ibid.*). Putnam's claim is that although Sally<sub>1</sub> and Sally<sub>2</sub> are in identical psychological states, the reference of 'water' is still different, even in 1750, since water was *always* H<sub>2</sub>O on Earth, and *always* XYZ on Twin Earth (despite no one in 1750 knowing that there was such a difference). The discovery on Earth that water is H<sub>2</sub>O, then, is both a discovery about the kind water, and the discovery, at least in our fictional example, that 'water' on Earth does *not* mean the same thing as 'water' on Twin Earth – for Putnam, recall, the notion of meaning is the conjunction of intension on the one hand, and extension on the other. Hence, since the extension of water on Earth is different to that on Twin Earth, 'water' does not mean the same thing on Earth as it does on Twin Earth. Of course, before the relevant scientific discovery (about the constitution of water) no language



user could have articulated (nor, indeed, would they have asserted) the difference in extension, and hence meaning, of ‘water’ as it was used on Earth, call it ‘water<sub>E</sub>’, compared to how it was used on Twin Earth, call it ‘water<sub>TE</sub>’. Nevertheless, the meanings were always quite distinct, whether speakers knew it or not.

What the Twin Earth thought experiment allegedly shows is that assumptions (i) and (ii) are false. Firstly, it purports to show that despite Sally<sub>1</sub> and Sally<sub>2</sub> being in identical psychological states, ‘water’ nevertheless refers to H<sub>2</sub>O on Earth, and XYZ on Twin Earth, and since H<sub>2</sub>O is not XYZ, that the psychological states of Sally<sub>1</sub> and Sally<sub>2</sub> underdetermine the reference of ‘water’. In other words, the conclusion is that rather than intension, or meaning, being determined by extension, extension determines, or rather *is*, meaning; thus assumption (ii) is false. Secondly, the example also shows that assumption (i) is false: even in 1750, when the facts about water on Earth were not known, there was still some physical fact of the matter that determined what it was for a substance to be water. Since the physical facts about the substance called ‘water’ on Earth are different to those on Twin Earth, but Sally<sub>1</sub> and Sally<sub>2</sub> have identical sets of beliefs about what they call ‘water’, to know the meaning of ‘water’ cannot consist of being in a certain psychological state. Thus assumption (i), like assumption (ii), is false.

The Twin Earth thought experiment is, in essence, a version of Kripke’s semantical argument – both purport to demonstrate that descriptions *always* underdetermine reference<sup>9</sup>. Where the two arguments differ is that Putnam’s version is explicitly directed towards the internalist assumption contained within descriptivism: the claim that meanings are the sorts of things we grasp via our psychological states. Nevertheless, the basic moral is the same:

---

<sup>9</sup> I argue, in chapter 5, that this conclusion is false.

descriptivist theories of meaning and reference are untenable. However, it is not just Putnam's objections to descriptivism that are similar to Kripke's, but his positive story also has much in common with, and can rightly be characterised as a version of, the causal theory of direct reference (*CDR*). In the next section we shall consider Putnam's metasemantic story, which, like Kripke's, has some dramatic consequences for cross-world reference and necessary truth.

## 2.2 Cross-world reference

Putnam, like Kripke, claims that the reference of natural kind terms is fixed via an 'ostensive definition' (Putnam 1973: 701). However, more explicitly than Kripke Putnam notes that the ostensive definitions of natural kind terms come with important 'empirical presuppositions: that the body of liquid I am pointing to bears a certain sameness relation (...) to most of the stuff I and other speakers in my linguistic community have on other occasions called "water"' (*ibid.*). Putnam cashes out the sameness relation as a *same liquid* relation, given his water example, but nothing significant is supposed to hang on the example *per se*. So, we can simply generalise the relation to a sameness of natural kind relation – *same<sub>NK</sub>*. The nub of the relation is that when we consider two particular substance samples, *x* and *y*, *x* and *y* are of the same kind if and only if *x* and *y* bear the relation *same<sub>NK</sub>* to one another. Similarly, where *x* is on Earth and *y* is on Twin Earth, the question as to whether *x* and *y* are samples of the same kind is, again, decided by whether or not *x* and *y* bear the *same<sub>NK</sub>* relation to one another. If they do (in either case), then they are samples of the same kind; if they do not, then they are *not* of the same kind. In effect, the 'necessary and sufficient condition for being [of the same kind as *x*] is bearing the relation [*same<sub>NK</sub>*] to the stuff' (*ibid.*) that was originally ostended and

named within our environment. Moreover, this  $same_{NK}$  relation is both an intra-world and an inter-world relation.

Consider, to illustrate, the relation of *being faster than* (or just *faster*). Let  $x$  be Usain Bolt, and let  $y$  be Linford Christie, and let their personal bests (for the 100 metre sprint) be 9.58 seconds and 9.87 seconds respectively. Thus we can say that  $x$  bears the relation of being *faster* to  $y$ , since  $x$  ran the 100 meters in 9.58 seconds, rather than 9.87 seconds. Now, imagine that Usain Bolt is on Earth and that Linford Christie is on Twin Earth. Is Bolt still faster than Christie? The answer, obviously, is yes, since the *faster* relation holds not only within a world, but across worlds. Now, had Christie run the 100 metres in 9.58 seconds, then he would bear the relation of *being as fast as* to Bolt, and just as the *faster* relation was a cross-world relation, so too is the *being as fast as* relation. By analogy, the  $same_{NK}$  relation is both an inter-world and an intra-world relation: if two objects  $x$  and  $y$  bear the relation  $same_{NK}$  to one another, then no matter where they are located (Earth or Twin Earth), provided they bear that relation to one another they will qualify as members of the same kind.

Since natural kind terms only refer to members of the relevant kind, and what it is to be a member of such-and-such a kind is dictated by bearing an appropriate physical sameness relation to some paradigm instance (an instance which is taken to be, with some degree of certainty, representative of the kind), the meaning and reference of natural kind terms is intimately connected with the members of natural kinds within the environment. On Putnam's model, when a natural kind term has been introduced by ostending some sample with which we are acquainted, the term can only truly apply to all those things which are suitably similar to the originally ostended sample, the archetypal member (or paradigm) of the kind. The account does not require speakers to know what this sameness amounts to, just to introduce the term with the empirical presupposition that there is some appropriate sameness relation.

Hence, there will be some physical fact of the matter, which will take more or less empirical investigation to discover, that determines whether or not a natural kind term has been correctly applied, once the initial introductions have been made. Further, this hidden fact of the matter ‘determines what it is to be a member of the natural kind, not only in the actual world, but in all possible worlds ... it determines what we can and cannot counterfactually suppose about the natural kind’ (Putnam 1975: 241).

The cross-world force of Putnam’s account begins by giving a meaning explanation of a particular natural kind term by ostending samples within the environment. Using Kaplan’s (1978) demonstrative version of ‘that’, *dthat*, the ostensive definition might be ‘*dthat* is water’. Now, consider two possible worlds: the actual world,  $w_1$ , and some other possible world,  $w_2$ . Putnam offers us two alternative theses about the meaning of ‘water’. Either, ‘water’ is ‘*world-relative* but *constant* in meaning’ (Putnam 1996: 16) such that the word ‘water’ means the same thing in  $w_1$  and  $w_2$ , where water is  $H_2O$  in  $w_1$  and is XYZ in  $w_2$ ; or, alternatively, water is  $H_2O$  in all possible worlds, and ‘water’ means one thing at  $w_1$  and something else at  $w_2$ . The Twin Earth thought experiment is intended to convince us of the latter thesis: ‘water’ means  $H_2O$ , since that is what water has turned out to be at the actual world: the relevant physical sameness relation consists in having the same molecular composition as the stuff with the composition  $H_2O$ . Hence, since XYZ is not  $H_2O$ , ‘water’ means something else on Twin Earth.

The nub of the thesis is that there is an indexical component to the ostensive definition that fixes the ostensive definition to the actual world, rather than any possible world. Putnam characterises the difference between the two options as one of scope, by formalising the two theses as follows:

(A) (For every world  $w$ ) (For every  $x$  in  $w$ ) ( $x$  is water  $\leftrightarrow x$  bears the relation  $same_{NK}$  to the entity referred to as ‘this’ in  $w$ ).

(B) (For every world  $w$ ) (For every  $x$  in  $w$ ) ( $x$  is water  $\leftrightarrow x$  bears the relation  $same_{NK}$  to the entity referred to as ‘this’ in the actual world  $w_1$ ). (Putnam 1996: 16)

In (A) the ostensive definition of the natural kind term is within the scope of the phrase ‘for every world  $w$ ’, whereas in (B) the ostensive definition is only within the scope of the *actual* world. In other words, as Putnam says, in thesis (A) the ostensive definition is within the scope of the bound variable, whereas in (B) the definition has ‘a reference *independent* of the bound variable “ $w$ ”’ (1996: 16).

As far as Putnam is concerned, that we intend (B) rather than (A) when we ostensively define out natural kinds indicates that we intend our natural kind terms to be what Kripke called *rigid designators*. Thus he says that if ‘we extend the notion of rigidity to substance names, then we may express Kripke’s theory and mine by saying that the term “water” is *rigid*’ (Putnam 1996: 16). The term ‘water’ designates  $H_2O$  in the actual world, and gets to refer to  $H_2O$  in all possible worlds since it was introduced with the empirical presupposition that all samples of the kind would bear some appropriate sameness relation to one another, and science has discovered that the property grounding that relation for water is being constituted of  $H_2O$  molecules.

### 2.3 *A posteriori* necessity

According to Putnam, the theory of meaning for natural kind terms that he and Kripke developed independently ‘has startling consequences for the theory of necessary truth’ (1996: 17). That consequence is that there are some truths which are necessary yet knowable only *a posteriori* (rather than *a priori*, as was traditionally supposed). Putnam’s explanation of this consequence runs as follows.

Firstly, we note that the theoretical relation  $same_{NK}$  (as stated above) is a cross-world relation: that is, for some object  $x$  in  $w_1$ , and some object  $y$  in  $w_2$ , if  $y$  has the same ‘important physical properties’ in  $w_2$  as  $x$  has in  $w_1$ , then  $y$  bears the relation  $same_{NK}$  to  $x$ , and thus  $x$  and  $y$  are of the same kind. So, whatever the ‘important’ properties of, say, water are in the actual world, for some object  $y$  in some arbitrarily selected possible world to be water, it must have the same important physical properties.

Whilst it might be debatable as to precisely what Putnam means by ‘important physical properties’, his discussion suggests a distinction between an operational definition (the description of some set of properties, which are standardly thought to be indicative of the kind e.g. falls from the sky, flows in the lakes and rivers) and a microstructuralist definition (the particular microstructure, e.g.  $H_2O$ , without which a sample of a substance could not be a member of the kind in question). The basic point is that provided we agree that the microstructure of a substance defines that substance, whether or not we know what the structure of a particular substance is, then we can agree that even in situations where a substance satisfies the operational definition, unless that substance has the relevant microstructure, it is of a different kind. To quote Putnam:

The operational definition, like the ostensive one, is simply a way of pointing out a standard – pointing out the stuff in the actual world *such that for x to be water, in any world, is for x to bear the relation [same<sub>NK</sub>] to the normal member of the class of local entities that satisfy the operational definition.* (Putnam 1996: 17)

The discovery that water is H<sub>2</sub>O, then, in conjunction with both the assumption that microstructure is what demarcates natural (substance) kinds, and the cross-world force of the *same<sub>NK</sub>* relation, amounts to the discovery that ‘*nothing counts as a possible world in which water isn’t H<sub>2</sub>O*’ (1996: 18). This effectively drives a wedge between the notion of conceivability and the notion of possibility, and as Putnam famously observed, conceivability turns out to be ‘no proof of logical possibility’ (*ibid.*). What this means is that whilst we might be in an epistemic situation such that *for all we know* water could be XYZ, since it is necessary that water is H<sub>2</sub>O (given that this is what water has turned out to be), this epistemically possible situation is not a metaphysically possible situation: since one cannot rule out *a priori* that water is not H<sub>2</sub>O, the truth that water is H<sub>2</sub>O is epistemically contingent, and hence knowable only *a posteriori*. However, since it turns out to be necessary that water is H<sub>2</sub>O, as a product of the indexical component of our ostensive definition and the empirical presuppositions we make when introducing terms for substances, there are no possible worlds in which there is water but no H<sub>2</sub>O: the truth that water is H<sub>2</sub>O, then, is metaphysically necessary.

The underlying empirical presuppositions about the microstructure of a substance (that the hidden structure of a substance determines what it is to be a member of a substance-kind), in cases where that presupposition is satisfied, generates a ‘defeasible necessary and

sufficient condition' (1975: 225) for being a member of a given natural kind. Moreover, where that condition is satisfied, since meaning explanations are given ostensively – for instance, *dthat* is water (again, to use Kaplan's indexical 'that') – the satisfaction of those conditions also determine whether or not we have successfully referred with our use of the term<sup>10</sup>. The hidden structural property that determines what it is to be a member of the natural kind is the what Putnam calls the 'essence' (1996: 20) of that kind, and the essence of the kind fixes the conditions captured by the *same<sub>NK</sub>* relation, which are (eventually) fleshed out by scientific investigation. Thus, when Putnam claimed that the 'this' in '*this* is water' was 'a *de re* "this" – i.e. the force of my explanation is that "water" bears a certain equivalence relation (...) to the piece of liquid referred to as "this" *in the actual world*' (1975: 231), plausibly what he intended was that ostensive definitions latch onto the essence of the kind in question, as possessed by each individual member of the kind with which a speaker is acquainted.

What precisely the essence of the kind is, on this view, is a matter for scientific investigation. Nevertheless, a speaker's ability to introduce a kind term does not require that the speaker *knows* what the essence of the kind is. All that is required is that they *assume* that there is such an essence to be found, after some indeterminate amount of empirical investigation, since it is this assumption that underpins the introduction of the kind term. In effect, one of the things rigid designators allow us to do is 'to believe in essences without having to know what they are' (de Sousa 1984: 567). Provided that the assumption is made, science, it turns out, is in the business of discovering metaphysical necessities, and

---

<sup>10</sup> Of course, what is meant here is not simply that I *manage* to refer to, say, *dthat* object (whatever *dthat* object happens to be) with any arbitrarily selected term. So, for instance, I manage to refer to my telephone with the word 'frog' insofar as I draw your attention to the object I intended to, with a random word. Rather, by successful reference I mean of the word that I am using, that I am using it correctly.



representing these claims via theoretical identity sentences, such as ‘water is H<sub>2</sub>O’, that are necessarily true, but knowable only *a posteriori*. Hence, if ‘a “logically possible” statement is one that holds in some “logically possible world”, *it isn’t logically possible that water isn’t H<sub>2</sub>O*’ (Putnam 1996: 18), since there is no possible world in which water exists without being H<sub>2</sub>O.

## 2.4 The Kripke-Putnam programme

The commonalities between Putnam’s and Kripke’s semantics are obvious: both have a similar metasemantics, both offer a version of (*CDR*), and both endorse the necessity and *a posteriori* of theoretical identities. Indeed, reference to the ‘Kripke-Putnam’ programme is commonplace (Donnellan 1983: 85; Mumford 2005: 428, Soames 2002: vii). However, Putnam’s motivations appear quite distinct from Kripke’s, and in his later work Putnam actually rejects an essentialist reading of his claim that theoretical identities are necessary yet knowable only *a posteriori*, effectively distancing himself from Kripke.

Firstly, there is a distinctly different motive driving Putnam’s semantic thesis about natural kind terms: while both Kripke and Putnam objected to what they saw as descriptivist psychologism, Putnam’s concerns are primarily with referential stability and theories of meaning, rather than with theories of necessary truth (Putnam 1996: 17). Consider, for instance, the way Putnam sets up the Twin Earth thought experiment. He says:

The fact that an English speaker in 1750 might have called XYZ “water,” whereas he or his successors would not have called XYZ water in 1800 or 1850 does not mean

that the “meaning” of “water” changed for the average speaker in the interval. In 1750 or in 1850 or in 1950 one might have pointed to, say, the liquid in Lake Michigan as an example of “water”. What changed was that in 1750 we would have mistakenly thought that XYZ bore the relation [*same*<sub>NK</sub>] to the liquid in Lake Michigan, whereas in 1800 or 1850 we would have known that it did not. (Putnam 1973: 702/3)

This version of the Twin Earth thought experiment speaks directly to the threat of Thomas Kuhn’s incommensurability thesis (Kuhn 1996: 117) – the claim that science undergoes so-called ‘paradigm shifts’ where the postulates of the previous scientific worldview,  $W_{v_1}$ , are incommensurable with the current scientific world view,  $W_{v_2}$ , and as such the words used to denote the scientific entities of  $W_{v_1}$  have no meaning in, and are untranslatable into, the language of  $W_{v_2}$ .

One standard diagnosis of this problem is that incommensurability is a product of an internalist notion of meaning (the thought that meanings are in the head), indicative of descriptivism. The plausibility of this interpretation is made more obvious by considering Putnam’s other variations on the Twin Earth thought experiment. Consider, first, the non-science-fiction example Putnam gives (Putnam 1973: 704), where he invites us to consider someone, call him Hilary (since it is Putnam himself that features in the example), who cannot tell the difference between elm and beech trees. Putnam then asks us to accept three plausible assumptions: (i) that the extension of ‘elm’ in Hilary’s idiolect is the same as anyone else’s, namely the set of all elm trees, (ii) similarly, that the extension of ‘beech’ in Hilary’s idiolect is the same as anyone else’s; the set of all beech trees, and (iii) that the extension of ‘elm’ in Hilary’s idiolect is distinct from the extension of ‘beech’ in everyone else’s idiolect. The

challenge, says Putnam, is to give a credible account of how Hilary's concepts are responsible for the difference in extension between 'elm' and 'beech', given the earlier concession that Hilary cannot, himself, differentiate between the two species of trees. The claim is that since Hilary's elm-concept is identical to his beech-concept, but that, nevertheless, the extension of 'elm' in Hilary's idiolect is not the same as the extension of 'beech' in his idiolect, Hilary's concepts are not responsible for the extensions of his (natural kind) terms.

The second, relevant, science fiction example distinguishes between sub-groups of speakers within the linguistic community: the ordinary group and the expert group. In this example we are asked to imagine that aluminium and molybdenum are indistinguishable to non-experts, at least as they occur when used to make kitchen pots and pans. We then suppose that on Twin Earth molybdenum is as abundant as aluminium is on Earth, and that aluminium on Twin Earth is as rare as molybdenum is on Earth. Moreover, the role that aluminium plays on Earth is played by molybdenum on Twin Earth: so, whereas on Earth kitchen pots are made from aluminium, on Twin Earth they are made from molybdenum. Finally, on Twin Earth 'aluminium' is the name of molybdenum, and 'molybdenum' is the name of aluminium. Unlike the initial Twin Earth case, however, an expert from either community can tell the difference between aluminium and molybdenum. As such, were one of these experts, call her Lisa, to be transported from Earth to Twin Earth, or vice versa, she would soon realise that what people call 'aluminium' on Earth is *not* what people call 'aluminium' on Twin Earth, and similarly for what they call 'molybdenum'. The distinction between experts and non-experts in the example means that although there *is* a decisive meaning of 'aluminium' known by a specialist within the Earthian linguistic community, normal (non-specialist) speakers on either Earth or Twin Earth will be in the same psychological state when they use the word

‘aluminium’, despite ‘aluminium’ having the extension aluminium on Earth and molybdenum in on Twin Earth.

Putnam takes the Twin Earth thought experiments to show that the psychological states of individuals cannot determine the reference of natural kind terms.<sup>11</sup> The crux of each variation of the thought experiment is that we can hold psychological states<sup>12</sup> constant while varying the extensions of our natural kind terms. The conclusion is that there is a general underdetermination of reference by psychological states. But how does this speak to the problem of incommensurability?

## **2.5 Reference and Kuhnian incommensurability**

Consider the shift from alchemy to chemistry, which might aptly be considered a paradigm shift (i.e. a radical change from one sort of research programme to another), and assume a descriptivist theory of meaning. On Kuhn’s view, although alchemists and chemists alike used the word ‘gold’, we might think that ‘gold’ meant one thing to the alchemist and another to the chemist. Let ‘gold’<sub>1</sub> be alchemical-gold, which has a meaning  $M_1$  (perhaps, ‘the yellowish, malleable, metallic, precious metal that can be transmuted from lead’). Let ‘gold’<sub>2</sub> be chemical-gold, with a meaning  $M_2$  (perhaps, ‘the element with atomic number 79’). Now, although ‘gold’<sub>1</sub> and ‘gold’<sub>2</sub>, as they appear in natural language (simply as ‘gold’), are syntactically and phonetically identical, since  $M_1$  is not identical to  $M_2$ , we can say that ‘gold’

---

<sup>11</sup> I say ‘Putnam’ here, explicitly, since the distinction between expert and non-expert speakers only goes to show that the extension of a natural kind term is not determined by the psychological states of non-expert speakers. It could still be the case that it is the psychological states of experts that determine the extension of natural kind terms, as noted by Mellor (1977), and discussed in chapter 5 of this thesis.

<sup>12</sup> Again, the psychological states of *non*-expert speakers.

meant one thing for the alchemists and another for the chemists. On a descriptivist theory of meaning and reference, since ‘gold’<sub>1</sub> and ‘gold’<sub>2</sub> mean different things they *literally* refer to different kinds of substances. The threat of Kuhnian incommensurability is that, as such, alchemists and chemists cannot communicate with one another about gold. When chemists discovered that gold is the element with atomic number 79 it was not that they discovered some additional facts about the stuff alchemists had been calling ‘gold’. Rather, they discovered a new substance which they called also called ‘gold’. Moreover, or so the Kuhnian claim goes, the alchemists’ notion of ‘gold’, what I have called ‘gold’<sub>1</sub>, is not even translatable to the chemists’ notion of ‘gold’, ‘gold’<sub>2</sub>.<sup>13</sup> In a sense, chemists and alchemists are as distant from one another as people on Earth and Twin Earth respectively.<sup>14</sup>

The Kuhnian worry as Putnam identifies it stems from two intuitions. The first, is the descriptivist thesis that meaning (or sense) determines extension – where M<sub>1</sub> is the meaning of ‘gold’<sub>1</sub>, M<sub>1</sub> determines the extension of ‘gold’<sub>1</sub>. The second, and in this case more significant, intuition is the antirealist intuition about the status of scientific theories. According to the antirealist intuition, where we have two theories like alchemy and chemistry, it is not the case that there are ‘two approximately correct descriptions of some fixed real or theory-independent entities’ (Putnam 1996: 20), because the notion of a ‘correct’ or ‘true’

---

<sup>13</sup> In some sense the example I have given is odd, since there was no substance which could correspond to the alchemists’ notion of ‘gold’, since gold cannot be transmuted from lead! However, I view this as actually somewhat helpful to the example. Provided we assume that alchemists had something like the notion I propose in mind, then it looks to be a postulate of their worldview that such a substance exists (much like, say, phlogiston). However, since gold cannot be transmuted from lead, the alchemists’ gold-postulate is nothing like the chemists’. Hence ‘gold’<sub>1</sub> and ‘gold’<sub>2</sub> are incommensurable.

<sup>14</sup> That isn’t to say that solutions to incommensurability have not been proposed – they have. Rachel Cooper (2007), for example, suggests that when paradigm shifts occur a pidgin language forms, facilitating communication across paradigms. The classic response, however, is from Donald Davidson (1973-4), who claims that we cannot make sense of the thought that a theory cannot be translated from one language into another, and hence that we can make no sense of incommensurability (see also Miller 2007 for an explanation of Davidson’s response).

description is intra-theoretic, not inter-theoretic. In other words, the antirealist intuition is that the notion of truth only makes sense within a particular theory. Putnam's general objection to this intuition is that 'scientists use terms as if the associated criteria were not *necessary and sufficient conditions*, but rather *approximately* correct characterizations of some world of theory-independent entities' (*ibid.*). They are, in effect, appealing to the notion of 'extension' when they deploy natural kind terms, and 'extension' is intimately connected to the notion of 'truth', since the extension of a term is precisely the set of objects of which that term is true.

If we accept the antirealist intuition, says Putnam, this leaves us in a peculiar position where we are unable to say of some object,  $x$ , that it does not lie within the extension of a particular natural kind term, 'tungsten' say, as used by a particular language user outside of our theory, and similarly we cannot judge whether that language user's statement ' $x$  is tungsten' is true or false (again, when they are outside of our theory). For Putnam, if attempts to make the notion of 'truth' compatible with antirealist intuitions about intra-theoretic kind terms makes for a theory of meaning that is incompatible with basic observations of scientific progress, then such a theory (of meaning) is untenable on that basis.

Take, for instance, Sir Martin Frobisher's hapless voyages between Canada and England, and his shipping of around 1500 tonnes of what he took to be gold ore, but which turned out to be mere iron pyrites, back to England. Intuitively, we wish to say of the set of objects  $x$ , filling the holds of Frobisher's ships, that despite Frobisher referring to them as gold, he was wrong to do so, since  $x$  was not within the extension of 'gold'. Similarly, we want to say that when Frobisher stated that ' $x$  is gold', where  $x$  refers to the cargo of his ship(s), he said something false. If we assume, plausibly, that we do not have the same theoretical outlook as Frobisher – after all, this all took place in the 16<sup>th</sup> Century – then we can imagine that Frobisher thought that anything satisfying the description 'heavy, metallic,

shiny, yellow-coloured substance' was gold. If we accept a paradigm-relative notion of truth, then it turns out that we are not in a position to say whether or not sentences like '*x* is gold', as uttered by Frobisher, are true or false, since truth is only intra-theoretic, rather than inter-theoretic. But this seems plainly false! It is not that Frobisher was right that he had gold relative to his paradigm, but the meaning of the word 'gold' changed sometime after he shipped the substance back to England. Rather, Frobisher was simply mistaken from the outset, despite no one within his community being in a position to say what gold actually was (the community were, remember, operating within the alchemy paradigm).<sup>15</sup>

Putnam appeals to the distinction between verification on the one hand, and truth on the other, to illustrate the objection. Consider Putnam's own, more extreme, gold example, involving Archimedes. Assume, for the sake of argument, that Archimedes has roughly the same operational definition of gold as we do, and consider some object, *x*, that satisfies that definition, but which is not gold but fool's gold. In Archimedes' time there may not have been the means to distinguish gold from fool's gold, and thus no means by which Archimedes could have reasoned that '*x* is gold' was false.

But, says Putnam, this is a problem with verification rather than truth (Putnam 1996: 22). There are various situations that we *could* have described to Archimedes or Frobisher to put the hypothesis that '*x* is not gold' back in the running. Regardless of whether either grasped the theory, we could illustrate our claim with empirical evidence. Once we had done

---

<sup>15</sup> This is not completely accurate. Experts were, eventually, able to tell that the samples Frobisher shipped back were not gold. However, that Frobisher went *back* to Canada for a second visit, with more ships, indicates that, at least initially, the cargo really was mistaken for gold by various members of the community. to Canada for a second visit, with more ships, indicates that, at least initially, the cargo really was mistaken for gold by various members of the community.

so, we could explain to them how the empirical results were produced via our atomic theory, leaving it in no ‘doubt that he would have agreed with us that  $x$  isn’t gold’ (*ibid.*).

The point, according to Putnam’s (*CDR*), is that when we use terms like ‘gold’ it is our intention that “‘we use the name *rigidly*” to refer to whatever things share the *nature* that things satisfying the description normally possess’ (*ibid.*). Since there is some fact of the matter about what that nature is, there will be some fact of the matter that determines whether a sentence such as ‘ $x$  is gold’ is true or false. Indeed, even when we are not in a position to *know* what that truth is – imagine, for instance, trying to discern how many electrons there are in the universe; just because we cannot verify how many there are does not entail that there is not some actual quantity – that does not change the point that there is some mind-independent fact of the matter, which makes the relevant sentence true or false.

The aim of Putnam’s thesis, then, is to respond to Kuhnian incommensurability, and he does so by claiming that natural kind terms, such as ‘gold’, were introduced with the intention that they refer to a particular natural kind, in this case the gold-kind, instantiated by samples with which a speaker was acquainted. As science has progressed, or so the story goes, we have come to discover additional facts about those kinds, and even discovered that some of the objects we thought to be within the extension of the kind are, in fact, not members of that kind (as in the Frobisher case). Putnam’s metasemantic story, the claim that kind terms are introduced by ostension to samples with which we are acquainted, in conjunction with his claim that true application of a kind term (to other samples/members of the kind) depends upon whether a sample bears the relevant sameness relation to the original, effectively moves the notion of meaning not only outside the head, but outside any particular scientific theory.



The meanings of kind terms, then, are *externalist* according to Putnam: whatever the kind turns out to be, members of which we find in our environment, since the meaning of a kind term just is its referent, it is the kind itself that determines the term's content. The upshot is that there is no change of reference from a pre-scientific to a scientific community. It might, of course, be that some of the objects that people *thought* were in the extension of a natural kind term turn out not to be – just think of poor Sir Frobisher! The empirical discovery of gold's essence allows us to distinguish between the samples of different kinds that happen to satisfy the same operational definition. Since we have always intended our use of the term 'gold' to apply to samples bearing the appropriate sameness relation, says Putnam, rather than satisfying the operational definition (which is more of a heuristic than anything else), where we have misapplied the kind term, we do not intend our usage to be accepted.

Plausibly, then, whilst Putnam's semantic thesis has consequences that are obviously parallel to Kripke's, the motivations of each are quite distinct: while Kripke's arguments aim to show us that epistemic and metaphysical necessity can come apart, Putnam's arguments aim to show us how we can avoid Kuhnian incommensurability, although his view has, as a corollary, the same implications as Kripke's (concerning the distinction between epistemic and metaphysical necessity). I shall leave an in-depth analysis of Putnam's mechanism for generating *a posteriori* necessities until chapter 4, where, following Nathan Salmon, I will identify the key essentialist assumptions within Putnam's thesis. The point of the preceding analysis is to demonstrate at least one sense in which reference to the 'Kripke-Putnam' programme is misleading.

There is one feature of Putnam's Twin Earth thought experiment that I have yet to mention: the thesis that the meanings of natural kinds terms 'depend upon a fact about language that seems, surprisingly, never to have been pointed out: that there is *division of*

*linguistic labour*' (Putnam 1973: 704). A discussion of the *division of linguistic labour (DLL)* will be relevant for two reasons. Firstly, it is often thought to be a premise in the argument for establishing an externalist theory of meaning, as mentioned indirectly above. We shall see, in chapter 5, that this is not the case. Secondly, once (*DLL*) has been shown *not* to be a premise in any argument for externalism, this will open it up for use within versions of descriptivism. Since this is the topic of chapter 5, I reserve my treatment until then. Now I will consider LaPorte's response to the Twin Earth thought experiment.

## **2.6 LaPorte on Twin Earth**

Recently, Joseph LaPorte (2004) has argued that there are real world counterexamples to Putnam's Twin Earth thought experiment. In LaPorte's example, although we discover that a particular natural kind term has been used to refer to two constitutionally distinct substances, we decide, regardless, to maintain our usage in spite of this discovery.

LaPorte (2004: 94-100) asks us to consider the history of the term 'jade'. In particular, LaPorte disputes Putnam's claim that 'although the Chinese do not recognise a difference, the term "jade" applies to two different minerals: jadeite and nephrite' (1975: 241). According to LaPorte, although Putnam is correct that 'jade' refers to chemically distinct substances, the history of jade is much more akin to Putnam's Twin Earth thought experiment, since the Chinese were well aware of the difference between nephrite and jadeite. The word 'jade', LaPorte tells us, was initially introduced to refer to the stuff with the chemical formula  $\text{Ca}_2(\text{Mg,Fe})_5\text{Si}_8\text{O}_{22}(\text{OH})_2$ , what chemists call 'nephrite'. Then, sometime towards the end of the eighteenth century a 'very similar stone with a totally unrelated composition,  $\text{NaAl}(\text{SiO}_3)_2$ , made its way for the first time to China' (LaPorte 2004: 95). In 1863 Alexis

Damour, LaPorte tells us, named this chemically distinct substance ‘jadeite’. Prior to this, however, the Chinese who worked with jade were able to discriminate between nephrite and jadeite (LaPorte 2004: 96). Putnam’s claim that they were unaware of the difference, then, is false. The important question, however, is why the Chinese, who used jade much like the west used gold, nevertheless refer to both nephrite and jadeite as jade? As LaPorte reports, when jadeite was discovered (after nephrite) the Chinese accepted both as instances of jade, and hence the term ‘jade’ was taken to apply to ‘jadeite as well as nephrite’ (LaPorte 2004: 96).

On LaPorte’s analysis there was, prior to the discovery of jadeite, a ‘hidden vagueness’ (2004: 97) in the application of the word ‘jade’, and with the later discovery of jadeite the meaning of ‘jade’ was refined. According to the Kripke-Putnam story, when we reflect upon our application of natural kind terms for substances, such as ‘water’, in scenarios where the superficial properties of two objects are identical, but their underlying microstructure is quite different, our *intuition* is that ‘water’ only applies to one of the microstructurally demarcated kinds. The jade example, however, appears to produce the opposite result. When Chinese speakers find themselves in a position analogous to Twin Earth, where there are two superficially similar (not quite identical) substances that are microstructurally distinct, they made the converse decision: they maintained that ‘jade’ should apply to *both* substances, despite their microstructural differences.

LaPorte’s conclusion is not that the standard Kripke-Putnam treatment of such cases is not open – speakers could indeed judge that where a term *T* turns out to refer to two microstructurally distinct kinds, that the use of *T* will be restricted to a kind with a single microstructure – but he does deny that it is the *only* available option. Speakers confronted with a Twin Earth style case are effectively confronted with semantic vagueness, and there are various precisification options open to them: they might decide to affix the term, *T*, to a single

microstructurally defined kind (as in the water case); alternatively, they might decide to affix the term, *T*, to a kind defined by more readily observable properties<sup>16</sup>. If that is right then given a case of semantic vagueness, it is open to us, as a linguistic community, to precisify our use of a term as we see fit. In other words, ‘we could go either way’ (LaPorte 2004: 100). If LaPorte is right then the force of the Twin Earth thought experiment is effectively undermined, since it is open to speakers to precisify their natural kind terms as they see fit (‘water’, then, *could* have been precisified to include samples of XYZ within its extension). However, LaPorte’s view is controversial (see, for instance, Bird 2010 for criticisms), and as he concedes (2004: 99-100) attitudes in the West towards the jadeite/nephrite debate have been somewhat closer to the original Kripke-Putnam treatment. At the very least the example is suggestive, and the idea of vague natural kind terms that are precisified is one I shall build upon later, in chapter 5.

## **2.7 Later Putnam**

Despite what has been said so far, it is worth noting that Putnam later came to reject much of the metaphysical import of the aforementioned view. The nub of Putnam’s change of heart, I hazard, stems from Salmon’s (2005) criticism that his theory of reference requires a commitment to a controversial brand of essentialism (much more on this in chapter 4).

---

<sup>16</sup> LaPorte’s jade case is suggestive, but there are a number of routes to the general idea that microstructural kinds need not be the kinds to which our natural kind terms are affixed. For instance, the genetic diversity of biological species and the assumption that species are natural kinds also undermines the requirement for microstructural kinds (see Dupré 1995; LaPorte 2004). Similarly, some philosophers have argued (Needham 2000; Van Brakel 2000) that classifications in terms of macroscopic properties are needed for chemical kinds, to unify microstructurally heterogeneous kinds of substance. Water, for instance, turns out to be microstructurally heterogeneous, and hence if chemical kinds were individuated by a unique microstructure, ‘water’ would not pick out a chemical kind.

However, Putnam sets up his own critique somewhat differently. Putnam begins by observing that there is an ‘incompatibility between metaphysical realism and the denial of intrinsic [essential] properties’ (Putnam 1983: 207). Metaphysical realism is the conjunction of a few ‘closely associated philosophical ideas about the relations between language and reality’ (Hale and Wright 1998: 472). These ideas are (i) that there is a mind-independent world of objects, (ii) that we can think and talk about those objects via ‘a “correspondence” relation between the terms in our language and some sorts of mind-independent entities’ (Putnam 1983: 205), and (iii) that there is one true description of the way the world is. The denial of essential properties goes back to John Locke (more on this in Chapter 6), and as Putnam construes it, it is the claim that ‘a thing is not related to any one of its properties (or relations) any more “intrinsically” than it is to any of its other properties or relations’ (1983: 206). The problem Putnam identifies is that a metaphysical realist must account for (ii), the correspondence relation between thought and thing, and that this is now typically done by appeal to the built in ‘similarities and dissimilarities between things in nature’, which is to say the essences of those things (Putnam 1983: 207).

The problem that Putnam identifies is that the metaphysical realist requires something like natural kind essentialism, since their view is incompatible with its denial. However, he says, this ‘modern mixture of materialism and essentialism’ is inconsistent (1983: 207). Here, however, I am not interested in Putnam’s critique of metaphysical realism. Rather, I am interested in his comments concerning the causal theory of direct reference (*CDR*), his rejection of essentialism, and his explicit disavowal of the revival of metaphysics which he attributes to Kripke (1983: 207), but which most of the philosophical community typically attribute to Putnam and Kripke. More simply, I wish to present later Putnam’s rejection of the apparent necessity of theoretical identities, which is to all intents and purposes a rejection of

essentialism. In chapter 4, when I present Salmon's analysis of (*CDR*), and his conclusion that there are hidden essentialist premises in the derivation of necessary *a posteriori* truths concerning natural kinds, I will highlight, following recent work by Ian Hacking (2007), key passages in early Putnam where he seems to connect any necessity to the *conventions* of our classification, as he certainly does in his (1983). For now, though, we shall focus only on Putnam's rejection of the conclusion that there can be metaphysical necessities about theoretical identities.

Later Putnam endorses Kant's attitude towards the rationalists, claiming that while medieval philosophers, for instance Duns Scotus (cf. Pini 2009: 255 - 272), thought we had some faculty of intuition that allowed us to latch on to the essences of things, 'there is no such faculty' (Putnam 1983: 209). Putnam's antiessentialist argument begins by identifying an important presupposition of the metaphysical realist position, namely that it is standardly combined with a correspondence theory of truth. This, says Putnam, results in the claim that there is one true (scientific) theory, and that the theory is made true by corresponding to actual facts of the matter. But, he goes on to say (following work by Nelson Goodman (1978)), if one believes that there is one true scientific theory – best physics, say – then with the correspondence theory of truth this implies that there must be a '*ready-made* world' (Putnam 1983: 211) possessing a determinate and built-in structure for the theory to latch on to. Thus, once our theorists discover what it is to be tungsten, then given the earlier framework common to Kripke and Putnam, not only does it turn out to be true that tungsten is the element with atomic number 74, but that truth is metaphysically necessary i.e. true in all possible worlds, yet knowable only *a posteriori*.

But here Putnam (or rather later-Putnam), whilst considering whether an ontology of objectives individuated by their modal properties can aid materialism, claims both that any

such ‘ontology *presupposes* essentialism’ and further that modal properties are not part of the furniture of the physical world (1983: 220). The most interesting feature of this discussion for my purposes is Putnam’s novel reinterpretation of what he intended when he originally spoke of ‘essences’ in his ‘The Meaning of “Meaning”’; now, he says, the discovering that the nature of a natural kind like tungsten is to have atomic number 74 is *not* to discover tungsten’s essence in any metaphysically robust sense. Rather, that the kind tungsten has any essence at all is the product of our original referential intentions when we introduced the name ‘tungsten’, namely that we intended it to apply to a (natural) kind of substance the members of which are unified by a shared atomic composition. When such a presupposition is borne out by the facts, there is an ‘essence’ of the kind; when it fails to be borne out, then the term failed to refer to a natural kind. Vivially, however, the so-called essence of the kind ‘is the product of our use of the word, [and] the kinds of referential intentions we have’ (Putnam 1983: 220), not any fundamental metaphysical fact about the world. I propose that we call this view (that essences are the product of referential intentions) *thin* essentialism, where the contrast is with *thick* essentialism, which I take to be the view that theoretical identities like ‘tungsten is the element with atomic number 74’ are necessary yet knowable *a posteriori*, but where the necessity is the product of some underlying, objective *essence* that tungsten has (rather than the product of how we decide to deploy the word ‘tungsten’).

The thin/thick essentialism distinction does not modify Putnam’s semantic theory *per se*, since he still claims that language users have certain sorts of referential intentions when they introduce natural kind terms. What the move to thin essentialism does is to undermine the metaphysical importance of Putnam’s original (1975) essence claims concerning natural kinds, namely the derivation of essentialism from (*CDR*) (again, we shall review this explicit project in Chapter 4). Putnam concedes that since his semantic theory *assumed*, from the

outset, that direct reference was unproblematic, it could not be employed to derive essentialist conclusions. The problem is that metaphysical realism holds that best physics is *the* true description of the world, and moreover that there is a correspondence relation between thought and thing. The causal theory of direct reference cannot provide support for such a view in a non-question begging fashion, since it presupposed the notion of reference from the outset i.e. it presupposed that there could be a straightforward relation between thought and thing.

The distinction between early and later Putnam, then, is effectively a distinction between an *essentialist theory of reference* that supposes a ready-made world, with a built-in modal structure<sup>17</sup>, which can be captured by one true theory, and a metaphysically neutral (*CDR*) that assumes reference is unproblematic, allowing that our initial intentions fix the reference of natural kind terms. On this view, what makes it necessarily true that water is H<sub>2</sub>O is (i) the fact that the stuff we call ‘water’ has the property of being composed of H<sub>2</sub>O molecules, and (ii) the referential intentions we have when we introduced the term. Since (ii) is dependent upon us, and our referential intentions do *not* have some special access to the word, the necessity of theoretical identities like ‘water is H<sub>2</sub>O’ are the product of both the physical facts and our classificatory and referential conventions.

However, Putnam is also quite deflationary about the role of convention, since he wants to avoid any implications that scientific classifications are purely subjective. Rather, he says, we should treat truth as an epistemic notion. The basic idea is that claims like ‘tungsten

---

<sup>17</sup> Putnam attributes the view that the world has a built in modal structure to Kripke (1983: 220). A contemporary version of this view has recently been expressed and defended by Brian Ellis (2001). In Chapter 7 I will illustrate just why such a thesis is untenable. Nevertheless, it is worth flagging from the outset that these views are closely related.



is the element with atomic number 79' are true, but where the notion of truth is 'corrigible, relative to task and technique, but not *subjective*' (Putnam 1983: 226). Thus, the truths about tungsten – such as that tungsten is an element, and that *qua* element there are 74 protons within each atomic nuclei of the atoms that belong to the tungsten-kind – depend upon how we define 'tungsten', which in turn depends upon the tasks and techniques of chemistry (or perhaps physics). Thus, since instances of tungsten have a variety of properties, there is no particular property that is, objectively, the essence of tungsten. Rather, it is only once we have *defined* tungsten as the element with atomic number 74 that the property of having 74 protons becomes essential to the kind. This, of course, is only the briefest of sketches of Putnam's preferred metaphysics. However, in Chapter 6 I argue that, in fact, this is Locke's empiricist view of essentialism; hence I reserve a full treatment of this notion until Chapter 6.

To put the issue back into semantic terms, when chemists use 'tungsten' their referential intention must be that 'tungsten' is to apply only to a simple substance (an element) that is defined by its atomic number. Similarly, the classification of elements via their atomic number is also a product of our intentions, but this time the intentions of the research programme concerned with classification. Hence, it is the conjunction of referential intentions, classificatory intentions (or decisions) and the physical facts, that make the sentence 'tungsten is the element with atomic number 74' necessarily true, and make the property of having atomic number 74 the essence of the kind. The physical furniture of the world, in and of itself, contains no modal properties, and thus is insufficient to ground the necessary truth of theoretical identities.

Although I have attributed this view to later Putnam, Putnam himself, as we have seen, reinterprets his initial talk of essence in line with thin essentialism, seemingly indicating that this was what he had in mind all along. Similarly Ian Hacking (2007) has argued that this

was *always* Putnam's intention. As we shall see in Chapter 4, this very much at odds with the now standard interpretation of Putnam as expounded by Salmon (2005). Since both interpretations are plausible, and the exegetical evidence is indecisive, I shall simply distinguish early-Putnam from later-Putnam, and concede, following Salmon (and contra Hacking 2007), that early-Putnam does employ essentialist assumptions in his derivation of necessary *a posteriori* truths concerning natural kinds (more on this in chapter 4). However, by distinguishing early from later Putnam, I will employ later-Putnam's arguments against essentialism, or at least against a semantics grounded in an essentialism that has not been independently justified, when, in chapter 8, I offer the conclusions to this thesis (a version of descriptivism without a metaphysically suspicious notion of 'essence').

## CHAPTER 3

In this chapter I want to return to some of the puzzles raised at the end of chapter 1. In particular, I will explore answers to the questions concerning phrases like ‘the element with atomic number 79’, and ‘H<sub>2</sub>O’ which feature in paradigmatic (Kripkean) theoretical identification sentences. In chapter 1 I asked how we should treat such phrases: are they analogous to proper names, or are they descriptions? If they are descriptions then why are only some descriptions rigid? Here I will consider Scott Soames’ complex semantic thesis concerning natural kind terms and theoretical identity sentences, which distinguishes between the semantics of phrases like ‘the element with atomic number 79’ and general names like ‘gold’, claiming that both types of phrase are rigid designators, but conceding that the former are descriptive.<sup>18</sup> I begin, in sections 1 and 2, by presenting, in more detail, the concerns outlined at the end of chapter 1, and contrast those with Soames’ own motivations for a further treatment of natural kind terms, since they are somewhat different. I present, in section 3, two potential accounts of natural kind terms that Soames rejects, but which have been raised elsewhere in the literature. I then proceed to present, in section 4, Soames’ positive semantic thesis. There is, as we shall see, a division between types of theoretical identity: firstly, simple-simple theoretical identities, containing two simple (single word) natural kind terms, for instance ‘Apatosaurus is Brontosaurus?; secondly, simple-complex theoretical identities, containing one simple natural kind term, and a complex theoretical term, for example ‘gold is the element with atomic number 79’. In section 5 I will consider the first of

---

<sup>18</sup> Despite also claiming that the similarity between proper names and natural kind terms that allows us to extend the conclusions concerning necessity and *a posteriori* from the former to the latter is their mutual nondescriptionality.

these categories, and in sections 6 and 7 respectively, consider and criticise two important metaphysical premises in Soames' argument for the necessity of theoretical identities. In section 8 I will argue that on one plausible reading of Soames' thesis, simple-simple theoretical identities are just straightforwardly false. In section 9 I will highlight the role that metaphysical essentialism is playing in Soames' account of the necessity of simple-complex theoretical identities. The upshot of this chapter will be that although I think Soames is right about the descriptive nature of phrases like 'the element with atomic number 79', his theory of types (section 6) and his special auxiliary assumption (section 7) are both false. I will conclude that simple-complex theoretical identities rely upon the independent metaphysical thesis of essentialism, which will set the scene for my defence, in Chapter 5, of descriptivism.

### **3.1 Natural kind terms, predicates and identities**

We know, from chapter 1, that theoretical identification sentences are supposed to be, when true, necessary, but knowable only *a posteriori*. This necessity, it is alleged by Kripke, is guaranteed by the semantic fact that natural kind terms, such as 'ammonia' and 'NH<sub>3</sub>' are *rigid designators* – terms that have the same reference in all and every possible world – plus the empirical fact that pairs of such terms are co-referential (making the identity sentence 'ammonia is NH<sub>3</sub>' true). The *a posteriority* is the result of the investigation required to discover the truth of such statements: the truth of 'ammonia is NH<sub>3</sub>' cannot be discovered from mere reflection on the terms within the identity statement as it can in, say, 'a bachelor is an unmarried man'.

The bulk of the detail of Kripke's semantic thesis, however, was developed in his discussion of proper names, and identity sentences containing them. His intention was that the

same considerations which applied to proper names and identity statements were to apply, via straightforward analogy, to natural kind terms and theoretical identity statements. Classic examples such as ‘Cicero is Tully’, ‘Hesperus is Phosphorus’ and ‘Bob Dylan is Robert Zimmerman’ are supposed to act as models for, and be analogous to, theoretical identification sentences such as ‘ammonia is  $\text{NH}_3$ ’, ‘gold is the element with atomic number 79’ and ‘heat is molecular motion’.

If theoretical identifications are analogous to identity sentences containing proper names we should expect them to share important linguistic similarities. Were the analogy straightforward we should expect theoretical identification sentences to contain two singular name-like terms, which are (i) the products of two separate baptismal events, (ii) connected to two distinct causal chains of reference, (iii) both rigid designators, and (iv) flanking the identity sign. Unfortunately almost none of the examples of natural kinds that Kripke gives fit this model.

Consider theoretical identifications concerning chemical kinds. These seem to come in at least two types: the first, typified by ‘water is  $\text{H}_2\text{O}$ ’, contains a *general* name (‘water’) and a *chemical formula* (‘ $\text{H}_2\text{O}$ ’) flanking the identity sign (provided we take ‘is’ to be the English language expression of the identity sign) e.g. methane is  $\text{CH}_4$ , sulphuric acid is  $\text{H}_2\text{SO}_4$ , Corundum is  $\text{Al}_2\text{O}_3$ . While the general name ‘water’ is arguably nondescriptive, and thus analogous to the proper name ‘Cicero’, ‘ $\text{H}_2\text{O}$ ’ is a chemical formula, and the semantic credentials of chemical formulae receive no treatment in Kripke.<sup>19</sup> As a result it is not immediately obvious that both terms within a theoretical identity sentence are rigid

---

<sup>19</sup> More recently Nathan Salmon (1987/88, 2005) has developed an account of the rigidity of chemical formulae where he argues that they are simply general term versions of proper names. See Chapter 4 for further details.

designators, and this is sufficient to raise questions about the modal status of true theoretical identification sentences containing chemical formulae.

There is also a second type of theoretical identification found in Kripke, exemplified in his discussion of gold. According to Kripke ‘a material object is (pure) gold *if and only if* the only element contained therein is that with atomic number 79’ (Kripke 1980: 138, *emphasis added*), and that fact that gold *is* such an element is metaphysically necessary. This second type of identity sentence has the distinctive logical form  $\forall x (Px \leftrightarrow Qx)$ . Formulated thus, the alleged identity sentence not only lacks the identity sign, but, as Scott Soames has noted (2002, 2004), has ‘gold’ and ‘the element with atomic number 79’ in the predicative position. Since Kripke’s account of rigidity was *only* for singular terms, it does not obviously extend to predicates. Hence we have no model for predicate rigidity, and consequently no reason to think that theoretical identity sentences constructed using predicates are necessary when true.

Scott Soames, in *Beyond Rigidity* (2002), recognises this lacuna in Kripkean semantics. He argues that we still require an account ‘specifying how the semantic model developed for proper names should be extended to a wider class of terms’ (Soames 2002: 242). Soames’ initial argument to establish that there is an unfinished semantic agenda in *Naming and Necessity* is essentially that which has been outlined above: that there are examples of theoretical identification sentences where the natural kind terms (i) appear in the predicative position, and (ii) appear to flank a universally quantified biconditional, rather than the identity sign. The important question, which will determine just how large Kripke’s lacuna is, is: how extensive is the class of natural kind terms that are predicates?

According to Soames (2002: 246-248) the *majority* of natural kind terms are, or should be treated as, predicates. He reasons as follows: the class of natural kind terms includes mass nouns, count nouns and adjectives, and there is no way to treat all of these as singular: Count nouns like ‘human’ and ‘mammal’, which are natural kind terms, naturally combine with quantifiers to occur in the predicative position e.g. ‘*some* mammals, but *all* humans, are less than 12ft tall’; this natural combination of count nouns with quantifiers constitutes a plausible argument that the predicative use of such terms is fundamental; Mass nouns too combine naturally with quantifiers e.g. ‘*all* sugar, but only *some* water, is sweet’, and occur in the predicative position: in unmodified usage they parallel count nouns, and as count nouns are treated as predicates so to should mass nouns be; similarly (and more obviously) adjectives are always predicative, and can occur naturally with quantifiers e.g. ‘*everyone* was amused, but *someone* was embarrassed, by the comedian’s political satire’.

If Soames is right then Kripke’s account of the necessary *a posteriori* does not obviously apply to natural kind terms. Since the account of rigidity, which grounds the necessity of true identity sentences, was only given for proper names, and natural kind terms are disanalogous to proper names, Kripke’s account of rigidity does not apply to them. Consider the identity sentence ‘Marie Curie is Maria Sklodowska’. Since names are always rigid designators, which is to say they have the same denotation in all possible worlds,<sup>20</sup> and the identity sentence is true, it will be true in all possible worlds and hence metaphysically necessary. But Kripke does not say that predicates are rigid designators. So if natural kind terms are predicates then given that Kripke thinks that theoretical identification sentences

---

<sup>20</sup> This is not completely accurate. It is possible to distinguish between *types* of rigid designator. So one might argue that rigid designators only designate in worlds where their denotation is present, and failed to designate in worlds where their denotation does not exist (cf. Besson 2010; LaPorte 2006; Salmon 2005). However, this distinction is not important for my discussion.

concerning natural kinds are also, when true, necessary and knowable only *a posteriori*, we urgently need some account of the rigidity of predicates: this is Soames' project in *Beyond Rigidity*.

It is worth noting, however, that the move to treat natural kind terms as predicates is not uncontroversial. Nathan Salmon has argued that Kripke's account of rigidity for proper names applies directly to his examples of natural kind terms, because natural kind terms are simply *general* names. The crux of this issue, which I will address in the next chapter, is whether or not the most fundamental use of mass nouns and count nouns is predicative. Nevertheless, it is clear from what Kripke says that he recognises that many of our natural kind terms can function as predicates (Kripke 1980: 127), and this is sufficient for our purposes here.

### **3.2 Revisiting Kripke**

There is one further argument, neglected by Soames, to demonstrate that Kripke's account of theoretical identification containing natural kind terms is not the straightforward analogue of his account of identity sentences containing proper names. Laporte, in his (2004: 36-37), makes brief reference to the similarity between Kripke's theoretical identifications and his account of the necessity of origin, which was first noticed by Salmon (2005): this was mentioned in chapter 1, but that connection is developed in this section, to illustrate the fact that not all of Kripke's claims concerning necessity are grounded in language.

For Kripke, theoretical identifications are 'identities between terms for substances, and also the properties of substances and of natural kinds' (Kripke 1980: 166). In the classic



discussion of gold, as a paradigm example of a theoretical identification, Kripke claims that ‘it is part of the nature of gold as we have it to be an element with atomic number 79’ (Kripke 1980: 125), and that ‘science attempts, by investigating basic structural traits, to find the nature, and thus the essence (in the philosophical sense) of the kind’ (Kripke 1980: 138). Taken together these claims suggests a reading of Kripke where theoretical identities are the identifications of a natural kind with its essence – they are ‘statements representing scientific discoveries about what this stuff *is*’ (Kripke 1980: 125), rather than statements of identity generated by distinct naming events. Theoretical identifications like the gold example ‘resemble “Cicero is a product of the egg and sperm that generated him” rather than “Cicero = Tully” in respect that they aim to expose essence’ (LaPorte 2004: 37). Standard identity sentences like ‘Cicero is Tully’, in stark contrast, are silent on metaphysical issues about the essence of Cicero. It is this observation, I claim, that represents the most significant difficulty for the supposed analogy between identities between proper names and theoretical identities between natural kind terms: if theoretical identifications are the analogue of Kripke’s claims about the essences of individuals, then although Soames might be right that they are semantically dissimilar to identity sentences featuring proper names, he may turn out to be wrong when he claims that theoretical identities are ‘linguistically guaranteed to be necessary when true’ (2002: 279), since the necessity of origins claims in Kripke stem from the essentialism already presupposed. Hence theoretical identities may well come with significantly more metaphysical baggage than Soames imagines.

The difficulty can be decisively illustrated with a brief analysis of the necessity of origin thesis. According to Kripke, there is a principle of origin for material objects such that ‘if a material object has its origin from a certain hunk of matter, it could not have had its

origin from any other matter' (Kripke 1980: n56).<sup>21</sup> This principle when applied to people, as it is in *Naming and Necessity*, boils down to the claim that for any particular person *P*, *P* necessarily comes from their particular origin – where their origin, for the sake of argument, is the sperm and egg from which they originated – and nothing lacking that origin could be *P*.

Consider, to illustrate, Sir Winston Churchill, and ask whether it is (counterfactually) possible that he was the son of Mr and Mrs Chamberlain. Given that Sir Winston Churchill really did come from his actual parents – Lord Randolph Churchill and Jennie Jerome – according to Kripke, he *necessarily* came from his actual parents. When we entertain a counterfactual scenario where 'Churchill' is the son of Mr and Mrs Chamberlain we are, at best, considering a scenario in which some other person possessing many of the properties that Sir Winston Churchill actually possessed, is the son of Mr and Mrs Chamberlain. As Putnam would say, we are considering a situation where someone else satisfies the operational definition we have of Sir Winston Churchill.

According to Kripke, whilst it is possible to consider situations in which Sir Winston Churchill was never actually born, or where he was born but did not become Prime Minister, or even where someone else with many of the properties of Sir Winston Churchill became Prime Minister, it is not coherent to consider him, the actual man, as being born of different parents. The point here is that people *necessarily* come from the origins that they do, but that their actual origins are not something we know about *a priori* (although the general principle that origin is necessary looks to be *a priori*).

---

<sup>21</sup> This principle is precisely the kind of principle that counts as nontrivially essentialist. However, more on this in chapters 7 and 8.

Kripke's necessity of origin thesis has been given numerous different treatments (see Hughes 2004: 108 – 118, Mackie 2006, McGinn 1976: 127 - 135), but the basic view is that 'a human being's parentage is one of that human being's essential properties' (Mackie 2006: 8). Some philosophers have distinguished between the *necessity* of origin and the *sufficiency* of origin (cf. Mackie 2006: 93 - 117), but for the purposes of this investigation I intend my use of the so-called 'necessity of origin thesis' to be understood in the following way. In Kripke's discussion of the necessity of origin thesis (Kripke 1980: pp. 110 – 115) he presents two types of counterfactual scenario. On the one hand he considers cases where an individual, say Churchill, is born of his actual parents, but lacks all the properties that he has in this world *except* his origin. So imagine that Churchill never became Prime Minister, never smoked cigars, and never won the Nobel Prize for Literature. Despite having none of these properties we can still say that it was Churchill who didn't do these things, because our example *gives us the man*, Churchill, by giving us exactly *his* origin. In this case it seems that origin is *sufficient* to guarantee that we are talking about Churchill.

The other type of case Kripke considers is where we have someone with all the properties we standardly associate with Churchill, but lacking Churchill's unique origin. In these cases, argues Kripke, it is impossible to imagine Churchill 'being born of different parents. It would seem to me that anything coming from a different origin would not be this object' (Kripke 1980: 113). In other words, it is a *necessary* condition of being Churchill that he have the particular origin he actually has. Any possible world where there is someone with all the properties we standardly associate with Churchill, but without his origin, does not count as a world where that person is Churchill: the *lack* of Churchill's origin guarantees that someone who is qualitatively identical to Churchill in all ways other than origin is *not* Churchill.

The two counterfactual scenarios Kripke sketches suggest that his intended notion of the ‘necessity of origin thesis’ is in fact one that states both a necessary *and* a sufficient condition for being a particular individual (Churchill, in my example). The ‘necessity’ in the ‘necessity of origin thesis’ is, I take it, intended to indicate that the origin of an individual is metaphysically necessary, which is to say, holds in all possible worlds: the claim, in effect, is that being a particular individual consists in possessing the relevant essence, i.e. the set of necessary and sufficient conditions, and these conditions holding across possible worlds. In other words, the sentence ‘someone is Winston Churchill if and only if they come from the (particular) gamete of Jennie Jerome, and the (particular) gametes of Randolph Churchill’ expresses the necessary and sufficient conditions for being Churchill, and further is, if true, metaphysically necessary.<sup>22</sup>

The important parallel between theoretical identification sentences and claims about the necessary origin of an individual, which makes discussion of the former important here, is that both are examples of the necessary *a posteriori*: the category of truths which, according to Kripke, have a ‘special character’ (Kripke 1980: 159) in that ‘[p]hilosophical analysis tells us that they cannot be contingently true, so any empirical knowledge of their truth is automatically empirical knowledge that they are necessary’ (*ibid.*). Since ‘the *origin* of an object is essential to it’ (Kripke 1980: 114 n57), and the origin of an object holds as a matter of metaphysical necessity, we should understand Kripke’s general notion of essence as the

---

<sup>22</sup> This interpretation of Kripke, however, is not uncontroversial: at one point Kripke states explicitly that ‘adequate necessary and sufficient conditions for identity which do not beg the question are very rare’ (Kripke 1980: 43) and that ‘I don’t know of such conditions for identity of material objects over time, or for people’ (*ibid.*).

*metaphysically* necessary set of necessary and sufficient conditions for the identity and existence of either kinds or individuals.<sup>23</sup>

Contrast now the identity statement (a) ‘Ehrich Weiss is Harry Houdini’, with the essence claim (b) ‘Ehrich Weiss is a product of the sperm and egg that generated him’. Both of these sentences, on Kripke’s view, are examples of the necessary *a posteriori*. But there are two clear points of disanalogy: firstly, there is a clear semantic dissimilarity between the names ‘Ehrich Weiss’ and ‘Harry Houdini’ on the one hand, and the descriptive phrase ‘being a product of the sperm and egg that generated him’ on the other; secondly, only the necessity of (a) is explicable via Kripke’s semantic story. The necessity of (b) hinges on whether or not we accept the necessity of origin thesis, which is an independent metaphysical hypothesis. Hence nothing in Kripke’s *semantics* motivates the claim that (b) is necessary.

But since (b) looks to be more obviously analogous to a standard example of a theoretical identification sentence, some account of why sentences with that structure are necessary (and knowable *a posteriori*) is required. Consider Kripke’s view of the natural kind gold, and its corresponding kind term:

[A] material object is (pure) gold if and only if the only element contained therein is that with atomic number 79. Here the ‘if and only if’ can be taken to be strict (necessary). In general, science attempts, by investigating basic structural traits, to find

---

<sup>23</sup> The intended contrast is a logical one: take the logical representation of necessary and sufficient conditions using a biconditional. The distinction between those that are metaphysically necessary and those that aren’t is just that the former, but not the latter, include a modal operator.

the nature, and thus the essence (in the philosophical sense) of the kind. (Kripke 1980: 138)

The passage quoted above indicates that theoretical identity sentences like ‘Gold is the element with atomic number 79’, taking the logical form of a biconditional, and represent scientific discoveries about the essences of kinds, much like claims about the origin of an individual represent discoveries about the essence of that individual. But the sort of discovery that makes a theoretical identity sentence *a posteriori* is as different from the proper names case as discoveries about the origin of an individual are. The *a posteriori* of standard identity sentences requires investigation into causal chains of reference, and, *prima facie* at least, involves no explicit metaphysical commitments.

Investigating the essences of individuals and kinds, on the other hand, requires empirical investigation into the properties of objects, in conjunction with the metaphysical notions of individual-identity and substance-identity respectively (and we have yet to even consider whether such criteria are *a priori*, and part of metaphysics proper, or *a posteriori*, being discovered, presumably, by scientists). According to Soames:

[S]cientific investigation may lead to the discovery of properties that are necessary and sufficient for membership in the kind. These properties are expressed in ‘theoretical identity sentences’ that are necessary but *a posteriori*. (Soames 2002: 244)

Theoretical identification sentences, then, merely represent important scientific discoveries in a language, but the empirical investigations that yield them concern a different type of discovery. In the proper names case, although the *a posteriori* is the product of empirical investigation, the discovery is that one object has been named twice, and was hence the subject of two distinct name acquiring transactions. The truth of an identity sentence, then, is guaranteed by simply substitution. This, is surely not the sort of discovery science makes when they find out that samples of tungsten have atomic number 74! Rather, the discovery was that the kind of substance called ‘tungsten’ had a certain physical property – the property of having 74 protons within the nucleus of a typical tungsten atom. Hence theoretical identities are much more similar to claims about the necessary origin of individuals, since both are concerned with the properties of particular objects (be they individuals or kinds).<sup>24</sup> Moreover, on the face of it at least, if descriptions like ‘the element with atomic number 74’ are rigid designators, one might reasonably expect Kripke to concede that ‘the person who originated from Jennie Jerome’s (particular) ovum and Randolph Churchill’s (particular) sperm’ is also a rigid designator. However, it is obvious that the latter is semantically nothing like the name ‘Winston Churchill’, although it might successfully refer to Churchill. Furthermore, one might also expect some alternative account of why such an expression is rigid in the first place, which would, presumably, appeal to the thought expressed by the necessity of origin: that the property of coming from Jennie Jerome’s (particular) ovum and Randolph Churchill’s (particular) sperm is a property that Winston Churchill, and only Winston Churchill, has necessarily. This represents a clear and decisive point of disanalogy between natural kind terms and theoretical identities, and proper names and identity

---

<sup>24</sup> The observation that theoretical identities are much more similar to necessity of origin claims highlights the implicit nontrivial essentialist principles underpinning Kripke’s enterprise. I will return to this issue in chapters 7 and 8.

sentences. The necessity of the former, but not the latter, appear to rely upon a particular conception of the world where individuals and kinds have some of their properties necessarily. Again, however, it must be noted that this is an independent metaphysical position, and nothing to do with semantics *per se*!

The foregoing analysis of Kripke's discussion of theoretical identification, then, makes a convincing case against the assumption that natural kind terms and theoretical identification sentences are straightforwardly analogous to proper names and identity sentences. Given the disanalogy, if the category of the necessary *a posteriori* is to apply to a wider class of cases, including the philosophically significant class of natural kind terms, then we require some alternative account of why Kripke's considerations concerning proper names and identity sentences apply to natural kind terms and theoretical identification sentences. It is with this in mind that we now turn to Soames' proposed extension of Kripke's semantics to natural kind terms.

### **3.3 Soames' positive account I**

Soames suggests two distinct initial strategies for extending Kripke's insights concerning proper names to natural kind terms. The first, Strategy 1, is to define a notion of rigidity for predicates that is a natural extension of Kripke's notion for singular terms, and then determine whether this notion of rigidity will guarantee the necessity of true theoretical identification sentences. The second, Strategy 2, is to relate each natural kind predicate to a corresponding singular term, and then define predicate rigidity in terms of singular term rigidity. Hence, a predicate will be rigid 'iff its associated singular term is rigid' (Soames 2002: 249). If the singular terms that correspond to the natural kind predicates are rigid then any true identity



sentence containing the singular terms will have a corresponding predicate analogue, and if the former is necessary then so is the latter.

In the end we shall see that Soames dismisses *both* strategies for extending Kripke's account, concluding that we should 'reserve the terminology of rigidity exclusively for singular terms' (Soames 2002: 263), and opting, instead, to argue for an analogy between proper names and natural kind terms via alternative semantic similarities. Nevertheless, I will sketch both of these ultimately rejected strategies since (i) they have received alternative treatments in the recent literature (see Gomez-Torrente 2006: 227-259 for a defence of essentialist predicates, and Ahmed 2009: 121-133 for a response), and (ii) they help to motivate Soames' final positive account.

The aim of Soames' positive thesis is to account for the modal intuitions we apparently have concerning theoretical identifications, namely that they are necessary when true. In the case of proper names the necessity of identity sentences containing them is grounded by the notion of rigid designation, and as such Soames formulates the following *desiderata*, which any satisfactory account of the semantics of natural kinds must satisfy:

- (1) The rigidity of predicates must be a natural extension of the rigidity of singular terms, as defined by Kripke.
- (2) It must have the consequence that (nearly all) natural kind predicates are rigid while (nearly all) other predicates are nonrigid.

- (3) It must play a role in explaining the necessity of true theoretical identification sentences that is analogous to the role of a rigid singular term in explaining the necessity of a true identity sentence.

Soames' first candidate for a predicate that can satisfy these criteria is an essentialist predicate.

### **3.3.1 Essentialist Predicates**

Soames defines an essentialist predicate (*EP*) as follows:

(*EP*) =<sub>df</sub> a predicate that, for all possible worlds *w* and objects *o*, if *P* applies to *o* with respect to *w*, then *P* applies to *o* in all worlds in which *o* exists.

The nub of the definition is that a predicate *P* is an essentialist predicate iff 'the property it expresses is an essential property of anything that has it' (Soames 2002: 251), and any individual that is *P* could not have existed without being *P*.

The notion of an essentialist predicate can be shown to satisfy (1) by first unpacking what Soames labels 'Kripke's test, TRS, for [the rigidity of] singular terms' (*ibid.*) and applying it to proper names, and then, by extension, to natural kind predicates. TRS states, most simply, that a singular term '*t*' is a rigid designator just when it is true that the individual

that is  $t$  could not have existed without being  $t$ , and nothing other than that individual can be the referent of ' $t$ '. Take a proper name, for instance 'Winston Churchill'. The name 'Winston Churchill' rigidly designates a particular individual iff that individual, Churchill, could not have existed without being Churchill, and no person other than Churchill can be the referent of the proper name 'Winston Churchill'.

The predicative version can now be straightforwardly articulated, in conjunction with some essentialist thesis about the properties of the individual or kind in question. To stick with the Churchill example, as we have already seen Kripke claims that individuals have their actual origin necessarily (the necessity of origin thesis). Thus we can say that given the facts about Churchill's actual origin, no other individual, in some other possible world, qualifies as being Churchill unless they have precisely the origin Churchill has in the actual world, and only Churchill has his particular unique origin.

Now, take Churchill's origin,  $O$ , and consider the predicate 'is the man with origin  $O$ '. We already know that 'Winston Churchill' rigidly designates the man with origin  $O$ , but now we can see that there is a corresponding predicate 'is the man with origin  $O$ ' that will refer in all and only worlds where there exists a person with the origin that Winston Churchill actually has. So, just as the name 'Winston Churchill' refers to Churchill in all worlds where the man with origin  $O$  exists, so too does the predicate 'is a man with origin  $O$ '. The predicate satisfies the predicative version of TRS, which is as follows: let a predicate  $p$  be a rigid designator just when it is true that the individual that has the property expressed by  $p$  could not have existed without that property, and nothing other than that individual can have the property expressed by  $p$ . Where  $p$  is the predicate that expresses the property that is Churchill's essence, (EP) satisfies (1), and is thus a natural extension of the rigidity of singular terms.

The extension of this thesis to natural kind predicates is fairly straightforward. In the case of natural kinds and theoretical identification sentences we have two essentialist predicates, for example ‘is gold’ and ‘is the element with atomic number 79’, where the theoretical identity takes the logical form of  $\forall x[Ax \leftrightarrow Bx]$ , where  $A$  is the predicate ‘is gold’ and  $B$  the predicate ‘is the element with atomic number 79’. According to the definition (*EP*),  $A$  and  $B$  are predicates that refer to essential properties of the objects that bear them, and are hence essentialist predicates. Further, we have discovered that both predicates apply to the same set of objects at this world, call it  $w$ . Suppose that for some object  $o$ , both  $A$  and  $B$  apply to  $o$  with respect to  $w$ , then by (*EP*) both  $A$  and  $B$  apply to  $o$  in every possible world in which  $o$  exists.

To satisfy (2), the claim that natural kind predicates are rigid, whilst all other predicates are not, the essentialist predicate account must presuppose ‘a substantive metaphysical claim about kinds – namely, that if  $K$  is a natural kind, then it is a necessary property of  $K$  that anything that is an instance of  $K$  is essentially an instance of it’ (Soames 2002: 254). Without the presupposition there is a risk that *all* kind terms, and all the predicates corresponding to them, will turn out to be rigid. Take the predicate ‘is coffee’. Everything to which the predicate ‘is coffee’ applies to – the extension – has the property of *being coffee*, which is expressed by the predicate. Moreover, the predicate will apply, across all possible worlds, to just the set of things that have that property.

But, this will hold for *all* predicates, since there will always be some property,  $P$ , expressed by the predicate, ‘is  $P$ ’, that holds trivially of just the set of things within the extension of that predicate. This problem has been labelled by Dan Lopez De Sa as ‘the

trivialization problem' (2007: 117).<sup>25</sup> The benefit of the metaphysical claim, then, is that since it is only *natural* kinds that have the necessary property that anything that is an instance of that kind is essentially an instance of it, only natural kind predicates will be essentialist predicates. The alleged contrast is with non-natural kinds, for instance the wristwatch kind, where no particular instance of the wristwatch kind is *necessarily* an instance of that kind. For instance, if it irrevocably stops to function as a wristwatch, and is used rather as an ornament, then it is plausible to suppose that it is no longer an instance of the wristwatch kind. If that is right, then as a corollary, since the object in question is no longer an instance of the wristwatch kind, it cannot have been a necessary property of that object that it was.<sup>26</sup> The cost, of course, is that such a theory of reference *requires* a highly controversial metaphysical assumption in order to work, and one might think (plausibly) that an account of the basic semantic function of a common set of terms within our language should be metaphysically neutral (cf. Salmon 2005).

The satisfaction of (3), the final *desideratum*, requires the notion of an essentialist predicate to play a role in the explanation of the necessity of true theoretical identification sentences (that is analogous to that for singular terms). As we noted earlier, however, Kripke's theoretical identifications have a different logical form to standard examples of identity between proper names, such as 'Hesperus is Phosphorus'. Consider some of Kripke's

---

<sup>25</sup> Both Soames (2002) and Schwartz (2002) notice problems with trivialisation, as does LaPorte (2004). However, the labelling of the problem is López De Sa's, so I have attributed it to him. For some further discussion of this problem see also Martí and Martínez-Fernandez in Beebe, H. and Sabbarton-Leary, N. (eds.) (2010).

<sup>26</sup> It should be noted that the claim that for some object *o*, where *o* is a member of the natural kind *NK* in virtue of possessing *E*, where *E* is the essence of *NK*, it is a further and far more controversial claim to say that *o* bears *E* necessarily. *Prima facie* it looks to be false. For example, where *o* is a member of the kind Thorium 232, it is possible for *o* to undergo radioactive (alpha) decay, emitting an alpha particle, and transmuting into a member of the kind Radium 228. However, if *o* can go from being an instance of Thorium 232 to Radium 228, then *o* cannot have been a member of the Thorium 232 kind necessarily!

paradigm theoretical identifications, for instance ‘light is a stream of photons’ (1980: 116), or ‘a material object is (pure) gold if and only if the only element contained therein is that with the atomic number 79’ (*ibid.*: 138). These theoretical identifications have the logical form of biconditionals:  $\forall x (Ax \leftrightarrow Bx)$ <sup>27</sup>. Why should we treat these as identity sentences?

Soames’ argument is simple: the absence of the identity predicate does not disqualify sentences with a distinct logical form from being *identity*-sentences. Sentences such as ‘an A is a B’ and ‘something is an A iff it is a B’ are still, in some sense, identifying As with Bs, and thus still qualify as a species of identity sentence. Nevertheless, they do require some alternate explanation of why they are, when true, necessary. With this in mind Soames produces the following argument for the necessity of theoretical identifications with this distinctive logical form:

- (i)  $\forall x (Ax \leftrightarrow Bx)$  is true.
- (ii) The predicates ‘A’ and ‘B’ are rigid designators – that is, for all possible worlds  $w$  and objects  $o$ , if the predicate applies to  $o$  in  $w$ , then it applies to  $o$  in all worlds in which  $o$  exists.
- (iii)  $\forall x (Ax \leftrightarrow Bx)$  is necessary – that is,  $\Box \forall x (Ax \leftrightarrow Bx)$  is true.

However, Soames identifies two problems with the above formulation. The first is that the argument is invalid. What follows from (i) and (ii) is not (iii), but (iv):

---

<sup>27</sup> Soames also considers, to begin at least, representing theoretical identities as universally quantified conditionals i.e.  $\forall x (Ax \rightarrow Bx)$ . However, since he eventually opts to use only biconditionals in his conclusion, I have omitted to former for simplicity.

(iv)  $\forall x \Box (Ax \leftrightarrow Bx)$  is true.

But now, although the argument from (i) and (ii) to (iv) is valid, (iv) does not guarantee the necessity of theoretical identification sentences since the modal operator has a more limited scope. The scope of the necessity operator in (iv) fails to guarantee that there is not some possible world 'w in which some object that doesn't exist in the actual world exists, and is an A in w, without being a B in w' (Soames 2002: 258). Consider the previous example, where A stands for 'is gold', and B stands for 'is the element with atomic number 79'. Fleshing out the story we found that since both A and B apply to an object *o* in this world, *w*, A and B apply to *o* in all possible worlds in which *o* exists. But nothing in Soames' argument on behalf of the (EP) model rules out there being some other object *o<sub>I</sub>* in some other possible world *w<sub>I</sub>*, such that it is an A in *w<sub>I</sub>* without being a B in *w<sub>I</sub>*. So, even when (iv) is true, the claim that true theoretical identifications sentences involving rigid designators are necessary will turn out to be false. Let a) be a definition of the compound predicate 'F', consisting of the disjunction of three essentialist (and thus rigid) predicates, and let b) be the definition of the predicate '...is a fundamental particle', to which the predicates in a) are related species terms:

a) *F* (...is a photon or ...is a muon or ...is an electron).

b) *G* (...is a fundamental particle)

Consider, then, the following argument, given the definitions stipulated in a) and b):

c)  $\forall x [Gx (x \text{ is a fundamental particle}) \leftrightarrow Fx (x \text{ is a photon or } x \text{ is a muon or } x \text{ is an electron})]$ .

Therefore,

d)  $\forall x \Box [Gx (x \text{ is a fundamental particle}) \leftrightarrow Fx (x \text{ is a photon or } x \text{ is a muon or } x \text{ is an electron})]$ .

Assuming that c) is true, and there really are only three species of elementary particle, then from c) we can derive d) since each actually existing fundamental particle is essentially a fundamental particle, and essentially a photon, muon or an electron. The problem with the argument is that it doesn't preclude *additional* species of fundamental particles at other possible worlds. So, imagine some other possible worlds where there are additional fundamental particles, say positrons. In these worlds there will be individuals that are fundamental particles but *not* any of the three species that constitute fundamental particles in our world. Thus the truth of c) does not guarantee the necessity of d), since there are worlds in which d) can be false. If this is correct then the notion of an essentialist predicate does not satisfy *desiderata* (3), since theoretical identities do not turn out to be necessary. Hence (EP) does not capture the analogue of the notion of rigidity for singular terms. This effectively makes strategy 1 unsuccessful as an extension of Kripke's thesis.

### 3.3.2 Correspondence rigidity



With strategy 1 unsuccessful Soames moves on to consider strategy 2: finding for each natural kind predicate a corresponding singular term which *is* rigid, and using the rigidity of that singular term to derive a corresponding rigidity for the related predicate. So, given a natural kind predicate  $P$ , we find the associated singular term  $st$ , and when it is both necessary and *a priori* that  $\forall x (x \text{ is } P \text{ iff } x \text{ is an instance of } st)$  then the rigidity of  $P$  can be defined in terms of the rigidity of  $st$  (Soames 2002: 259). So, if  $P$  is the natural kind predicate ‘is water’ and  $st$  is the corresponding singular term<sup>28</sup> ‘water’, provided the latter is a rigid designator we can derive the rigidity of the predicate ‘is water’ from the rigidity of the singular term ‘water’ – the predicate ‘is water’ only applies to instances of the kind water, and nothing else.

The clear benefit of strategy 2 is that it allows the construction of theoretical identity sentences like ‘(the kind) water is (the kind) H<sub>2</sub>O’ that are analogous to identity sentences containing proper names, for instance ‘Ehrich Weiss is Harry Houdini’. Once the necessity of the theoretical identity containing singular terms is established we reason as follows: if  $P$  is the predicate corresponding to the singular term ‘water’ and  $Q$  is the predicate corresponding to the singular term ‘H<sub>2</sub>O’, and singular terms are, as Kripke said, rigid designators, then it will follow from the truth of the identity sentence ‘water is H<sub>2</sub>O’, the rigidity of those singular terms, and the rule for defining predicate rigidity in terms of the rigidity of some corresponding singular term, that *necessarily* for all  $x$ ,  $x \text{ is } P \text{ iff } x \text{ is } Q$  i.e.  $\Box \forall x (Px \leftrightarrow Qx)$ .

The key here is the truth of the identity of the singular terms denoting the kind. Where  $P$  and  $Q$  are predicates that refer to their extensions, and  $o$  is an object that is referred to by both  $P$  and  $Q$ , it still might be the case that  $P$  and  $Q$  are merely coextensive, rather than identical. The identity must be formulated in terms of the corresponding singular terms, such as ‘the kind

---

<sup>28</sup> Soames actually uses ‘the kind water’ as his singular term, but we lose nothing by using ‘water’ instead, since ‘water’, if it designates anything, plausibly designates the kind water.

water' and 'the kind H<sub>2</sub>O'. If an identity between two such terms is true, then since the corresponding predicates *P* and *Q* only refer to objects that are instances of the kind water and the kind H<sub>2</sub>O, theoretical identities involving predicates will turn out to be necessary.

The main problem with the account (as Soames notes) is that it will not just be natural kind predicates for which we can locate corresponding singular terms. Consider the predicate 'is a philosopher' and the corresponding singular term 'the kind philosopher'. Since the predicate 'is a philosopher' has a corresponding singular term 'the kind philosopher', and given that singular terms are always rigid designators, it will follow that the predicate 'is a philosopher' is also a rigid designator. Strategy 2, then, is also liable to Lopez De Sa's 'trivialization problem', and hence fails to satisfy *desiderata* (2): it does not maintain an asymmetry between natural kind terms (predicates or otherwise) and other general terms. Thus it is not a successful extension of Kripke's semantic thesis.

Since both initial strategies are unsuccessful, Soames concludes that the notion of rigidity cannot carry the analogy between proper names and natural kind terms, and that there is no satisfactory notion of rigidity for natural kind predicates. This invites further analysis of the semantics of natural kind terms to reveal an alternative mechanism that satisfies *desideratum* (1) – (3), and which does justice to our apparent confidence in the necessity of true theoretical identification sentences.

### **3.4 Soames' positive account II**

Given the two unsuccessful strategies Soames proposes to 'reserve the terminology of rigidity exclusively for singular terms', and ground the relation between singular terms and natural

kind predicates via some other semantic similarity (2002: 263). As Soames reads Kripke, the central thesis in *Naming and Necessity* is not that names and natural kind terms are both rigid designators, but that names and natural kind terms are both *nondescriptive*. The nub of Kripke's thesis on this reading is that since names are nondescriptive the denotation of a name (an individual) is not required to satisfy any descriptive conditions in order to qualify as the referent of that name. Thus nondescriptive terms get to be rigid precisely because they (a) have no descriptive content, and (b) are introduced with the *intention* that they are to refer to precisely this individual. Hence, the 'doctrine that names are rigid designators may be viewed as a corollary of the more central thesis that they are nondescriptive, together with an account of how their reference is fixed in the actual world' (Soames 2002: 264).

Soames' account of natural kind predicates maintains that the important similarity between names and natural kind terms is that they are both nondescriptive, and that both have broadly the same metasemantics – so a speaker, *s*, associates a predicate *p* with a kind *K* by ostending a sample of individuals, *x*, *y* and *z*, all of which are presumed to be instances of *K*; once *K* has been determined 'it is understood that for any world *w* the extension of the predicate at *w* is to be the set of instances of *K* at *w*' (Soames 2002: 267). The thesis that both proper names and natural kind terms (predicates or otherwise) are nondescriptive is what we can call *Extended Millianism*:

*Extended Millianism (EM)*: the thesis that 'the meaning (semantic content) of a simple natural kind predicate is the natural kind it designates' (Soames 2002: 278).

According to (*EM*) a speaker associates a natural kind predicate *p* with certain individuals with which she is acquainted, and which she presumes to be members of the same kind.<sup>29</sup> The introduction of the predicate comes with the explicit intention that *p* apply to all and only things of the same kind as the individuals with which the speaker is acquainted. Alternatively, a speaker may employ a description to help affix the natural kind predicate to the individuals which she presumes to be members of the same kind, but in such cases although the description is taken to be indicative of the kind, it is not thought to be synonymous with the introduced predicate. Consider the predicate ‘is gold’. The actual introduction of this predicate is lost to history, but imagine the Romans were right, and the god Mercury discovered gold. Mercury, as our typical language user, introduced the predicate ‘is gold’ with the intention that it apply to all and only stuff of the same kind as *that* stuff (the ostended samples), with which he was acquainted.

The thesis (*EM*) stems, in part, from what Soames takes to be part of Kripke’s ‘unfinished semantic agenda’ (Soames 2002: 3) from *Naming and Necessity*, namely to give an account of the semantic content of proper names and, by extension, simple natural kind terms. Kripke, as we have already seen in chapter 1, endorsed a modified version of Millianism, the thesis that (logically proper) ‘names have denotation but not connotation’ (Kripke 1980: 25). A Kripkean semantics is essentially Millianism, extended to natural kind terms (which Mill thought of as connoting expressions). Kripke construes Mill as ‘more or less right’ (Kripke 1980: 127) about names, but wrong about general terms. Where Mill treated general terms as connotative, Kripke extends Mill’s thesis about names to general

---

<sup>29</sup> The individuals, presumably, get to be members of the kind by being instances of it. So, for some individual *x*, *x* is a member of the gold-kind if (and only if) it instantiates the property of *being gold*. In effect, then, the implicit assumption here is that kinds are property like abstract objects. The assumption is made explicit later in Soames, and will be discussed here in due course.

terms, claiming that they, like names, have denotation but no connotation. Since neither proper names nor (simple) natural kind terms have connotation, but only denotation, the semantic content of either category of term is simply the referent of the relevant term. Both are what we might call purely referential expressions.<sup>30</sup>

Soames, then, follows Kripke in endorsing a modified Millianism, extended from proper names to natural kind terms. It is important to note, though, that as Soames interprets Kripke, Kripke never really constructed a *theory* of the semantic content of either names or natural kind terms. On this point, Soames view is plausible; after all, Kripke states quite explicitly ‘You may suspect me of proposing a theory; but I hope not, because I’m sure it’s wrong too if it is a theory’ (Kripke 1980: 64). But if that is right, it is far from obvious that Kripke really did have a solution to Frege’s Puzzle (see Chapter 1, section 5), and hence not even obvious that Kripke *justified* that identities and theoretical identities express propositions that are actually knowable only *a posteriori*. After all, if the meaning of a term just is its referent, then since, for example, ‘Batman’ and ‘Bruce Wayne’ are co-referential, the identity sentence ‘Batman is Bruce Wayne’ expresses the same proposition as ‘Batman is Batman’, and so should be knowable *a priori*. Putting an answer to this worry to one side, this observation makes Soames’ account quite unlike Kripke’s insofar as it attempts to give an account of the semantic *content* of proper names (and natural kind terms), and argues for an analogy between proper names and natural kind terms on the basis of their nondescriptive semantic content. He then uses this similarity to ground our modal intuitions about theoretical identification sentences. In the following section I will sketch his account of the necessity of

---

<sup>30</sup> It is not completely obvious *what* Kripke thinks about the semantic content of names and natural kind terms. However, what he says about Mill at least suggests that buys Mill’s view of content.

theoretical identification sentences, but first there are some problems and confusions to address, concerning (*EM*).

Firstly, the confusion that stems from Soames' perplexing articulation of (*EM*). Prior to the introduction of (*EM*) as a semantic thesis, Soames is fairly consistent in claiming that natural kind predicates have, as their extensions, sets of instances of the kind with which the introducer of the predicate was acquainted. Natural kind predicates are the sorts of term that a speaker can use 'to say something about members of a kind, even if he lacks the ability to accurately describe either the kind itself or its instances' (Soames 2002: 266). Moreover, when natural kind predicates are introduced into a language, they are stipulated to 'apply to all and only instances of the unique natural kind' of which the samples employed to introduce the predicate are instances (*ibid.*: 267). Hence it sounds as though natural kind predicates designate their extensions.

However, Soames also talks about natural kind predicates *designating* kinds (*ibid.*: 272), and indeed defines (*EM*) in these terms. For Soames, natural kind predicates are semantically similar to proper names, like 'Mercury', in that they are denoting (rather than connoting) expressions, and fundamentally nondescriptive. The predicate 'is gold' nondescriptively *applies* to all and only the stuff which is of the same kind as the stuff with which Mercury is already acquainted – instances of the kind – and *designates* the kind itself.<sup>31</sup> But as Michael Devitt notes 'it is not obvious that they [natural kind terms] designate at all' (2005: 139). Similarly, as Nathan Salmon notes, there 'are very forceful reasons – due

---

<sup>31</sup> There is perhaps an obvious parallel here between Soames' account and Putnam's account of natural kind terms. Indeed Nathan Salmon (2003) raises this point as an objection to Soames, observing that Soames' account relies upon the same metaphysical essentialism that Putnam's does, and that this is objectionable since a semantics should be metaphysically neutral on these sorts of issues.

independently to Church, Gödel, and ultimately Frege – for taking predicates to designate their semantic extensions’, rather than as designating the kinds themselves (Salmon 2003: 480). Admittedly, the extension of a kind is *not* a good candidate for the reference of a predicate if we are trying to capture something like the notion of rigidity, whilst also endorsing (*EM*), since general terms, like ‘tungsten’ or ‘bear’, cannot rigidly designate their extensions, when the class of objects that is the extension of a predicate varies from world to world. Moreover, extension tends to vary within a world over time. Consider the extension of ‘is a human’, which is the class of humans. Since the population is constantly changing, if the meaning of a predicate is its extension, ‘is human’ will mean something different every time the population changes – this surely cannot be right.

How, then, should we understand Soames’ thesis? It is clear from Soames’ rejection of strategy 1, the essentialist predicate strategy, that he is aware of the problems associated with marrying rigidity to extension: whilst all actual samples of a kind may be within the extension of the corresponding kind predicate, there are additional possible samples that are not. Hence, the notion of a rigid predicate is undermined. Similarly, Soames dismissed strategy 2, the correspondence rigidity strategy, on the grounds that for the vast majority of predicates there is a corresponding singular term (e.g. ‘is a coffee mug’ and ‘the kind coffee mug’), which undermines the required asymmetry between kind terms and *natural* kind terms.<sup>32</sup>

The upshot of the two failed strategies, recall, was to reserve the notion of rigidity for singular terms. Does that mean that natural kind predicates are not rigid designators? Well, since they are nondescriptive and introduced by ostension and stipulation – making them analogous to proper names – and, for Soames, rigidity is a corollary of being nondescriptive,

---

<sup>32</sup> Inviting the so-called ‘trivialisation problem’, which I discuss in chapter 4, section 8.

presumably predicates are rigid (Soames 2002: 264). What Soames thesis appears to be, then, is that for any arbitrarily selected natural kind predicate  $p$ , introduced with the appropriate metasemantics (by ostending samples of some natural kind  $NK$ , or via reference fixing description),  $p$  applies to all instances of the kind  $NK$  –  $x$ ,  $y$  and  $z$  say – but designates the kind  $NK$ ; moreover, since  $p$  is nondescriptive,  $p$  rigidly designates  $NK$ . That this reading of Soames is correct is given additional support by his claim that ‘properties are the meanings of predicates’ and his initial treatment of natural kinds as properties (2002: 278-9).<sup>33</sup> So, if  $p$  is the predicate ‘is tungsten’,  $p$  applies to all and only instances of the tungsten kind. Since objects qualify as members of the tungsten kind,  $NK$ , only in virtue of possessing the property *being tungsten*, and the content of  $p$  is defined by (EM), the meaning of ‘is tungsten’ will be the property *being tungsten*.

But this raises a second, obvious problem. According to Millianism, the meaning of a nondescriptive proper names is just its referent. Soames has extended this thesis of semantic content to natural kind predicates. Hence, if two simple natural kind predicates are co-referential then they have the same semantic content, since the meaning of a natural kind predicate just is the natural kind it designates. Take the theoretical identity sentence ‘for all  $x$ ,  $x$  is a Woodchuck iff  $x$  is a Groundhog’, which contains two simple natural kind predicates. Since the two natural kind predicates are co-referential, the proposition expressed by ‘for all  $x$ ,  $x$  is a Woodchuck iff  $x$  is a Groundhog’ is the same proposition as that expressed by ‘for all  $x$ ,  $x$  is a Woodchuck iff  $x$  is a Woodchuck’. In effect, the two theoretical identification

---

<sup>33</sup> Although Soames goes on to change his mind slightly, claiming that although the meanings of natural kind predicates are properties, the natural kinds themselves are not properties but *intensions* (functions from worlds to extensions). He then goes on to distinguish simple natural kind predicates from complex natural kind predicates, claiming that the former designate natural kinds (intensions) directly, whilst the latter designate natural kinds *indirectly*, since the meaning of a complex predicate is the property that determines the kind (2002: 279). I will discuss this distinction in significantly more detail in section 9 of this chapter.



sentences are synonymous. However, since the latter is knowable *a priori* so too must the former be. Hence it is a consequence of endorsing (*EM*) that identity sentences and theoretical identity sentences containing two simple natural kind predicates turn out to be knowable *a priori*, rather than being knowable only *a posteriori* as Kripke argued.<sup>34</sup>

Soames' effectively concedes the point that the proposition expressed by sentences containing two co-referential names or natural kind terms are knowable *a priori*, but maintains that this is consistent with the claim that in certain contexts an assertive utterance of such a sentence could well communicate *a posteriori* information (2002: 279). Consider the following two theoretical identity sentences: (a) 'for all x, x is a Moose iff x is an Elk', and (b) 'for all x, x is a Moose iff x is a Moose'. Given (*EM*), the thesis that the meaning of a term just is its referent, the proposition expressed by (a) is the same as that expressed by (b), since 'is a Moose' refers to the same species as 'is an Elk'. Thus, it is a logical consequence of (*EM*) that just as the proposition expressed by (b) is knowable *a priori*, so too is the proposition expressed by (a), since (a) and (b) express the same proposition. However, observes Soames, 'in different contexts [assertions made of one] would standardly convey, and even result in the assertion of, a posteriori propositions that would not be' (Soames 2002: 279) asserted of the other.

The crux of the move is to concede that both names and natural kind terms are Millian, and that when such terms are co-referential they are synonyms, but to maintain that when speakers utter simple sentences containing such terms, they assert more than just Russellian

---

<sup>34</sup> It is worth noting that this consequence stems from endorsing Mill's account of the semantic content of denoting expressions. Thus if Kripke's original intention *was* to endorse this Millian account of semantic content, then identities like 'Hesperus is Phosphorus' are not obviously *a posteriori*. However, Kripke did not give an explicit account of the *meaning* of names and/or natural kind terms, so it is not obvious that the criticism applies to Kripke.

structured propositions, i.e. a proposition that has an individual (the referent) as its constituent. So, when a speaker utters a sentence about, say, an elk, she does not merely assert the singular proposition that has the referent of the natural kind predicate 'is an elk' as a constituent, such that the substitution of 'is a moose' for 'is an elk' would have no impact on the content of her utterance. Rather she asserts, in addition to the singular proposition, various general or descriptive propositions that, as far as she is concerned, may be true of everything within the extension of 'is a moose', but not true of things within the extension of 'is an elk'. For instance, our typical speaker might know that a bull moose sheds its antlers seasonally, but that if castrated it will shed its antlers, and then grow a single misshapen set that it keeps for the rest of its life (Geist 1998). However, they might not associate this information with things within the extension of 'is an elk'.

Let us represent the descriptive information that our language user, *Lu*, associates with the predicate 'is a moose' using the sentence letter *D*; according to the story told above, *Lu* associates *D* with 'is a moose', but not with 'is an elk', and hence when she asserts a sentence containing 'is a moose' part of what she intends to convey (the content of her sentence/utterance) is *D*. The solution, then, is to concede that although (a) is *a priori*, assertions containing the natural kind predicates 'is a moose' or 'is an elk' will tend to convey *a posteriori* information such that standard language users will not think that (a) expresses the same proposition as (b). Whilst this does not save the Kripkean insight about the *a posteriori* of identity and theoretical identity sentences, it does offer some account of why we think that (a) is *a posteriori*, despite (a) expressing the same proposition as (b).<sup>35</sup>

---

<sup>35</sup> It should also be noted that a proponent of Millianism is committed to conceding that 'Hesperus is Phosphorus' expresses the same proposition as 'Hesperus is Hesperus', since the two names are co-referential, and hence both contribute the same referent to any proposition in which they figure. It was this problem with

However, as we shall see in section 3.6, Soames does *not* adopt this solution to the second class of theoretical identification sentences, namely those containing complex natural kind terms such as ‘H<sub>2</sub>O’ or ‘the element with atomic number 79’.

### 3.5 Simple theoretical identification sentences

My central line of objection to Soames’ thesis is to his theory of types, and his assumption that such a notion of ‘type’ is true. In this section I show that Soames’ notion of ‘type’ is actually false, which will effectively undermine his argument for the necessity of theoretical identities, since the notion of ‘type’ is doing the heavy lifting in establishing that necessity. I will begin by showing how the argument is clearly invalid without Soames’ notion of ‘type’, and then use this to analyse precisely what Soames has in mind when he introduces the notion. The conclusion of this section will be that Soames fails to establish the necessity of theoretical identities linguistically.

Theoretical identities, according to Soames’ analysis, come in two semantic varieties. The first, which I call the *simple-simple type*, involve a pair of simple natural kind predicates, for instance ‘ $x$  is a Woodchuck iff  $x$  is a Groundhog’. The second, what I call the *simple-complex type*, involves a simple natural kind predicate and a complex natural kind predicate, for example ‘ $x$  is tungsten iff  $x$  is the element with atomic number 74’. Here I focus only on the former, the simple-simple type, but Soames’ notion of type plays the same role in arguments for the necessity of simple-complex theoretical identities.

---

Millianism that generated Frege’s Puzzle, and encouraged the development of descriptivist theories of meaning and reference.

A simple-simple theoretical identity sentence consists of two simple natural kind predicates. A simple natural kind predicate, typically, consists of a single count noun, for instance ‘Apatosaurus’, modified with the copula ‘is’, to form the predicate ‘is an Apatosaurus’. Using LaPorte’s (2004: 36) example to illustrate: let  $A$  stand for the predicate ‘is an Apatosaurus’ and let  $B$  stand for the predicate ‘is a Brontosaurus’. Assume that  $A$  was introduced via ostension with the explicit intention that it apply to all and only instances of the unique natural kind of which *this* is an instance, where ‘*this*’ picks out a sample within the environment. Let  $B$  be introduced more loosely, by ostension or reference-fixing description, but let  $B$  designate a kind of the same type as  $A$  (so, if  $A$  designates a species kind then so too does  $B$ ). Now, suppose that after some empirical investigation we discover that everything within the extension of  $A$  is also within the extension of  $B$ . How, given co-extension, do we firstly conclude that ‘all  $A$ s are  $B$ s’, and secondly move from this discovery to the modal version of the claim i.e. that *necessarily* all  $A$ s are  $B$ s?

Soames employs the following argument, call it the *Necessity Argument*, to extend Kripke’s insights into proper names and identity sentences to natural kind predicates and theoretical identity sentences:

...we reason as follows: (i) From the assumption that the ostensive natural kind predicate  $A$  has successfully been introduced it follows that there is a unique natural kind  $K_A$  (of a certain type  $T$ ) of which nearly all members of the sample associated with  $A$  are instances, and  $A$  applies (at a world) to all and only instances of  $K_A$  (at the world). (ii) From the assumption that the natural kind predicate  $B$  has successfully been introduced, it follows that there is a natural kind  $K_B$  which is such that  $B$  applies (at a world) to all and only

members of  $K_B$  (at the world). (iii) By hypothesis, the two predicates designate kinds of the same type  $T$ ; thus both  $K_A$  and  $K_B$  are species, or both are substances, or both are kinds of some other category. (iv) If the theoretical identity sentence [All  $A$ s are  $B$ s] is true, then (since nearly all objects in the  $A$ -samples are  $A$ s) nearly all of the objects in the  $A$ -sample are  $B$ s, and hence they are instances of kind  $K_B$  as well as kind  $K_A$ . (v) Since the  $A$ -sample determines a single kind (of the given type  $T$  – species, substance, etc.), of which nearly all members of the sample are instances, it follows that kind  $K_A = K_B$ . (vi) But this means that in addition to [All  $A$ s are  $B$ s], [All and only  $A$ s are  $B$ s] must also be true. (vii) Moreover, both must be necessary, since from steps (i), (ii), and (v) it follows that for all worlds  $w$  the extension of  $A$  at  $w$  = the set of instances of  $K_A$  at  $w$  = the set of instances of  $K_B$  at  $w$  = the extension of  $B$  at  $w$ . (viii) In short, if  $A$  and  $B$  are natural kind predicates of the sorts indicated, then a theoretical identity sentence [All and only  $A$ s are  $B$ s] involving those predicates is necessary, if the corresponding sentence [All  $A$ s are  $B$ s] is true. (Soames 2002: 268)

If we ignore premise (v) of the preceding argument, then we can see that the argument is obviously invalid. Without premise (v) there is an inference from ‘All  $A$ s are  $B$ s’ to ‘All and only  $A$ s are  $B$ s’ – i.e. from  $\forall x[Ax \rightarrow Bx]$  to  $\forall x[Ax \leftrightarrow Bx]$  – that is erroneous. Since the argument is supposedly valid with premise (v), let me first represent the argument more formally without (v), in a bid to cash out precisely what (v) amounts to. The argument is as follows:

(a) Let  $P$  and  $Q$  be natural kind predicates that have been successfully introduced.

- (b) If  $P$  and  $Q$  have been successfully introduced, then there is a unique kind  $K_P$  corresponding to  $P$ , and a unique kind  $K_Q$  corresponding to  $Q$ .
- (c) Let  $P$  and  $Q$  be natural kind predicates designating kinds of the same type  $T$  (i.e. species predicates, or substance predicates).
- (d) Let  $\forall x [Px \rightarrow Qx]$  be true (on the basis of empirical investigation).

Given (a) and (b),

- (e) The extension of  $P$  at a world  $w$  is the set of individuals that are members of  $K_P$ , and the extension of  $Q$  at  $w$  is the set of individuals that are members of  $K_Q$ .

From (e) and the truth of (d) it follows that,

- (f) since (nearly) all  $P$ -samples are  $P$ s, and (nearly) all  $P$ -samples are  $Q$ s, that (nearly) all  $P$ -samples are instances of  $K_P$  and  $K_Q$ .

Thus, from (b), (c) and (f) it follows that,

- (g)  $K_P = K_Q$ .

And therefore,

- (h)  $\Box \forall x (Px \leftrightarrow Qx)$ .

The invalidity of the argument is now clear. The derivation of (g) and consequently (h) from (a)-(d) just does not follow. The biconditional at (h) is effectively being derived from a single conditional, rather than from the conjunction of two conditionals:  $\forall x [Px \leftrightarrow Qx]$  only

follows from the conjunction of  $\forall x [Px \rightarrow Qx]$  and  $\forall x [Qx \rightarrow Px]$ . Since the inference to (h) is straightforwardly invalid, Soames' argument devoid of its notion of 'type' fails.

The problem stems from Soames' claim that 'finding out that everything in the extension of  $[P]$  is in the extension of  $[Q]$  is tantamount to finding out that the kind associated with  $[P]$  is identical with the kind associated with  $[Q]$ ' (2002: 269). This is not the case. It fails to follow from the fact that everything within the extension of  $P$  is within the extension of  $Q$  that everything within the extension of  $Q$  is within the extension of  $P$ . For instance, if  $P$  is the predicate 'is nickel' and  $Q$  the predicate 'is metal', then whilst everything within the extension of 'is nickel' is indeed within the extension of 'is metal' not everything within the extension of 'is metal' is within the extension of 'is nickel'.

Moreover, even when we have two predicates that are actually co-extensional, an inference to identity would still not be warranted. If  $P$  is the predicate 'is a renate' and  $Q$  the predicate 'is a cordate' then despite everything within the extension of  $P$  being within the extension of  $Q$ , and everything within the extension of  $Q$  being within the extension of  $P$ , this is insufficient to ground an inference to the identity of the renate-kind and the cordate-kind (since there will be close possible worlds where there are creatures with hearts but no kidneys).

What the forgoing analysis reveals, then, is just what work Soames notion of 'type' is doing. In the original argument we are offered, as a paradigm case, a situation where we discover that one predicate,  $P$ , has total extensional overlap with another,  $Q$ , but not that the latter,  $Q$ , has total extensional overlap with the other,  $P$ . As I have demonstrated, a version of the argument without (v) invalidates the inference to identity, hence (v) is vital for deriving the necessity of true theoretical identities. But what manner of premise is (v)? Premise (v)

states that ‘individuals in the sample associated with the ostensive predicate [ $P$ ] are members of a single natural kind of a given type’ (Soames 2002: 269). From this, and the observation that the two predicates,  $P$  and  $Q$ , designate kinds of the same type, it allegedly follows that the kinds are identical. Premise (v), as Soames concedes, is a ‘very substantial non-linguistic claim’ (2002: 269) about kinds that accompanies the introduction (or metasemantics) of a natural kind predicate. With the analysis in place I now propose to distil Soames’ notion of ‘type’, and show why it is false.

### 3.6 Kinds of a type $T$ ?

Soames claims, explicitly, that:

If the theoretical identity sentence [All As are Bs] is true, then (since nearly all objects in the A-sample are As) nearly all of the objects in the A-sample are Bs, and hence they are instances of kind  $K_B$  as well as kind  $K_A$ . (v) Since the A-sample determines a single kind (of the given type  $T$  – species, substance, etc.), of which nearly all members of the sample are instances, it follows that kind  $K_A = K_B$ . (Soames 2002: 268)

According to the quoted passage, if we have the true conditional  $P \rightarrow Q$ , and there is a kind  $K_P$  and a kind  $K_Q$  to which the predicates correspond, then all  $P$ -samples are instances of  $K_Q$  as well as  $K_P$ . So far, so good. The nub of the inference to identity is that since the  $P$ -samples determine a single kind of a type  $T$  of which all correctly identified members of the sample



are instances, it follows that the kinds  $K_P$  and  $K_Q$  are identical. But how does being a kind of a type  $T$  do so much work?

Soames says little of what being a kind of a particular type consists in, saying only that where two kinds, say  $K_P$  and  $K_Q$ , are of the same type, then both are, for example, substance kinds (for substances) or species kinds (for species). Presumably what Soames has in mind is something like the following: where  $A$  is the predicate ‘is an Apatosaurus’ and  $B$  is the predicate ‘is a Brontosaurus’, and  $K_A$  is the kind designated by the former, whilst  $K_B$  is designated by the latter, since both kinds are of the same type – species kinds – the discovery that if something is within the extension of  $A$  then it is also within the extension of  $B$  is sufficient to infer the identity of  $K_A$  and  $K_B$ . But this, as it stands, does little to plug the invalidity of the argument form without (v).

Consider a variation on an earlier example: let  $P$  be the natural kind predicate ‘is gold’, let  $K_P$  be the natural kind gold, and assume that  $P$  has been successfully introduced (which means that it does pick out individuals that are members of a natural kind) by ostending samples within the speaker’s environment. Let  $Q$  be the natural kind predicate ‘is metal’, let  $K_Q$  be the natural kind metal, and assume also that  $Q$  has been successfully introduced. We have discovered that ‘all  $P$ s are  $Q$ s’ is true, since all instances of gold are also instances of metal, and as such can say that all the samples with which a speaker is acquainted that are instances of  $K_P$  are also instances of  $K_Q$  (all objects within the  $P$ -sample are also within the  $Q$ -sample). Both  $K_P$  and  $K_Q$  are kinds of the same type insofar as both are substance kinds. Hence, according to Soames’ model, we can infer that  $K_P = K_Q$ , and derive the further conclusion that *necessarily* something is  $P$  iff that something is  $Q$ . But this is obviously false. All samples of gold are samples of metal, but not all samples of metal are samples of gold. The mere fact the everything within the extension of  $P$  is within the

extension of  $Q$  does not warrant an inference to the identity of the kinds associated with the predicates. Even with a stipulation that the kinds be of the same type, where by ‘type’ we mean substance or species (the kinds *gold* and *metal* are, after all, both substance kinds), the argument is still invalid.

We can identify just what work the theory of types must do by considering precisely why the counterexample works. The answer, in this case, is fairly obvious: the counterexamples works by appealing to kinds at different levels of generality. Any case where there is a natural kind  $K_P$ , and  $K_P$  is an instance of a higher-order kind  $K_Q$ , it will follow automatically that any instance of  $K_P$  will also be an instances of  $K_Q$ . Of course, what will not automatically follow is that any instance of  $K_Q$  will also be an instance of  $K_P$ .

Soames’ notion of ‘type’ must therefore restrict the class of inferences to necessary theoretical identities, specifically excluding cases where one *kind* is a member of another (higher-order) *kind*. What, if anything, makes this restriction plausible? Well, Soames’ appeal to so-called substantial non-linguistic claims must include something like the idea that natural kinds come in a hierarchy, and that kinds of the same type are *not* kinds across different levels within the hierarchy, but only on a given level.<sup>36</sup> Hence, the gold-kind is not a kind of the same type as the metal-kind, since the former is a member of the latter, despite both kinds clearly being substance kinds.

Moreover, it appears as though this thought is supposed to be implicit in the introduction of the relevant natural kind predicates in the first place. According to the necessity argument, note, when a speaker introduces a natural kind predicate they do so with

---

<sup>36</sup> The thesis that natural kinds form a unified hierarchy is something I shall criticise on empirical/metaphysical grounds in Chapter 7. In this section, for the sake of argument, I simply accept it.

some notion of what *type* of kind they are denoting. But that just looks implausible! How are typical language users supposed to be able to judge at what level the kind they are picking out occurs upon? Moreover, even if they were able to do so, this renders at least part of Soames' account implausible, since he purports to be presenting a theory concerning the basic semantics of natural kind terms.

Consider, for instance, the following (somewhat artificial) example: a technical language user *Lu* discovers a fundamental particle, and introduces the natural kind predicate 'is a muon' successfully. Now imagine, further, that there are also protons in the environment, and *Lu* introduces the natural kind predicate 'is a proton' accordingly. Since the introduction of kind terms requires, as one of its non-linguistic claims, a stipulation about the level of generality at which the kind occurs, we are effectively being asked to suppose that *Lu* implicitly recognises that 'is a muon' applies to things that are *less* general than 'is a proton', since the latter is only a subatomic particle, whilst the former is a more fundamental elementary particle. But this is highly implausible. When *Lu* introduces 'is a muon', how could she know at what level of generality the muon-kind occurs? Indeed, even if she does, that *Lu* is required to know at what level of generality the muon-kind occurs within a hierarchy only serves to illustrate just how absurd Soames' assumption is. Soames' theory of natural kind predicate semantics purports to be an account of a basic part of our language, but turns out to rely fundamentally upon a rich and complex knowledge of science. This makes it highly implausible as an account of our basic semantics.

### **3.7 The Special Auxiliary Assumption**

Even if, despite what has been said, we decide to accept Soames' restriction, it is still not strong enough to rule out counterexamples. The earlier renate/cordate example, recall, still poses a problem. Let  $P$  be the predicate 'is a renate' and  $Q$  the predicate 'is a cordate'. We have discovered that the conditional  $\forall x[Px \rightarrow Qx]$  is true, since if anything is a renate then it is a cordate. Further, since there is no more to the extension of  $P$  than  $Q$  – there is no more to the class of renates than those that are also cordates – it follows that  $\forall x[Qx \rightarrow Px]$ , and hence that  $\forall x[Px \leftrightarrow Qx]$ . However, it would still be a mistake to equate this with identity, since the property of having a heart is surely distinct from that of having a kidney. Thus, we can imagine worlds where although  $\forall x[Px \rightarrow Qx]$  is true – where, if a creature has a kidney it has a heart – it might well be the case that there are creatures that possess hearts but no kidneys i.e. where  $\forall x[Qx \rightarrow Px]$  (or vice versa) is false, and hence  $\square \forall x[Px \leftrightarrow Qx]$  – the conclusion of Soames' necessity argument – is also false.

This counterexample is interesting since it highlights a case not ruled out by Soames' initial restriction on types. In this case, since  $K_P$  and  $K_Q$  appear to be kinds of, roughly, the same type (kinds defined by the possession of particular organs), the inference is not blocked by appeal to the notion of 'type' discussed earlier.<sup>37</sup> Soames, presumably with something like the preceding (timeworn) example in mind, observes that his original argument did 'not assume that an individual cannot be a member of two kinds of a given type' (2002: 269-270). As such, he introduces a special auxiliary assumption:

---

<sup>37</sup> Since Soames does not provide any criteria for what qualifies as a natural kind, the example is *prima facie* acceptable.

*Special auxiliary assumption (SAA):* An individual cannot be a member of two different kinds of the same type  $T$  (e.g. two species, two substances). (Soames 2002: 270)

What (SAA) is intended to do is show that for some object,  $o$ , if that object is a member of one kind, the gold-kind say, then it cannot also be a member of another kind of the same type, for instance the nickel-kind. The restriction has some initial plausibility, and its utility is obvious. The renate/cordate counterexample, for instance, is apparently ruled out by (SAA): since I am clearly a member of both the renate-kind and the cordate-kind, but no individual can be a member of two kinds (presumably *natural* kinds) of the same type, either the renate-kind and the cordate-kind are *not* kinds of the same type, or they simply are not kinds at all.

Given (SAA), says Soames (2002: 270), we can construct the following argument for the necessity of theoretical identities:

(I)  $\exists x (Px \ \& \ Qx)$  is true.

(II)  $P$  and  $Q$  are natural kind predicates of the same type  $T$  (e.g., both species predicates, both substance predicates, etc.). Moreover,  $P$  ‘designates’ the unique natural kind (of type  $T$ ) instantiated by nearly all members of its associated samples. Where  $K_P$  and  $K_Q$  are the kinds associated with (‘designated by’)  $P$  and  $Q$ , respectively, the extensions of  $P$  and  $Q$  at a world  $w$  are the sets of individuals that are members of  $K_P$  and  $K_Q$ , respectively.

(SAA) An individual cannot be a member of two different kinds of type  $T$  (e.g. two species, two substances, etc.).

Therefore,

(III)  $\forall x (Px \leftrightarrow Qx)$  is necessarily true.

Premise (I) is obviously, as Soames admits, empirically *slimmer* than the initial  $\forall x (Px \rightarrow Qx)$ , since it claims that you only require one sample,  $x$ , to which both predicates,  $P$  and  $Q$ , truly apply, to infer the necessity of  $\forall x (Px \leftrightarrow Qx)$ . Whether or not we think that (III) can actually be ‘validly drawn’ (2002: 270) as Soames claims, the argument reveals just how many metaphysical assumption Soames is making about natural kinds: firstly, he assumes that natural kinds at different levels of generality are not kinds of the same type, secondly he assumes that natural kinds occur in an ordered hierarchy, and thirdly that no individual can be a member of two different kinds of the same type.

It is only when all of these assumptions are combined with Soames semantic story that finding out that anything ‘in the extension of  $[P]$  is in the extension of  $[Q]$  is tantamount to finding out that the kind associated with  $[P]$  is identical with the kind associated with  $[Q]$ ’ (Soames 2002: 269). In other words, in order to derive the identity of the kinds from the semantics of the corresponding predicates Soames assumes enough metaphysics to guarantee that it follows from either the conditional  $\forall x (Px \rightarrow Qx)$  or merely the conjunction  $\exists x (Px \& Qx)$  that the kinds associated with  $P$  and  $Q$  are identical. Indeed, the addition of (SAA) to the latter argument seems only to highlight a suppressed premise in the original argument form – in order to rule out non-identical but co-extensive predicates from being employed to show

the invalidity of the argument, it must be the case that no individual can be a member of two kinds of the same type. But this notion of type is patently false.

In the first instance it looks perfectly reasonable to treat both the predicate ‘is metal’ and the predicate ‘is gold’ as kind predicates of the same type, namely the substance-kind type. So, let  $P$  be the predicate ‘is gold’, let  $Q$  be the predicate ‘is metal’ and let  $R$  be the predicate ‘is a substance’. For some object,  $x$ , if  $x$  is gold then  $x$  is metal. More formally, this means that it is true that  $\forall x (Px \rightarrow Qx)$ . Similarly, if  $x$  is metal then  $x$  is a substance, and thus  $\forall x (Qx \rightarrow Rx)$ . Given transitivity, it is also true that  $\forall x (Px \rightarrow Rx)$ . Hence, it is reasonable to treat both  $P$  and  $Q$  as predicates of the same kind in at least one important sense – again, they are both substance predicates. If that is right then one way to view Soames’ inflated notion of ‘type’ is as an *ad hoc* attempt to rule out false inferences to necessary theoretical identities from mere conditionals and conjunctions, given his *Necessity Argument*. Since a more natural notion of ‘type’ recognises  $P$  and  $Q$  as kinds of the same type, but does not construe the discovery that either everything within the extension of  $P$  is also within the extension of  $Q$ , or that one object is both a  $P$  and a  $Q$ , as tantamount to a discovery of identity, which is surely too strong, the derivation of necessary theoretical identities, or so it seems to me, is predicated on both a notion of ‘type’ that is false, and a highly suspicious use of the material conditional (or conjunction). Standard notions of ‘type’ do *not* have the restrictions that Soames places on his notion. Moreover, even co-extensive predicates are not normally considered to be indicative of kind identity.

Similarly, (SAA) also appears to be false. Consider the following example from Muhammad Ali Khalidi (1998: 33 – 50): take the category *parasite*, of which fleas and tapeworms are both members, and the category *insect* of which fleas and flies are both

members. Since the categories cross-cut, neither *parasite* nor *insect* is obviously a subset of the other, and hence we do not breach the first condition placed upon the notion of a ‘type’. The point is more obvious if we represent the example formally as follows: let  $F$  be is a flea, let  $G$  be is a tapeworm, let  $H$  be is a fly, let  $I$  be is an insect and  $J$  is a parasite. Now, both the following conditionals are true:

$$(a) \forall x ((Fx \vee Gx) \rightarrow Jx)$$

$$(b) \forall x ((Fx \vee Hx) \rightarrow Ix)$$

Since both fleas and flies are insects, and fleas and tapeworms are parasites, but tapeworms are not insects and flies are not parasites, the categories *insect* and *parasite* cannot be subsumed, one within the other, since they imperfectly overlap. That one category cannot be subsumed within the other means that the two categories cannot be ordered into a hierarchy.<sup>38</sup> If the categories *insect* and *parasite* cannot be ordered into a hierarchy, reasoning about their members cannot be subject to the hierarchy restriction introduced via Soames’ notion of ‘type’.

Hence fleas, as members of both the insect and parasite category, operate as a counterexample to (SAA), the claim that something cannot be a member of two kinds of the same type. Since a flea is both an insect and a parasite, and the categories cannot be ordered into a hierarchy, it turns out to be false that something cannot be a member of two kinds of the same type. The *Necessity Argument*, then, is unsuccessful.

---

<sup>38</sup> Khalidi, note, uses this as an *objection* to the so-called ‘hierarchy’ thesis, that all natural kinds are ordered into a single hierarchy.



### 3.8 False theoretical identities

Although (SAA) looks to be false it is worth noting that if you construe it as true (as Soames does), then in conjunction with a further, uncontroversial assumption of taxonomic naming, many of Soames' paradigm examples, by his own lights, turn out to be straightforwardly false, and hence not illustrative of his overall thesis (that there are necessarily true theoretical identities).

According to (SAA), no individual *i* (or object *o*) can be a member of more than one kind of the same type. So, for instance, if an individual, Billy, is a member of the species *Bear* then Billy can be a member of no other species – he cannot be a member of the species *Tiger*, say – since membership of one species precludes membership of another. Now, the purpose of introducing natural kind terms, broadly speaking, is to classify individuals into kinds. Let us assume that in some cases, when a natural kind term is introduced, it is accompanied by the intention that it (the natural kind term) denotes a new kind – that is, one which was previously unnamed.<sup>39</sup>

Typically, once a kind has acquired a name (at least within a modern system of classification), unless there turns out to be some radical error about the existence of the kind, as in cases like 'phlogiston', then it has that name for good. If, for some reason or other, a kind term is *replaced* then presumably the new kind term is introduced with that intention, and is accompanied by some further rationale. For instance, if the names of chemical kinds are required to communicate information about the structure and composition of that kind, but

---

<sup>39</sup> The role of *intention* here is not such a large departure from Soames general project. Even on Soames metasemantic story, speakers are required to have some sort of intention when they ostend samples that they take to be indicative of some natural kind, namely that their proposed natural kind predicate refers to the kind of which those ostended are instances.

the kind already possess a pre-scientific name, the latter might replace with the former. Take the classic example of water, where the term ‘water’ is taken to denote the kind with the chemical composition  $H_2O$ . Using the specific rules for generating names for chemical kinds,  $H_2O$  comes out as ‘dihydrogen oxide’. Since the name ‘dihydrogen oxide’ communicates information about the structure and composition of the kind, inorganic chemists might opt to use it instead of water, given the additional information it contains. However, I shall discuss these sorts of cases further in Chapter 5. For now, I wish merely to propose that in some cases the naming of natural kinds comes with the intention that the natural kind being named is a newly discovered kind, and has not been named before.

If we accept that such is possible, and combine this with Soames’ model for generating necessary theoretical identities that contains (SAA) then there is an alternative interpretation of paradigm simple-simple theoretical identities (which were used to motivate the simple-complex theoretical identities). Take LaPorte’s ‘Apatosaurus is Brontosaurus’ example: as LaPorte tells it, particular fossils, *a*, *b*, and *c* say, were found, and a genus term (nothing hangs on these being genus, rather than species, terms) ‘Apatosaurus’ was introduced to pick out the genus of the type of which the (ostensively identified) fossilised remains were an example. Later, in a different location, a similar but much smaller skeleton was found, call it *d*, and *mistakenly* a new genus term ‘Brontosaurus’ was introduced to denote the genus of which *this* fossilised skeleton was an example. When *d* was found *d*, unbeknownst to the person who found it (in both cases this was O.C. Marsh, according to LaPorte), was already a member of the Apatosaurus genus. Marsh’s mistake was to think that he had discovered a member (or the remains of a member) of a previously undiscovered genus, and thus to introduce a new genus-name. While the discovery of *d* was obviously considered reasonable evidence of a hitherto undiscovered genus, and led to the introduction

of a new genus term ‘Brontosaurus’, *d* was, in actual fact, already a member of the Apatosaurus-genus of which *a*, *b* and *c* were members.

If we accept (SAA) then *d* cannot be a member of any other genus, and *a fortiori* cannot be a member of the Brontosaurus-genus, since *d* is already a member of the Apatosaurus-genus. If we add this to the second credible assumption that when some natural kind terms are introduced they are intended to denote *new* kinds (and where they do not denote a new kind the speaker would not intend their usage to be accepted), and note that *d* was the paradigm sample ostended by O. C. Marsh to introduce the new term (for a new kind), we can reasonably conclude that, in actual fact, ‘Brontosaurus’ was never successfully introduced. If *d* cannot be a member of any other genus other than the Apatosaurus genus then it is implausible to suggest that *d* could be successfully employed as the prime (and only!) exemplar of some new genus. Moreover, a common sense account of Marsh’s rationale in introducing the genus term ‘Brontosaurus’ is that he erroneously thought he had discovered a new genus. Had Marsh know that he was *not* naming something new with his term ‘Brontosaurus’ then presumably he would *not* have attempted to introduce a new natural kind term.

The standard treatment, found in both LaPorte and Soames, of this sort of scenario is to formulate a theoretical identity sentence, ‘Brontosaurus = Apatosaurus’, or the predicate analogue, treat it as straightforwardly true, and claim that it is also necessary. For LaPorte, the necessity follows from the fact ‘[that] “*Brontosaurus = Apatosaurus*” is true, [that] it must be necessarily true, assuming that “*Brontosaurus*” and “*Apatosaurus*” are rigid designators, as they seem to be’ (2004: 38); similarly, Soames claims that theoretical identity sentences of this form (his example is ‘all and only woodchucks are groundhogs’) are ‘linguistically determined to be necessary if true’ (2002: 278).

The argument for the necessity of ‘Brontosaurus is Apatosaurus’, on Soames’ account, should run something like this: ‘is a Brontosaurus’ was introduced in much the same way as ‘is an Apatosaurus’, via ostensive definition to a sample with which a speaker (O.C. Marsh) was acquainted. Furthermore ‘is a Brontosaurus’, just like ‘is an Apatosaurus’, successfully designated a unique natural kind of a given type. However, it turns out that the kind designated by ‘is a Brontosaurus’ was the Apatosaurus-kind. Hence everything within the extension of ‘is a Brontosaurus’ is also within the extension of ‘is an Apatosaurus’ (or, more minimally, the sample *d* used to introduce ‘is a Brontosaurus’ is within the extension of ‘is an Apatosaurus’ as well). Via the necessity argument, it follows that necessarily, for all *x*, *x* is an Apatosaurus if and only if *x* is a Brontosaurus.

But this is not the only available treatment of this (type of) example. The alternative, derived from the argument above, is that theoretical identification sentences like ‘Apatosaurus is Brontosaurus’ are straightforwardly false, since at least one of the natural kind terms (or predicates) is, in fact, *not* a term for a unique natural kind, but was introduced by mistake. The argument for such a conclusion runs as follows: assume that some taxonomic names are introduced with an implicit presupposition that they are denoting a new kind. Let natural kind terms be introduced via a broadly Kripkean metasemantic story about name acquiring transactions (ostension, etc.), and let us also endorse Soames’ special auxiliary assumption (SAA), that nothing can be a member of two kinds of the same type. Now, since *d* is a member of the Apatosaurus kind it could not have been a Brontosaurus, given (SAA). Since *d* could not have been a Brontosaurus, then despite the role *d* played in the introduction of the term ‘Brontosaurus’ (it was *the* paradigm sample), ‘Brontosaurus’ does not qualify as a natural kind term since it was not successfully introduced: thus (a) the natural kind term ‘Brontosaurus’ is an empty kind term since it did not pick out a unique kind – the so-called

Brontosaurus kind would not be the only kind of its type since ‘Apatosaurus’ had already been used to pick out that kind – and (b) given the semantic presupposition that new kind terms are introduced to denote new kinds, and there was no new Brontosaurus kind, only an existing Apatosaurus kind, ‘Brontosaurus’ does not qualify as a *bona fide* natural kind term.

If picking out a *unique* kind is one of the conditions on successful naming for at least some natural kinds, and we read ‘unique’ as meaning the only one of its kind, then trying to pick out a *new* kind using something that is a member of an existing kind appears almost self stultifying. The genus term ‘is a Brontosaurus’ failed to pick out a unique kind (the kind represented by *d*, and not by *a*, *b* or *c*), since it picked out an existing kind, the Apatosaurus genus (which includes *a*, *b*, *c* and *d* among its members). The predicate ‘is a Brontosaurus’, then, is an empty kind term, since it fails to denote anything. Consequently, given the theory of semantic content indicative of (*EM*) – that the semantic content of a natural kind predicate is just its referent – it turns out that ‘is a Brontosaurus’ is empty, much like ‘is a Pegasus’ or ‘is a Griffin’. This, in turn, invites various problems concerning the substitution of empty names, although these are beyond the remit of this thesis.<sup>40</sup>

The point is that there is a fairly natural treatment of typical theoretical identity sentences that uses Soames’ model for the derivation of necessary theoretical identities, in conjunction with reasonable assumptions about the introduction of taxonomic names, where such identities are just false. If that is correct then an important class of Soames’ paradigm cases of necessarily true theoretical identification sentences drop out of the picture, leaving us without a satisfactory account of necessary *a posteriori* truths concerning natural kind terms.

---

<sup>40</sup> Briefly, if we construe ‘Apatosaurus is Brontosaurus’ as true, and ‘Brontosaurus’ has no semantic content (being, as it is, an empty name) then ‘Apatosaurus is Pegasus’ is also true, since both ‘Pegasus’ and ‘Brontosaurus’ have the same meaning (which is to say, *no* meaning).

### 3.9 Complex natural kind terms and essentialism

We have, so far, only discussed the simple-simple type of theoretical identity i.e. those consisting of two simple natural kind predicates (or their singular term counterparts). The extension of the Kripkean insight concerning (metaphysical) necessity and *a posteriori* was the most likely in such cases given their obvious resemblance to standard identity sentences: ‘Apatosaurus is Brontosaurus’ clearly resembles ‘Ehrich Weiss is Harry Houdini’ insofar as there are two name-like terms flanking the identity predicate. However, as we have seen, Soames’ account of the necessity of theoretical identities has not, so far, been successful – or at least, if it is successful then it is predicated on a highly counterintuitive notion of ‘type’ and a decidedly odd use of the material conditional.

But, as I noted at the beginning of chapter 3, section 5, there are *two* types of theoretical identity: simple-simple and simple-complex. Now, then, we turn to the simple-complex theoretical identity sentences, and consider Soames’ separate argument for their necessity. This second brand of theoretical identity is, I claim, the most interesting; however, it is also the most controversial since it is, or so I shall conclude, predicated on metaphysical essentialism.

The simple-complex theoretical identity sentence consists, semantically, of a simple natural kind predicate, for instance ‘is gold’ or ‘is water’, and a second complex natural kind predicate, for example ‘is the element with atomic number 79’ or ‘is H<sub>2</sub>O’. This second species of theoretical identity has, as I observed in Chapter 1 (and as we shall see further in Chapter 4), much more in common with Kripke’s claims about the necessity of origin for individuals in that both ‘aim to expose essence’ (LaPorte 2002: 37). Hence simple-simple

theoretical identities and simple-complex theoretical identities are both semantically dissimilar and metaphysically dissimilar: they are semantically dissimilar since the latter (but not the former) contain semantically complex phrases that appear to be descriptions; they are metaphysically dissimilar insofar as the latter (but not the former) appear to be describing essences, and thus come with significant metaphysical commitments.<sup>41</sup>

As Soames carves up the semantic geography, a simple-complex theoretical identity consists of a simple natural kind predicate  $P$ , for example ‘is water’, and a complex natural kind predicate  $Q$ , the chemical formula (modified with the copula ‘is’) ‘is  $H_2O$ ’, which Soames claims is synonymous with the description ‘substance molecules of which contain two hydrogen atoms and one oxygen atom’ (2002: 274). This description, as Soames observes, is the product of scientific theory, and belongs to ‘a class of formulas specifying different possible molecular structures that are understood as determining possible physically constitutive kinds’ (*ibid.*). The empirically discovery that water is  $H_2O$ , then, is the discovery of the basic physical constitution of the water-kind. The question now is what implications the complexity of  $Q$  has on its semantics. More specifically, since the semantically complex phrases that specify the alleged essences of natural kinds, for instance the molecular structures of chemical compounds, are being treated as descriptions, are they Millian (i.e. nondescriptive marks, with a semantic content that just is the referent)? If  $Q$  does fall under the purview of ( $EM$ ) – the thesis that the meaning of a kind term is just the kind it designates – then ‘ $x$  is water iff  $x$  is  $H_2O$ ’ (or, ‘water is  $H_2O$ ’) will express the same proposition as ‘ $x$  is water iff  $x$  is water’, since ‘water’ is Millian, and both terms designate the same kind. But, if the latter proposition is knowable *a priori* then so too is the former, and the Kripkean insight

---

<sup>41</sup> Although as we have seen, Soames’ simple-simple theoretical identities have significant metaphysical baggage.

concerning *a posteriority* is lost. If *Q* does not fall under the purview of (*EM*) since it has a meaning other than the kind to which it refers (it has a connotation in addition to its denotation) then we might be sceptical as to just how Kripkean Soames' semantics really is. Moreover, we should wonder why *some* descriptions (the essentialist ones) apparently maintain the necessity (and *a posteriority*) of theoretical identities whilst others (the non-essentialist ones) do not.

The first question, concerning whether or not descriptive predicates like 'is H<sub>2</sub>O' are Millian, turns on the question of what natural kinds *are*, and is motivated by concerns over the *a posteriority* of theoretical identities. If natural kinds are property-like entities, and the predicate 'is H<sub>2</sub>O' describes the essence of that kind, then since 'is water' and 'is H<sub>2</sub>O' designate the same kind, and the meaning of such a kind term is just the kind it designates, their respective meanings must be identical. More simply, if natural kind predicates designate properties, and the 'property of being water is the property of being a substance molecules of which contain two hydrogen atoms and one oxygen atom' (Soames 2002: 276) then the natural kind predicates 'is water' and 'is H<sub>2</sub>O' mean the same thing, since both mean 'the property of being a substance, molecules of which contain two hydrogen and one oxygen atoms'. The upshot is that we lose Kripke's 'conventional wisdom about the *a posteriority*' (Soames *forthcoming*: 19) of necessarily true theoretical identity sentences.

Soames' solution is to concede that 'the kind (...) water is identical with the kind (...) substance molecules of which contain two hydrogen atoms and one oxygen atom' (2002: 277), but to claim that 'the two predicates do not express the same properties' (*ibid.*). Simple natural kind predicates, such as 'is gold', 'is water' or 'is a muon', are like proper names insofar as they are nondescriptive, directly referential and have their reference fixed by stipulation (*modulo* certain empirical presuppositions). So, just as 'the referent of a name at a



world is not semantically determined by the satisfaction of any descriptive condition at the world' (Soames 2002: 265), neither is the referent of a simple natural kind predicate.

Complex natural kind predicates, for instance 'is the element with atomic number 79', 'H<sub>2</sub>O' or 'is an unstable subatomic particle with a negative charge, a mass of 105.7MeV/c<sup>2</sup> and spin ½', are semantically quite different. Let  $Q$  be a typical complex natural kind predicate. According to Soames,  $Q$  can be analysed as the complex  $\langle I, D \rangle$ , where  $I$  is the copula meaning roughly 'the relation of being an instance of', and  $D$  is the descriptive semantic content of  $Q$ . In the case of 'is H<sub>2</sub>O' then, the descriptive content,  $D$ , is 'substance, molecules of which contain two hydrogen atoms and one oxygen atom'. The descriptive content of  $Q$ ,  $D$ , denotes the property  $F$ . In the water example that property is, again roughly, the property of *being a substance, molecules of which contain two hydrogen and one oxygen atom*. Since  $D$  always designates this property, in all possible worlds,  $Q$  turns out to be rigid. More important, however, is Soames' claim that the property  $F$ , described by  $D$  and indirectly denoted by  $Q$ , is the property that determines the kind  $K_Q$  in every possible world, but is not itself identical with it. Thus not only is the predicate  $Q$  – which is a typical complex natural kind predicate – *not* a Millian predicate, since its reference to the natural kind is mediated by description, the natural kind  $K_Q$  is *not* the property of *being a substance, molecules of which contain two hydrogen and one oxygen atom*.

Since complex natural kind predicates 'have meanings that are not identical with the objects they designate, but rather are properties in virtue of which the descriptions designate what they do' (Soames 2002: 279), the proposition expressed by 'for all  $x$ ,  $x$  is  $P$  iff  $x$  is  $Q$ ' is *not* identical to that expressed by 'for all  $x$ ,  $x$  is  $P$  iff  $x$  is  $P$ ', and thus it fails to follow that the former is *a priori* just because the latter is. Typical simple natural kind predicates, like 'is water' designate natural kinds directly, although we have yet to say precisely what they

(natural kinds) are, whilst complex natural kind predicates, like ‘is H<sub>2</sub>O’, are ‘analogous to singular definite descriptions’ (Soames 2002: 279) whose reference is fixed by discovery (rather than stipulation), and whose use *is* satisfied via an object’s meeting of certain descriptive conditions i.e.  $Q$  only applies to things that satisfy the description  $D$ , the only things that satisfy the description  $D$  are those things that possess the property  $F$ , and  $F$  determines the kind  $K_Q$ . Now Soames must show that the kind denoted directly by  $P$ , call it  $K_P$  is identical to that which is *determined* by  $F$ , which we have called  $K_Q$ . But what sort of thing does  $P$  denote?

According to Soames natural kind predicates denote natural kinds, where ‘we identify them [natural kinds] with intensions – that is functions from [possible] worlds to extensions’ (2002: 277). The thesis, then, is broadly as follows. Take, for example, the nickel-kind. The nickel-kind,  $K_N$ , is to be identified with an intension  $I_N$  (a function from worlds to extensions), and the simple natural kind predicate ‘is nickel’, or  $P$ , is directly referential, nondescriptorial and denotes  $I_N$  across all possible worlds. The complex natural kind predicate ‘is the element with atomic number 28’, abbreviated to ‘is Ni28’ and represented formally as  $Q$ , is a definite description (the descriptive content of which we have called  $D$ ) that refers to anything with the property of *being an element with atomic number 28*. This property, call it  $F$  (which is described by  $D$ ) is what *determines* the intension  $I_N$ .

The interesting (and controversial) element of Soames’ thesis is the determining relation that holds between the property  $F$  and the natural kind, which is an intension,  $I_N$ . An intension, as I have said, is a function from worlds to extensions. Roughly understood it is a function that assigns, at some possible world, a value, either positive or negative, to objects which either (i) are instances of a natural kind, or (ii) are truly referred to by the relevant natural kind predicate. For instance, the natural kind predicate ‘is nickel’ refers to the function

$I_N$  that assigns, at any possible world, a value of 1 (positive) to all samples of nickel, and a value of 0 (negative) to all other objects. The complex natural kind predicate ‘is the element with atomic number 28’, more importantly, describes the property *being an element with atomic number 28*, or  $F$ , and it is this property which determines the intension  $I_N$  that we have identified with the nickel-kind. The vital question now is why is it  $F$  that determines  $I_N$ , rather than some other property, not just in the actual world, but in all possible worlds?

Soames, in his (2002) at least, gives no explanation, but it is easy enough to reconstruct one. Firstly, we need a *reason* to think that  $F$  is *the* property that determines the nickel kind, rather than some other property,  $G$  say. Since Soames has already been explicit about the source of complex natural kind descriptions – that they are the product of scientific theory – it is to science that he would presumably turn for some sort of justification as to why  $F$ , but not  $G$ , determines the nickel-kind. Presumably then Soames would say something like ‘since science classifies substances like nickel in terms of their atomic number, indeed they define elements on that basis, we philosophers should construe that property as *determining the nickel-kind*’. Indeed, in later work (2008) this is roughly what he does say. But note that for Soames, and arguably *not* for the scientist, the upshot of the empirical discovery that nickel is the element with atomic number 28 is that *necessarily* nickel is the element with atomic number 28. But the claim that science is in the business of discovering necessary truths is controversial to say the least (see, for instance, Dupré 1986, Mellor 1977, Putnam 1983; see Leplin 1988 for a response).

In order for it to be the case that *necessarily* nickel is the element with atomic number 28 it must be the case that the property *being the element with atomic number 28* is *essential* to nickel. Put in terms of intensions, it must be the case that the property *being the element*

*with atomic number 28* is the only property that determines the intension  $I_N$  which is the nickel-kind. But why think that?

For Soames the key is in the metasemantics of simple natural kind predicates, specifically the empirical presuppositions that accompany their (ostensive) introduction. When simple natural kind predicates are introduced they are introduced with the *intention* that they apply to anything that possesses the same properties that explain the salient characteristics of paradigmatic instances of the kind. The empirical discovery of what this property (or set of properties) is, is tantamount, says Soames, to learning ‘empirically that instances of the kinds have these characteristics at every possible world-state in which the kind exists’ (*forthcoming*: 8). It seems, then, that built into the initial enterprise is an assumption that certain ‘properties are essential to the kinds’ (*ibid.* 7). Moreover, the assumption seems to be that when simple natural kind terms are successfully introduced they latch on to these essences, by being nondescriptive (which means that we don’t need to *know* what the essence is when the term is introduced), and that the stability of essences across possible worlds stabilises the reference of the natural kind term. Hence, when science eventually discovers what underlying property it is that explains the salient features of the kind, and describes that property, this complex natural kind term also manages to denote the unique kind, and can be formulated into a theoretical identity sentence. As Soames says, it is ‘a feature of any genuine substance S that whatever its molecular structure turns out to be, all possible instances of S share that structure (and all possible instances of that structure are instances of S)’ (2002: 273).

Thus it is from the conjunction of a belief about substances (that they have essences), the truth of a theoretical identity sentence – such as ‘nickel is the element with atomic number 28’ – and the metasemantics of simple natural kind terms like ‘nickel’, that we derive the

necessity of the relevant theoretical identity. Such theoretical identities are both (metaphysically) necessary and *a posteriori*, and the ‘necessary aposteriority of these examples is thereby grounded in essentialist claims about kinds which can be known only empirically’ (Soames *forthcoming*: 5).

The problem, perhaps obviously, with this account is that it is predicated upon the truth of essentialism – an independent metaphysical thesis which receives no independent justification in Soames’ theory. On that basis, not only does the proposed extension of Kripke’s insight rest upon an unjustified, highly controversial assumption (of metaphysical essentialism), but it is incompatible with anti-essentialism. Moreover, like Kripke’s original account, it simply assumes that reference is unproblematic, and this, I think, is less obvious when we are talking about natural kind terms (more on this in chapter 5). As we shall see in the next chapter, however, an account of our basic semantics (the causal theory of reference) should not be incompatible with any particular metaphysical thesis. Further, as we shall see in chapter 5, if some versions of Millianism are committed to essentialism and countenance at least *some* rigid descriptions, then there is an obvious version of descriptivism that is equally as tenable. Underlying both such theses, however, will be some version of *a posteriori* essentialism, and the notion of an explanatory essence that accounts for the salient features of natural kinds. It is to this that I turn in chapters 6, 7 and 8.

As a final thought, it is worth noting that Soames’ account appears not to satisfy one of its own key desiderata. In section 3 of this chapter, I identified Soames’ three key desiderata for an extension of Kripke’s insight into proper names to natural kind predicates. Condition (2), recall, was that it must be a consequence of the account that there is an asymmetry between natural kind predicates and all other predicates insofar as only the former are rigid. However, it cannot be the case that what determines whether something is a natural

kind predicate is whether or not it is rigid, since this would make (2) completely trivial. For instance, it seems straightforwardly false to suppose that determining whether or not the general term 'philosopher' picks out a natural kind hinges upon whether 'is a philosopher' is a rigid predicate. Hence, given that Soames assumes that there is an asymmetry between natural kind predicates and all other predicates, he must be assuming that somehow the rigid predicates will pick out that natural kinds. But then it seems he should say much more about what the natural kinds are. Presumably what Soames has in mind is that the natural kinds are just those kinds marked out in our best scientific classification. However, not only is this an unargued assumption, but it seems totally implausible to suppose that rigid predicates, miraculously, are just those predicates that latch onto the best scientific classification. The kind term 'phlogiston', at some point, was part of the best scientific classification, but failed to latch onto a legitimate joint in nature. Hence there is no reason to think that everyday naming practices latch onto nature's joints, as revealed by scientific investigation. To put it another way, there are no *a priori* reasons to think that general terms (be they predicates or otherwise) that are rigid designators should correspond exactly to those employed by modern science (physics, chemistry and biology, say). Given these problems I now turn to an alternative attempt to extend the Kripkean thesis from proper names to natural kind terms.

## CHAPTER 4

In this chapter I will present Nathan Salmon's alternative version of the Causal Theory of Direct Reference, (*CDR*) for short, and argue that although it avoids some of the problems that beset Soames' version, it is unsuccessful insofar as it ignores the function of natural kind terms within classification and taxonomy. I begin, in section 1, by articulating the debate between Salmon and Soames over the status of rigidity in the project of extending Kripke's conclusions from names to natural kind terms. In section 2 I present Salmon's version of (*CDR*), and in section 3 discuss the Kripkean objections to descriptivist natural kind terms, along with Salmon's important caveats. In section 4, following Salmon, I distinguish between the two distinct routes to necessary *a posteriori* truth in Kripke. In sections 5 I identify the essentialism implicit in Kripke, and in section 6 present Salmon's distinction between Kripke and Putnam in terms of their respective essentialist commitments. In section 7 I present Salmon's account of the semantics of theoretical terms, such as 'H<sub>2</sub>O'. In sections 8 and 9 I criticise Salmon's positive proposal. I claim that his account of the semantics of natural kind terms is implausible on two fronts: firstly, concerning the function of natural kind terms within science, and secondly, given the so-called 'trivialisation problem' (Lopez De Sa 2007). In section 10 I pause to briefly take stock of how the debate has progressed, before moving on, in chapter 5, to defend a novel version of descriptivism. I will also, here, present Salmon's famous objections to the Kripke-Putnam (i.e. early-Putnam) programme to derive metaphysical conclusions from semantic premises, offer some qualifications to Salmon's objections, and illustrate how they also apply to Soames' more recent account. I will conclude, however, that the most plausible version of (*CDR*) for natural kind terms contains a

*prima facie* essentialist premise. This will set up the defence of a novel version of descriptivism in chapter 5, using the implicit essentialism of (*CDR*).

It is worth noting from the outset of this chapter that the discussion concerning natural kind terms is no longer being framed in terms of predicates. For Salmon, the fundamental use of natural kind terms is as names for natural kinds. I shall indicate a naming use of a natural kind term throughout this chapter by capitalising the first letter, as one would do with a proper name.

#### **4.1 Salmon vs. Soames**

Nathan Salmon's explication of (*CDR*) differs sharply from Soames', as presented in the previous chapter. Indeed, Salmon objects to Soames' entire project, claiming that 'Soames creates the misimpression that *N&N* [*Naming and Necessity*] stands somehow incomplete and so falls short of meeting its own objectives' (2003: 447).

Among the various points of disagreement between Salmon and Soames there are two that are pertinent to my project concerning the metaphysical necessity and *a posteriori* knowability of theoretical identities: firstly, the issue of whether general terms (primarily mass and count nouns) are fundamentally predicative or not; and secondly, the alleged 'demotion of the status of rigidity' (Soames 2002: 264) as the fundamental semantic property, which is responsible for maintaining the analogy between proper names and natural kind terms. Soames' view, recall, is that the predicative use of mass and count nouns is fundamental, rather than the denoting use (so 'is a bear' is more fundamental than 'Bear'). Further, he argued for the demotion of rigidity, claiming that the analogy between proper



names and natural kind terms (and the extension of Kripke's insights from the former category to the latter category) is grounded in an alternative semantic similarity, namely their mutual nondescriptiveness.

Salmon, in contrast, claims that Soames has failed to 'distinguish sharply between a general term [like] "tiger" and its corresponding predicate, "is a tiger"' (2005: 386), and as such has falsely demoted, and failed to appreciate the philosophical significance of, rigidity. Salmon's basic claim is that it is the rigidity of count and mass nouns that makes them parallel to proper names, and that it is this parallel that allows us to extend Kripke's insights concerning proper names to natural kind terms. In this chapter I will be particularly interested in Salmon's analysis of this parallel, and in his distinctive treatment of so-called theoretical identification sentences.

#### **4.2 Salmon's (CDR)**

Salmon claims that there is an 'imposing analogy' (2005: 43) between proper names and common nouns, in that both sorts of terms are *designators*: proper names designate individuals, common nouns designate kinds. The account runs as follows: Consider, first, two distinct uses of a general term, for instance 'Billy is a bear' and 'Billy is a member of the species Bear'. The difference between the two sentences is that whilst the former contains but one denoting phrase, the latter contains two. In the former sentence, we take it that an object is the referent of the proper name 'Billy', and that we predicate of that object, Billy, that he 'is a bear'; hence, while we are ontologically committed to Billy when we assert that 'Billy is a bear', the occurrence of 'bear' in that sentence carries no explicit ontological commitment, being, as it is, merely a property that we are attributing to Billy.

In the latter sentence, however, both the proper name ‘Billy’ and the general term ‘Bear’ are most naturally read as denoting expressions. Here we say of Billy that he is a member of a particular species, which we have designated ‘Bear’, and take the general term ‘Bear’ to ‘denote a certain *kind* of animal, a certain abstract entity or *universal*’ (Salmon 2005: 43). Since the second sentence contains two denoting expressions we garner (when we assert it) two obvious ontological commitments, one to Billy, and one to the species *Bear*.

Salmon, following Kripke, claims that single world general terms like ‘Bear’ are ‘nondescriptonal label[s]’ (2005: 45) that designate natural kinds – in other words, they are *names* for natural kinds.<sup>42</sup> But, if we treat general terms, such as ‘Bear’, as nondescriptonal labels (making them analogous to proper names), we need to explain why it is that they can be truly applied to a range of individuals. For instance, imagine Billy the bear has a family, Betty, Belinda and Ben, all of whom are members of the species *Bear*. If ‘Bear’ is a nondescriptonal label then it cannot be the case that the word ‘Bear’ gets to apply to these individuals in virtue of them satisfying some description, which is encoded within, and is the content of, the term ‘Bear’. Thus it must be that the individuals themselves that fall within the extension of the term have something in common, which explains why the term ‘Bear’ does (truly) apply to any and all bears.

There is also an important, but well known, difference between general terms and proper names, when we distinguish between the extension (the object(s) to which a term

---

<sup>42</sup> I will, to begin with, simply restrict the focus of the discussion to general terms that designate natural kinds, since this is the topic of my overall investigation. However, as we shall see, Salmon’s account generalises across a range on non-natural kind cases, including, for instance, general terms for artefact kinds. This will be particularly important when we come to the metaphysics sections of this thesis, where articulations of natural kind essentialism typically appeal to Kripkean insights. If rigidity applies to all general terms, rather than just natural kind terms, then there is no special Kripkean insight into natural kind terms. Hence it cannot be employed to motivate metaphysical essentialism.

applies) and the denotation (the object to which a term refers) of a term. The distinction is best illustrated by example. Take the proper names ‘Bruce Wayne’ and ‘Batman’. Since Bruce Wayne is Batman, the two names are co-referential. When two proper names are co-referential they have the same extension – that is, they apply to the same person. Moreover, when two proper names have the same extension, this is sufficient to infer a corresponding identity sentence, in this case ‘Bruce Wayne is Batman’.

However, when two general terms have the same extension this is insufficient for theoretical identity. For the sake of argument, treat natural kinds as abstract universals. Following Salmon, we can say that a natural kind term like ‘Bear’ denotes or *designates* the natural kind or species *Bear*, i.e. the *Bear*-universal. The natural kind *Bear* also has what Salmon calls a *metaphysical extension*, which is ‘the class of instances of an abstract universal’ (2005: 46). Since the natural kind term ‘Bear’ designates the kind *Bear*, and the metaphysical extension of the *Bear*-kind is the class of instances of bears, the *semantic extension* of the term ‘Bear’ is also the class of instances of bears. Strictly speaking, then, no particular bear is the denotation of the term ‘Bear’, although it is, of course, clearly possible to refer to an individual bear using that term.<sup>43</sup> Hence we can produce the following definitions (Salmon 2005: 46):

*Metaphysical extension* =<sub>df</sub> the class of instances of an abstract universal.

---

<sup>43</sup> Perhaps obviously, the extension of a Millian thesis about proper names (the claim that the meaning of a term is its referent) extended to natural kind terms does require some care. For instance, it must not be the case that the actual *extension* of a general term is the *referent*, since extensions are fairly dynamic sorts of things. So, if the referent of ‘human’ was the actual extension of the kind *human*, then since the population of humans is constantly changing, so too would the meaning of the general term ‘human’, and this is surely wrong.

*Semantic extension* =<sub>df</sub> the class of individuals to which a term (truly) applies.

But there are obvious cases of co-extensional general terms that are insufficient for an inference to identity. Adapting an example from Salmon (2005: 45-8), consider the term ‘Member of a species that serves as a mascot for Chicago’s NFL team’. This general term denotes the (non-natural) kind *Member of a species that serves as a mascot for Chicago’s NFL team*, and, like the term ‘Bear’, applies to individual animals on the basis of their being members of the species that serves as a mascot for Chicago’s NFL team. The extension of the phrase ‘Member of a species that serves as a mascot for Chicago’s NFL team’, in this world, is the class of bears. Hence, the kind terms ‘Bear’ and ‘Member of a species that serves as a mascot for Chicago’s NFL team’ have the same semantic extension (i.e. both terms apply to the same class of individuals). But there is an obvious difference between the denotations of the respective kind terms. While the term ‘Bear’ must apply to the *Bear*-kind in all possible worlds, it is pure coincidence (and thus contingent) that the term ‘Member of a species that serves as a mascot for Chicago’s NFL team’ also applies to the *Bear*-kind. For instance, we can clearly imagine a scenario where Chicago’s NFL team had an alternate mascot, a buffalo say, and as such the extension of the general term ‘Member of a species that serves as a mascot for Chicago’s NFL team’ is the class of buffalos, rather than the class of bears. In such a world, call it  $w_1$ , the extensions of the two general terms would be quite different since the extension of ‘Member of a species that serves as a mascot for Chicago’s NFL team’ in  $w_1$  would be the class of buffalos, rather than the class of bears as it is at the actual world.

What this example illustrates is that theoretical identities require more than just co-extensionality, else ‘A Bear is the member of a species that serves as a mascot for Chicago’s

NFL team' would be necessarily true. Thus, following Salmon (2005: 46), we can draw the following, additional semantic and metaphysical distinction, and note that theoretical identities require co-intensionality.

*Metaphysical intension* =<sub>df</sub> the function that assigns, for any abstract universal *K* and any possible world *w*, the class of instances (*m*-extension) of *K* at *w*.

*Semantic intension* =<sub>df</sub> the function that assigns, for any (general) term *T* and any possible world *w*, the class of instances (*s*-extension) of *T* at *w*.

With the addition of these pieces of terminology there is an obvious distinction between 'Bear' and 'Member of a species that serves as a mascot for Chicago's NFL team'. The general term 'Bear' designates the *Bear*-kind, or more accurately the species (or family) *Ursidae*, and hence the semantic intension of 'Bear' coincides with the metaphysical intension of the *Bear*-kind, and does so in all possible worlds. The metaphysical intension of the non-natural kind *Member of a Species that Serves as a Mascot for Chicago's NFL Team*, on the other hand, clearly has a different metaphysical intension to the *Bear*-kind, since only the former could have assigned a class of instances consisting of all and only buffalos.

As a brief aside, the forgoing discussion also reveals how, for Salmon, the notion of rigidity applies to general terms. On Salmon's view the rigidity of a general term is *not* simply a product of that term's nondescriptiveness, since general terms can be rigid regardless of 'whether the general term is nondescriptive, relationally descriptive, or thoroughly descriptive' (Salmon 2005: 70). For Salmon, provided a general term is a

designator it follows trivially that it is also a rigid designator. There are, however, two conditions on general term designation: for some general term  $T$  to count as the designator of a kind  $K$ , (i) the semantic extension of  $T$  must coincide with the metaphysical extension of  $K$ , and (ii) the semantic intension of  $T$  must also coincide with the metaphysical intension of  $K$ . The second condition places ‘severe restriction[s] on the sorts of entities that can be designated by a general term’ (*ibid.*) since it restricts the reference of general terms to universals. The Kripkean notion of rigidity, recall, is sameness of reference across possible worlds i.e. if  $a$  rigidly designates  $x$ , then  $a$  designates  $x$  in all possible worlds. Now, since on Salmon’s view general terms only qualify as designators when the semantic intension of  $T$  coincides with the metaphysical intension of  $K$ , the semantic intension of  $T$  only coincides with the metaphysical intension of  $K$  when  $T$  designates  $K$  (where  $K$  is a universal and  $T$  a general term), and when  $T$  designates  $K$ , given that  $K$  is an abstract object,  $T$  automatically does so in all possible worlds (perhaps with the qualification, where  $K$  exists), it turns out that ‘the distinction between *designation* and *rigid designation* disappears’ (Salmon 2005: 71-2).

For instance, the general term ‘Bear’ designates the species *Bear*, which is a universal, instances of which are actual bears. In contrast, general terms like ‘the colour of the sea’ are thought to be non-rigid designators, since (a) there is no *colour of the sea* kind, and (b) the colour of the sea varies from possible world to possible world (in some worlds it is green, in others blue, and so on). However, as I shall argue later, following work by López De Sa (2007), the distinction between rigid and non-rigid general terms turns out to be spurious, given the abundance of available properties to correspond to the appropriate general term. This, I will show later (in Chapter 7), has consequences for parallel views in metaphysics concerning natural kind essentialism. However, in this chapter I will focus on the consequences it has for distinguishing theoretical identities from predications. For instance,

typical language users predicate certain properties of objects regularly: ‘my shirt is blue’, ‘this saucepan is hot’ or ‘my stomach is aching’. However, if general terms like ‘shirt’, ‘blue’, ‘saucepan’, ‘hot’, ‘stomach’ and ‘aching’ all designate properties, which are treated as abstract universals, then the distinction between sentences that predicate properties of objects, and those that identify two kinds, seems to disappear. I will return to this issue in detail in section 8. For now, however, I propose the following principle for restricting what sentences count as theoretical identities, which I take to be an implication of Salmon’s thesis. The principle – call it the *Theoretical Identity Principle* – is intended to express the restriction required to make an inference from co-extension to identity, namely, that the semantic intensions of the kind terms must also coincide (by designating kinds with the same metaphysical intension).

Theoretical Identity Principle =<sub>df</sub> for any two general terms  $T_1$  and  $T_2$ , the theoretical identity  $T_1 = T_2$  can only be true when the (metaphysical) intension of the kind  $K_1$ , denoted by  $T_1$ , coincides with the (metaphysical) intension of  $K_2$ , denoted by  $T_2$  (i.e. when  $K_1 = K_2$ ).

### **4.3 Objections to descriptivism about natural kind terms**

In chapter 1 I outlined the extension of Kripke’s objections to a descriptivist semantics to cases of natural kind terms. These arguments, recall, formed a putative *reductio ad absurdum* of the descriptivist position by highlighting the absurd metaphysical, epistemological and

semantic consequences of maintaining a descriptivist semantics, which were clearly inconsistent with our (intuitive) everyday practices.

The conclusion of these arguments was a return to a version of naive Millianism, the thesis that proper names have denotation by no connotation, extended to natural kind terms. However, Salmon notes (2005: 61-5) that these (now classic) objections only ever warranted a weaker conclusion in the case of general terms. Consider, for instance, Salmon's (2005: 59 – 68) discussion of the species term 'Tiger'. A descriptivist, says Salmon, might 'associate with the term ['Tiger' the description] *being a four-legged animal, being cat-like in appearance, having a tawny yellow coat with black traversing stripes ...*' (2005: 59). From this we can formulate (1):

- (1) Something is a tiger iff it is a four-legged, meat eating, cat-like animal with a tawny yellow coat and black traversing stripes.

Let (1), for the sake of argument, be taken to give the meaning of the word 'Tiger'. Since (1) is analytic (for the descriptivist), necessarily (this is the modal argument) anything that satisfies the description 'four-legged, meat eating, cat-like animal with a tawny yellow coat and black traversing stripes' is a tiger. But, we can imagine a world where there is a species of feline that satisfies this description, but which could not interbreed with a tiger. Thus we appear to have two *distinct* species of feline<sup>44</sup>. Hence, it is (metaphysically) possible

---

<sup>44</sup> The example is predicated on the assumption that being able to interbreed is a necessary condition of being a member of the same species. However, this isn't always true. Consider, for instance, the offspring of a zebra stallion and a horse mare, a *zorse*. The *zorse*, as is typical of most hybrids, is sterile. Hence a male and female



that non-tigers satisfy the description said to define the word 'Tiger', and (1) cannot be necessary, and *a fortiori* cannot be analytic.

Now consider (1) in an alternate counterfactual scenario (the epistemological argument), where, as a result of some mass hallucination, we mistakenly attribute to tigers the properties described in (1). For instance, imagine that individual tigers have only ever been observed in dense jungle, and so the stripes we think they have are actually just shadows, and their tawny yellow colour is just a trick of the light (perhaps they are actually green). If this is possible then the word 'Tiger' cannot include, as part of its meaning, 'tawny yellow' and 'traversing stripes' since tigers are actually green and not striped. Hence (1) cannot be knowable *a priori*, and *a fortiori* cannot be analytic.

Finally, imagine a species of animal qualitatively identical to tigers, but inhabiting a planet far, far away (the semantical argument). Since we are not acquainted with these creatures, but they exactly resemble our tigers, do we refer to instances of them when we use the word 'Tiger', as defined by (1)? According to the semantical argument we do not. The general term 'Tiger' is 'a label that applies only to animals of a certain familiar kind. It does not apply to everything whatsoever that might happen to have the general features that we associate with tigers' (Salmon 2005: 61). Hence the existence of a qualitatively identical species with which we are not acquainted has no bearing on our use of the word 'Tiger'.

What Salmon notes, however, is that the upshot of the Kripkean arguments is *not* that descriptivism is false (as is perhaps the case for proper names), and that general terms are (semantically) nondescriptional *tout court*. Rather, the conclusion is that the general terms

---

zorse would be unable to breed. However, this would not rule out either of them as being members of the species zorse.

considered *so far* ('Water', 'Gold', 'Tiger' etc.) cannot be descriptive in terms of the properties considered *so far* (*ibid.*).

Salmon thus considers, on behalf of the descriptivist, the thesis that the term 'Tiger' refers to all and only those things that have the property of *being a tiger* (rather than the property of *being tawny, yellow, striped four-legged felines*). The proposal, then, is that the description 'is a tiger' is that which fixes the extension, and is the meaning, of the term 'Tiger'. The obvious attraction of the proposal is that it is obviously 'analytic that all and only tigers have this property [the property of *being a Tiger*]... and that the term "tiger" applies to all and only those things that have this property' (Salmon 2005: 61). *Prima facie*, then, the proposal appears to avoid Kripke's three objections.

Salmon's objection to the descriptivist rejoinder stems from his analysis of what the property of *being a tiger* amounts to. For Salmon, the 'property denoted by "*being a tiger*"... involves direct reference to the species *Tiger* and is necessarily equivalent, in some sense, to the intrinsically relational property of *being an animal of this very kind*, said with direct reference to the species *Tiger*' (2005: 63). In other words, the property of *being a tiger* consists of the conjunction of the species *Tiger* (well, *Felis Tigris* actually) and a 'kind-membership relation' (*ibid.*). Salmon's objection is that *even if* it is conceded that 'Tiger' is synonymous with the description 'is a tiger', the latter still involves direct reference to the species *Tiger* (note the use of 'this' in 'this very kind', which is intended to be ostensive). Hence such a concession does not affect the overall conclusions of (*CDR*).

Salmon uses Putnam's Twin Earth thought experiment to demonstrate that the property of *being  $\phi$* , where  $\phi$  is any natural kind, is an intrinsically relational property, and as such of no use to the descriptivist. Salmon treats the Twin Earth thought experiment as a type

of semantical argument (2005: 68) – which is to say, an argument that the general properties associated with a natural kind do not constitute a logically sufficient condition for the use of the corresponding natural kind term. Consider two speakers *s1* and *s2*, and imagine that *s1* is an inhabitant of Earth, and *s2* is an inhabitant of Twin Earth, and both *s1* and *s2* have identical mental biographies. Imagine, further, that *s1* and *s2* find themselves in qualitatively identical circumstances, being pursued by a large, furry, four legged animal with claws, and a rather antisocial disposition, say. Now, both speakers have the thought ‘there is a Bear chasing me!’. For *s1* ‘Bear’ functions as it standardly does on Earth, for any typical language user. For *s2* ‘Bear’ denotes a species *k* which cannot (and could not) interbreed with the species that *s1* denotes using the term ‘Bear’. Now, since *s1* and *s2* have identical mental biographies, the conceptual content of their respective *there is a Bear chasing me*-thoughts must be identical. But since the species denoted by *s1*’s use of the term ‘Bear’, *Ursidae*, is not the same as the species denoted by *s2*’s use of the term ‘Bear’, the conceptual content *cannot* fix the reference of a general term.

For the descriptivist (as noted in chapter 1) there must be some description that both is the conceptual content of a word, and plays the unique denotation-determining role. Salmon takes the Twin Earth thought experiment to demonstrate that no single description can play both of these roles, and hence that descriptivism is untenable. The crux, then, is that the descriptions a speaker might come to associate with a kind, and treat as the conceptual content of the corresponding kind term, can ‘always underdetermine the denotation or extension’ (Salmon 2005: 68) of a natural kind. The constant denotation of the kind term in the Twin Earth experiment is a product of the direct reference to the kind in question, and since ‘properties that involve direct reference to external things of a certain kind... cannot be part of

a psychological state' (*ibid.*) the thought experiment shows that the property of *being  $\varphi$*  'cannot be a general property involving no direct reference' (*ibid.*: 69).

What the descriptivist requires, then, is that the property of *being a tiger* is a 'general property that does not involve direct reference to the species *Tiger*' (Salmon 2005: 65). If they can show this to be the case, then 'the direct reference theory [is] neatly refuted, by the simple observation that "tiger" may be said to express the property of *being a tiger*' (*ibid.*). Thus descriptivism cannot analyse the property of *being a tiger* as the conjunction of the species *Tiger* and the kind-membership relation, since this is 'only a small step away from the claim that "tiger" is altogether nondescriptive' (*ibid.*: 64). Rather, it must show that there is some general property that defines what it is to *be* a tiger, but does not, itself, involve direct reference to the species *Tiger*. I will not pursue this line of argument on behalf of the descriptivist here, but reserve a proposed rejoinder for chapter 5, where I will argue that such a strategy is readily available for the descriptivist. Now I turn to Salmon's analysis of the derivation of *a posteriori* necessities concerning natural kinds.

#### **4.4 Theoretical identity and *a posteriori* necessity**

Ignoring, for the time being, any problems that might afflict Salmon's own account of the rigidity of general terms and theoretical identity, and the point raised in Chapter 3 that it is a commitment of a Millian view of general terms that theoretical identities containing co-referring terms are knowable *a priori*, not *a posteriori*, I now return to Salmon's analysis of Kripke and Putnam, and their alleged derivation of *a posteriori* necessities. Once this is complete, I turn to my own criticisms of Salmon's account.

As I demonstrated in Chapter 1, the necessity of certain identity sentences is, according to Kripke, entailed by the conjunction of their truth and the philosophical semantics of (*CDR*), in particular the notion of rigid designation. Identity sentences such as ‘Hesperus is Phosphorus’ or ‘Ehrich Weiss is Harry Houdini’ contain two proper names, which are always rigid designators, flanking the identity predicate (or that natural language version of it) and are guaranteed to be necessary when true. They also allow us to derive additional (metaphysically) necessary truths, for example it follows from the necessity of ‘Ehrich Weiss is Harry Houdini’ that, necessarily, if Ehrich Weiss is human, then Harry Houdini is a human. Despite the apparent necessity of these truths Kripke argued, convincing almost the entire philosophical community, that such truths were knowable only *a posteriori*, and hence a new category of necessary truth was introduced into the philosophical literature – the so called necessary *a posteriori*.

However, there are two clearly distinct categories of necessary *a posteriori* truth within Kripke. The first kind, sketched above (and in Chapter 1), involve singular reference to an individual, and result from the substitution of co-referential terms within an identity sentence. Their necessity, as stated above, is the uncontroversial result of (*CDR*), an empirical (entirely uncontroversial) claim that the identity sentence is true, and some logic. As such, there is no special ‘metaphysical theory involving modality’ (Salmon 2005: 80) invoked to guarantee their necessity.

The second category is more controversial. A theoretical identification sentence like ‘Water is H<sub>2</sub>O’ also, according to Kripke, falls within the purview of the necessary *a posteriori*. However, such a sentence is clearly disanalogous to ‘Hesperus is Phosphorus’ in that the term on the right hand side of the identity appears (a) to ‘be descriptive’ (Salmon 2005: 82), and (b) bears a clear resemblance to other examples of the necessary *a posteriori*

that ‘involve singular reference to an individual but which do not result from substitution’ (*ibid.*). The most notable example from this latter category, which has nothing to do with semantics *per se*, is Kripke’s necessity of origin thesis. As others have noted, theoretical identifications appear to ‘resemble “Cicero is a product of the egg and sperm that generated him” rather than “Cicero = Tully” in respect that they aim to expose essence’ (LaPorte 2004: 37), and are thus not the product of the substitution of co-referential terms.

The necessity of origin thesis, as is well known, is an essentialist thesis concerning the origin of an individual. Kripke claims, about the origin of an object or person, that ‘anything coming from a different origin would not be this object’ (Kripke 1980: 113), and that ‘the *origin* of an object is essential to it’ (*ibid.*: 114, n57). Precisely how we should treat this thesis, as noted in chapter 3, section 2, is a matter of some controversy (cf. Hughes 2004: 111-7; Mackie 2006), however all we need note here, as Salmon does, is that Kripke employs a ‘nontrivial general principle of essentialism’ (Salmon 2005: 195) in his derivation of necessary *a posteriori* truths concerning the origins of individuals.<sup>45</sup>

Salmon’s distinction between trivial and nontrivial essentialism (2005: 82-7) is roughly as follows. Consider the identity sentence ‘Ehrich Weiss is Harry Houdini’, and the claim, following (*CDR*), that both ‘Ehrich Weiss’ and ‘Harry Houdini’ are rigid designators. We know from Kripke that the truth of the identity sentence guarantees its necessity given the rigidity of the two names flanking the identity predicate. But we can also say that the identity sentence ‘Ehrich Weiss is identical to Harry Houdini’ is necessary iff *that* man, Ehrich, could not have failed to have the property of *being identical to Harry Houdini*. In effect, Ehrich Weiss has the property of *being Harry Houdini* essentially, since it is impossible for Ehrich to

---

<sup>45</sup> I will discuss the distinction between trivial and nontrivial essentialism in greater detail in chapter 7, section 1.

exist and lack that property. But the property of *being Harry Houdini* is a property which we can apply to Harry Houdini *a priori*, since it is analytic that Harry Houdini has that property. Similarly with the property of *being Ehrich Weiss* and the man Ehrich Weiss. Once we discover *a posteriori* that Ehrich Weiss is Harry Houdini, we discover that the property of *being Ehrich Weiss* and the property of *being Harry Houdini* are properties that both Ehrich and Harry have necessarily. But these properties, as it goes, are pretty uninteresting properties. The property Ehrich Weiss has necessarily of *being Ehrich Weiss* is, to all intents and purposes, just the property of being self-identical. Hence the derivation that Ehrich Weiss has the property of *being Harry Houdini* necessarily is, at best, a trivial form of essentialism, and as such metaphysically uncontroversial.

Nontrivial essentialist conclusions, on the other hand, are metaphysically controversial. Take the case of individuals: following Kripke's model we might say that since Winston Churchill has his (actual) origin necessarily (say, being the son of Randolph Churchill and Jennie Jerome), he could not have existed and yet failed to have precisely the origin that he actually has, although he might not have done the things we commonly associate him with (being Prime Minister, for instance). But such a conclusion goes beyond the claim that Winston Churchill has the property of *being Winston Churchill* necessarily. It states, rather, that there is some particular origin that Churchill has, and that Churchill could not have existed whilst lacking that very origin. Such a conclusion is 'Aristotelian essentialism *par excellence*' (Salmon 2005: 86), and thus has significantly more metaphysical bite.<sup>46</sup>

---

<sup>46</sup> I say more about the distinction between trivial and nontrivial essentialism in chapter 7, section 1. The basic distinction is just that nontrivial essential properties require some additional theory of essentialism, such that the

The distinction between trivial and nontrivial essentialism is important here as theoretical identity sentences, such as ‘Water is H<sub>2</sub>O’, have more in common with nontrivial essentialist claims concerning the necessity of origin, than typical identity sentences. For example, if ‘water’ picks out a kind, and ‘H<sub>2</sub>O’ describes that kind’s chemical composition (as it does for Kripke, Putnam and Soames), being shorthand for ‘being composed of two parts hydrogen and one part oxygen’, then the claim that ‘necessarily water is H<sub>2</sub>O’ is a claim about water’s composition. Thus the claim that necessarily water is H<sub>2</sub>O is, in effect, a claim that it is impossible for water to exist and *not* be so composed. This suggests that ‘the chemical composition of a compound is an essential feature of that compound’ (Salmon 2005: 82), in much the same way as Kripke claims that origin is an essential feature of an individual.

But this is a far more controversial conclusion than the more general semantic conclusion that true identities, theoretical or otherwise, between two co-referring names, are guaranteed to be necessary when true. The only properties that you can infer about Ehrich Weiss, that he bears necessarily, once you discover the Ehrich Weiss is Harry Houdini are properties like *being Harry Houdini*. There is no obvious entailment from the fact that Ehrich Weiss is Harry Houdini that Ehrich *necessarily* has the parents he actually has. However, if there is such an entailment, then these essentialist ‘facts’, about the origins of individuals and the composition of substances, were allegedly uncovered via a ‘relatively simple theory about the reference of single words’ (Salmon 2005: 87). The key question, then, is whether this second category of necessary *a posteriori* truth (theoretical identities and claims about the origins of individuals) are derivable from (*CDR*) in conjunction with some uncontroversial empirical observations. Salmon’s answer, as we shall see, is that they are not.

---

origin of an individual is an essential property of that individual, whilst trivial essential properties do not. However, more on this in chapter 7.



#### 4.5 Kripke's essentialist premise

The majority of Salmon's discussion of the derivation of necessary *a posteriori* truths focuses on Putnam, and Donnellan's (1983) elucidation of Putnam. His criticism, in a nutshell, is that the only essentialist conclusions that follow from (*CDR*) are those that stem from nontrivial essentialist premises. Here, however, I wish to focus on how this criticism might apply to Kripke, especially in light of chapter 2, where I distinguished early-Putnam from later-Putnam, and sketched Putnam's rejection of the metaphysical reading of his conclusions (viz., the necessary *a posteriori*). Thus, the following explication of Salmon's account of, and objection to, Putnam's mechanism for generating *a posteriori* necessities for natural kinds and individuals should be read as applying only to early-Putnam. The purpose of this exposition, however, is to use it, in conjunction with an analysis of Kripke's necessity of origin thesis, to illustrate the clear similarities between necessity of origin claims and theoretical identities, and to show that Kripke endorsed the self same mechanism that Salmon attributes to early-Putnam (and Donnellan).

Salmon presents the following argument, labelled the *OK*-mechanism<sup>47</sup> ('*O*' for ostension, '*K*' for kind), as Putnam's (and Donnellan's) attempt to derive the necessary, nontrivial essentialist conclusion, such as 'necessarily, gold is the element with atomic number 79', from (*CDR*), uncontroversial empirical observations, and no nontrivial essentialist premises. The mechanism (see Salmon 2005: 166), using water as an example, is:

---

<sup>47</sup> The *OK*-mechanism is a restricted version of a more general mechanism, the *K*-mechanism. However, since my target here is Kripke, and Kripke's thesis critically includes ostension rather than Putnam's operational definition, I will only present the *OK* version.

(T1) It is necessarily the case that: something is a sample of water if and only if it is a sample of *dthat* (the same substance as *this* is a sample of).

(T2) *This* (liquid sample) has the chemical structure H<sub>2</sub>O.

(T3) Being a sample of the same substance as something consists in having the same chemical structure.

Therefore,

(T4) It is necessarily the case that: every sample of water has the chemical structure H<sub>2</sub>O.

The *OK*-mechanism, then, consists of an ostensive definition, (T1), a description of the hidden nature of the paradigm sample, referred to via the ostensive definition, (T2), a claim about what being the same kind (as the kind in question) consist in, (T3), and the conclusion asserting the necessity of the generated truth, (T4). Since (T1) is clearly a consequence of (*CDR*), and (T2) is an uncontroversial empirical observation, the focus of the investigation is (T3).

Salmon settles on the following formalisation of (T3), in a bid to determine precisely what ‘consists in’ means: ‘if  $x$  exists in  $w_1$  and  $y$  exists  $w_2$ , and if  $x$  is a sample in  $w_1$  of the same substance that  $y$  is a sample of in  $w_2$ , then whatever chemical structure  $x$  has in  $w_1$ ,  $y$  has that same chemical structure in  $w_2$ , and vice-versa’ (Salmon 2005: 179-180). On this interpretation the ‘consists in’ relation in (T3) is effectively being cashed out as a cross-world

sameness (of chemical structure, in this instance) relation, which fits with the analysis of Putnam given in Chapter 2.

However, it is worth noting, briefly, that Salmon treats this sameness relation as merely a necessary condition. Thus he says that ‘the third premise schema of the *OK*-mechanism must assert, at a minimum, a necessary condition of the cross-world *same-K* relation’ (2005: 181). However, this fails to capture what Kripke and (early) Putnam intended. Whilst I agree with Salmon that (T3) *at a minimum* must assert a necessary condition, it seems to me that the original force of the argument is that it be a necessary *and* sufficient condition. Indeed, the most natural interpretation of ‘consists in’ seems, obviously, to be a necessary and sufficient condition. Moreover, there is decisive exegetical evidence illustrating that both Kripke and Putnam were interested in necessary and sufficient conditions – that is, essences rather than essential properties.

The exegetical evidence in Putnam (1973) quite clearly demonstrates his intention that the notion of ‘consists in’ captures a necessary and sufficient condition. When he defines what sameness of substance (or liquid, in the original article) consists in, he states quite clearly that the ‘necessary and sufficient condition for being [x] is bearing the relation *same<sub>L</sub>* to the stuff’ (Putnam 1973: 702) identified as the paradigm sample. Similarly, Kripke’s statement that:

a material object is (pure) gold if and only if the only element contained therein is that with atomic number 79. Here, the “if and only if” can be taken to be *strict* (necessary)... the coextensiveness of the predicates is *necessary*, but not *a priori*’ (Kripke 1980: 138)

also clearly indicates a stronger reading of what ‘consists in’ amounts to: namely, as a necessary-and-sufficient condition, rather than merely a necessary condition. Salmon’s interpretation of (T3), then, seems clearly implausible given the exegetical evidence. However, since Salmon’s criticisms apply to Kripke and early-Putnam given the weaker reading of (T3), they apply *a fortiori* to a stronger reading of (T3), hence my observation has no impact for the overall argument. Still, it is worth noting that both Kripke and early-Putnam claim that H<sub>2</sub>O is *the* essence of water, rather than merely an essential property of water. Similarly, Soames recognises the use of the biconditional in Kripke, and analyses theoretical identities as having the logical form  $\forall x [Fx \leftrightarrow Gx]$ . Salmon is just wrong to treat ‘consists in’ as merely a necessary condition.

#### **4.6 Essentialism, *a priori* or *a posteriori*?**

According to Salmon, there is one particular difference between Kripke’s derivation of necessary *a posteriori* truths concerning natural kinds, and the Putnam-Donnellan derivation, which it is worth noting. The second half of Salmon’s *Reference and Essence* is dedicated to showing that nontrivial essentialist conclusions cannot be derived merely from semantic premises.<sup>48</sup> The key premise identified by Salmon as that which enables nontrivial essentialism to be derived, is (T3).

Salmon’s criticism is that (T4) critically relies upon (T3), and (T3) is a ‘nontrivial general principle of essentialism’ (Salmon 2005: 184). Hence, the derivation of necessary *a*

---

<sup>48</sup> Although I have said a little concerning the distinction between nontrivial and trivial essentialism, I am well aware that I have left much to be said. However, I will deal explicitly with what essentialism, and claims to the effect that H<sub>2</sub>O is the real essence of water, amount to in Chapters 6, 7 and 8.

*posteriori* truths concerning natural kinds (and, by analogy, the necessity of origin thesis) are *not* merely the product of philosophical semantics. To claim that essentialism can be derived just from linguistic premises is just to beg the question of essentialism; in other words, the only essentialism you get out from the philosophy of language is that which you put in, since the necessity of ‘water is H<sub>2</sub>O’, at least as it is presented in Kripke and Putnam, ‘depends crucially on a premise, which is a contestable essentialist principle that does not in any way seem to come from the philosophy of language’ (Salmon 2005: 186).

Salmon observes that Putnam and Donnellan appear to claim that (T3) is a scientific principle, and mistakenly take the derivation of necessary truths like ‘water is H<sub>2</sub>O’ to be the product of philosophical semantics and uncontroversial empirical observations/discoveries. But (T3) is ‘a principle whose truth is established in good part by conceptual analysis and the reflective methods of the metaphysician’ (Salmon 2005: 185), rather than scientific theorising. In a sense, the claim that philosophical semantics in conjunction with (T3) entails nontrivial essentialism is trivially true since (T3) itself is a statement of nontrivial essentialism, and as such when combined with *any* semantic theory (including a descriptivist semantics, as we shall see in chapter 5) it will generate essentialist truths.

Kripke, by contrast, does not appear to treat (T3) as a scientific truth. Rather, again as Salmon notes (2005: 195), he appears to treat it as a conceptual truth, knowable *a priori*. Thus we know *a priori* that if a chemical kind *K* has a molecular structure/constitution *C* that it is necessary that *K* have *C*. If *K* is the kind water, and we discover (empirically) that water has the molecular constitution H<sub>2</sub>O, then it follows from our *a priori* essentialist principle (T3), that necessarily water is H<sub>2</sub>O. This is why the discovery that water is H<sub>2</sub>O is automatically the discovery that necessarily water is H<sub>2</sub>O. The difference between Kripke on the one hand, and Putnam and Donnellan on the other, at least as Salmon sees it, is that Kripke ‘explicitly

recognizes the essentialist import of the major premise in his modal arguments generating necessary *a posteriori* truths' (Salmon 2005: 196), and avows the ambition to *prove* essentialism from a theory of reference (Kripke 1980: 1, n1).

The point that Salmon is really intending to make is that Putnam, but not Kripke, claimed to have derived necessary truths knowable only *a posteriori* entirely from philosophical semantics. However, the distinction between Putnam and Kripke on this basis seems obviously spurious. As Salmon quite rightly notes, (T3) must be established by conceptual analysis. However, Putnam's Twin Earth thought experiment is one of modern philosophy's paradigms of conceptual analysis. Moreover, inasmuch as Kripke offers *a priori* arguments for (T3), they are (independently developed) variations on the Twin Earth thought experiment. Furthermore, Kripke claims explicitly that 'science discovers the nature[s], and thus the essence[s]' of natural kinds (Kripke 1980: 138). Whilst Putnam is surely right that neither Kripke nor Putnam manage to derive essentialism from (*CDR*), since both employ (T3), it seems that both offer some arguments for (T3) in the form of conceptual analysis.<sup>49</sup> Hence, if there is a distinction between the projects of Kripke and Putnam, it is not on this basis.<sup>50</sup>

The important point for this thesis is that there are two apparent routes to necessary *a posteriori* truth: the first being the product of the substitution of co-referential names and (*CDR*), such as 'Ehrich Weiss is Harry Houdini', the second being the product of nontrivial

---

<sup>49</sup> In some sense, then, it is not clear why Salmon labels (T3) as an essentialist assumption, since it is rather the conclusion of independent *a priori* argument.

<sup>50</sup> In chapter 2, section 5, I observed that Putnam was concerned with the threat of Kuhnian incommensurability, whilst in chapter 1 it was obvious that Kripke's concerns were with descriptivism. This, or so it seems to me, is where the two projects differed. Not, as Salmon would have it, in that Kripke's essentialism is *a priori* whilst Putnam's is *a posteriori*.

essentialist premises like (T3), and including truths like ‘necessarily, water is H<sub>2</sub>O’. In the literature the two categories are not always properly distinguished, but the distinction suggests a possible (novel) descriptivist rejoinder, utilising an essentialist premise like (T3), and arguing that only descriptions of the essential properties of kinds can guarantee determinate reference. This is the topic of the next chapter.

#### **4.7 Salmon’s account of theoretical terms**

Before moving on to a defence of descriptivism, it is worth presenting, and ultimately rejecting, Salmon’s alternative account of necessary *a posteriori* theoretical identity sentences. Salmon, as I have indicated, recognises a division between two types of necessary *a posteriori* truth in Kripke: (i) those like ‘Erich Weiss is Harry Houdini’ and ‘Hesperus is Phosphorus’, and (ii) those like ‘water is H<sub>2</sub>O’ and ‘Winston Churchill is the son of Randolph Churchill and Jennie Jerome’. As we have seen (ii) is the product of a not uncontroversial nontrivial essentialist premise, whilst (i) is metaphysically neutral.

Salmon’s alternative account (cf. Salmon 1987/8: 193-217) attempts to maintain Kripkean insights into proper names and apply them straightforwardly to natural kind terms by treating natural kind terms *tout court* as names for abstract universals. The basic move is to restrict the sorts of things that general terms can designate. For Salmon, where *T* is a general term, *T* cannot be taken to designate its actual extension, since a Millian theory of meaning says that the meaning of a term is its referent, and the actual extension of a general term is liable to change over time. For instance, were the general term ‘Human’ to designate the extension of the human-kind, then since new humans are born every day, and other humans die, the fluctuating extension of the human-kind would entail a fluctuating meaning of the

general term 'Human'. Hence, Salmon claims that general terms should be thought of as designators for abstract universals. So, where  $T$  is the mass noun 'Water', we should construe  $T$  as a designator of the water-kind, which is an abstract universal. It follows from this that where  $T$  successfully designates a kind,  $T$  rigidly designates that kind, in precisely the way that all successfully introduced names are, in virtue of being designators, rigid designators.

The problem cases for Kripke were descriptions like 'H<sub>2</sub>O', – where 'H<sub>2</sub>O' is read as being synonymous with 'the substance consisting of two hydrogen atoms and an oxygen atom' (as it is by Soames) – 'the element with atomic number 79', or statements about the origin of an individual. In the case of natural kind terms, Salmon opts to treat 'H<sub>2</sub>O' and 'the element with atomic number 79' in much the same way as he does 'Water' or 'Gold', which is to say as substance *names* designating abstract universals. On such a view a term like 'H<sub>2</sub>O' is 'a general-term version of a proper name whose reference is fixed through a scientific convention concerning chemical-compound terms' (Salmon 2003: 488). Consequently, the apparent necessity of a true theoretical identity sentence, for instance 'Water is H<sub>2</sub>O', really is a product of philosophical semantics since *both* general-terms are the analogues of logically proper names, and the two general terms are co-referential.

The main argument for this claim, if indeed it can be called such, actually occurs in Salmon (1987/8), in footnote 5, where he contrasts the occurrence of the '2' in the phrase '2 meters' with its occurrence in the proper name 'R2-D2', and notes that only in the former case does '2' seem to be replaceable by a variable in an existential generalisation. In a later paper he invites us to 'note the oddity of  $(\exists x)(\exists y)[x = \text{the most abundant element} \ \& \ y = \text{the smallest prime integer} \ \& \ \text{water} = x_y\text{O}]$ ' (2003: 488, n25). When we break down a general term such as 'H<sub>2</sub>O' into what we take to be its component parts, and substitute them as in the above logical formula, we generate highly odd results where 'H<sub>2</sub>O' is synonymous with the



conjunction of ‘the most abundant element’, ‘the prime integer’ and ‘oxygen’. Terms like ‘H<sub>2</sub>O’, then, are just technical names, and analogous to any other proper name.

It is difficult to know what to make of this ‘argument’. In the first instance, the ‘2’ in ‘H<sub>2</sub>O’ clearly has more in common with the ‘2’ in the phrase ‘2 meters’ than it does ‘R2-D2’, since it really does mean two: there are *two* hydrogen atoms, and *one* oxygen atom. The point is even more obvious when we consider phrases like ‘the element with atomic number 79’ where, again, the number ‘79’ is doing precisely the work one would expect, rather than the slightly odd work that numbers are doing in the name ‘R2-D2’. Hence I am not entirely sure that I even understand what Salmon is trying to say. The argument for treating all referring general terms as the analogue of proper names, then, seems to rely upon two concerns about treating them as descriptions: (i) that to treat only some descriptions as rigid seems to require the controversial metaphysical thesis of essentialism, and (ii) the apparent oddity of breaking down terms like H<sub>2</sub>O into their constituent parts. However, (ii) seems plainly false, since empirical formulae like ‘H<sub>2</sub>O’ only make sense when we realise that they *literally* tell us something about the ratios of the constituent atoms of a molecule, and whilst (i) maybe *prima facie* plausible, when I have established, in Chapter 5, that the combination of essentialism with descriptivism makes for a defensible thesis, in Chapter 8 I will give a deflationary account of essences that will also highlight that (i) is false.

Given that Soames’ alternative (discussed in Chapter 3) is to treat general terms like ‘H<sub>2</sub>O’ as synonymous with descriptions, in this case ‘the chemical compound consisting of two parts hydrogen and one part oxygen’, the choice seems to be whether to admit that some descriptions can be rigid, but restrict the class of rigid descriptions via the claim that the kind in question ‘has an essential property’ (Salmon 2003: 488). But this means that such a theory of reference comes with substantive metaphysical baggage. Alternatively, we can endorse

Salmon's proposal that all referring general terms are like proper names. But this comes with the immediate consequence that such an account looks plainly false for theoretical terms like 'H<sub>2</sub>O' and 'the element with atomic number 79'. Furthermore, the view appears to be motivated by a rather dubious and apparently flimsy argument. However, there are further negative consequences entailed by Salmon's thesis that I have yet to mention.

#### **4.8 The trivialisation problem**

One important objection to Salmon's proposal is aimed at his claim that general terms only designate abstract universals, and the consequence of this claim that every general term designator is a rigid designator. The objection, simply, is that it trivialises the notion of rigidity (cf. Schwartz 2002, Salmon 2002). Following Dan López De Sa, let us call this problem – the problem that it follows trivially from any general term being a designator that it is a rigid designator – the 'trivialization problem' (López De Sa 2007: 117). The effect of the trivialisation problem is that there is no semantic distinction to be drawn between natural kind terms like 'Water', 'Gold' and 'Tiger' on the one hand, and general terms like 'Member of a species that serves as a mascot for Chicago's NFL team', since, *prima facie* at least, each can be taken to designate a property, and it follows trivially from being a designator that it is rigid.

Salmon has responded to concerns over trivialisation, at least as they are articulated by Soames (2002: 250), in the following fashion. While it is true that many general terms *do* turn out to be rigid, including terms like 'Bachelor' which we might not think of as standard examples, there are still plenty of general terms that fail to be rigid, and hence rigidity is not trivial. Consider, for instance, the sentence 'A Bear is a member of the species that serves as a

mascot for Chicago's NFL team'. According to Salmon (2005a: 124), this sort of sentence is most naturally read as a kind of identity sentence. This identity sentence is 'true if and only if the terms flanking the "is" are co-designative' (Salmon 2005a: 124). However, the term on the left, 'Bear', is *not* co-designative with 'Member of a species that serves as a mascot for Chicago's NFL team', since the intension of the latter, but not the former, is such that it might have assigned, at a world like  $w_1$ , a class of instances consisting of all buffalos, rather than all bears. Hence the metaphysical intension of the kind *Member of a species that serves as a Mascot for Chicago's NFL team* is *not* the same as the *Bear* kind, and hence the corresponding general terms are *not* co-designative. Rather, as stated earlier, they are contingently co-extensional, and as we have already established this is insufficient for the truth of a theoretical identity. However, what the example also shows, in addition to the falsity of the alleged identity, is that the definite description 'Member of a species that serves as a Mascot for Chicago's NFL team' is not rigid, since it applies to bears in the actual world, but buffalos at nearby possible worlds, such as  $w_1$ . Thus there are still 'non-rigid general terms everywhere' (Salmon 2005a: 132).

But Salmon's response is unsuccessful. Consider the following example: let the general term 'Object that weighs the weight of three packets of peanuts' pick out the kind *Object that weighs the weight of three packets of peanuts*, and let the extension of that general term be any object that is an instance of something with the weight of three bags of peanuts. Given Salmon's response above, the general term 'Object that weighs the weight of three packets of peanuts' looks to be a non-rigid designator. The point is easily illustrated: if a packet of peanuts in the actual world,  $w_@$ , weighs 27g, then the combined weight of three packets of peanuts would be 81g. Hence the kind referred to be the general term 'Object that weighs the weight of three packets of peanuts' has, as its extension, anything weighing 81g.

However, go to a nearby possible world,  $w_I$ , where the weight of a packet of peanuts is 30g, and the combined weight of three packets is 90g. In  $w_I$  the general term ‘Object that weighs the weight of three packets of peanuts’ will apply to anything weighing 90g, rather than 81g. In other words, the general term has a different extension at  $w_I$  than it has at  $w_@$ , and cannot therefore be a rigid designator.

However, all that this shows is that the general term ‘Object that weighs the weight of three packets of peanuts’ does not rigidly designate either 81g or 90g. It does *not* show that it is not a rigid designator at all. For Salmon, being a rigid designator follows trivially from being a designator. Hence, provided there is ‘an appropriate property (whose extension necessarily tracks the extension of the general term)’ (Lopez de Sa 2007: 118) to designate, then it won’t be that the general term ‘Object that weighs the weight of three packets of peanuts’ *flexibly* designates objects with different weights in different possible worlds. Rather, the general term ‘Object that weighs the weight of three packets of peanuts’ will designate anything with the property of *being an object that weighs the weight of three packets of peanuts* in all possible worlds, regardless of how heavy those objects contingently happen to be.

Let me put the point more formally: let  $f$  be the function that assigns a value of 1 to all objects that bear the property of *being an object that weighs the weight of three packets of peanuts*, and 0 to all other objects. Let the general term ‘Object that weighs the weight of three packets of peanuts’, abbreviated to  $T$ , designate the function  $f$ . Let  $w_@$  be the actual world, where the weight of three packets of peanuts is 81g, and let  $w_I$  be a nearby possible world where the weight of three packets of peanuts is 90g. Now consider what it is that the class of objects that weigh 81g, call that  $O$ , in  $w_@$ , have in common with that class of objects that weigh 90g, call that  $O^*$ , in  $w_I$ ? The answer, perhaps obviously, is that both classes of

object,  $O$  and  $O^*$ , are the product of the function  $f$  that assigns 1 to all objects that have the property of *being an object that weighs the weight of three packets of peanuts*. It does not matter that the class of objects  $O$  weigh 81g, while the class of objects  $O^*$  weigh 90g. What is important is that both  $O$  and  $O^*$  have the property of *being an object that weighs the weight of three packets of peanuts*. Hence the general term ‘Object that weighs the weight of three packets of peanuts’ *does* rigidly designate, it is just that it rigidly designates anything with the property of *being an object that weighs the weight of three packets of peanuts*. Hence, this still appears to trivialise the notion of rigidity.

According to López De Sa, it must be conceded that the trivialisation problem only arises if we have ‘a very “liberal” view about the entities general terms are said to designate’ (2007: 119). However, it seems to me that he need not be quite so concessive. Salmon, by his own admission, claims that examples like ‘A Bear is the Member of a species that serves as a mascot for Chicago’s NFL team’ are most naturally read as identity sentences. Since there is no substantive difference between Salmon’s own example, and the one I have employed to articulate the trivialisation problem, the concession is unnecessary. Salmon’s view of general terms designation entails the trivialisation problem, since he cannot distinguish between rigid and non-rigid general terms.<sup>51, 52</sup>

---

<sup>51</sup> There have been some more recent responses in the literature, which I have not had time to discuss here. For instance, Martí and Martínèz-Fernandez (2010) argue that we can distinguish between rigid and non-rigid readings of general terms. However, since my overarching aim in this thesis is to show that there is no interesting connection between the semantics and metaphysics of natural kinds (and natural kind terms), I have not included that discussion here. In any event, my main objection to Salmon’s account, in chapter 4 section 9, is that his interpretation of the semantics of theoretical terms is implausible in light of actual practices.

<sup>52</sup> Similarly, it is worth noting that Soames might also be liable to the trivialisation problem. At the end of chapter 3, I observed that Soames does not really give a story about why there is an asymmetry between natural kind predicates and other predicates. Moreover, any story he does give appears to rely upon some significant metaphysical assumptions about natural kinds. Hence, either Soames owes us a story about natural kinds such

#### 4.9 The semantics of science

The second (rarely discussed, to the best of my knowledge) problem with Salmon's account is that it simply fails to cohere with our actual practices. While it is certainly true that general terms like 'H<sub>2</sub>O' are introduced by scientists, and are (in part) the product of scientific theorising, that they are not synonymous with certain descriptions is far from obvious, and indeed I will claim that in many cases they are. However, it should be conceded that Salmon is quite correct that *if* we treat 'H<sub>2</sub>O' as the general-term analogue of a proper name, then 'Water is H<sub>2</sub>O' is necessary. Nevertheless, this still leaves work to be done carving out a distinction between rigid and non-rigid designators. Furthermore, if Salmon is endorsing Millianism, and the meaning of a term just is its referent, then 'Water is H<sub>2</sub>O' should be *a priori* if 'Ehrich Weiss is Harry Houdini' is (which, recall, is what Soames conceded in Chapter 3). However, there are compelling reasons *not* to treat 'H<sub>2</sub>O' as the analogue of a logically proper name.

Since general terms such as 'H<sub>2</sub>O' are the product of scientific theorising, we should look to how scientists *actually* use such terms, in order to get a better sense of their semantics. The first thing to note is that chemists standardly distinguish between types of chemical formula, which is what 'H<sub>2</sub>O' undoubtedly is. An *empirical* formula merely expresses the proportion of distinct elements within a compound. The empirical formula of butane, for example, is C<sub>2</sub>H<sub>5</sub>, but it would be a mistake to infer from this that butane molecules actually have two carbon atoms and five hydrogen atoms. The *molecular* formula of butane, in

---

that they ground the asymmetry between natural kind predicates and other predicates, or he too is victim of the trivialisation problem. However, I do not have the space to address that issue here.

contrast, is  $C_4H_{10}$  and this *does* tell us precisely how many atoms of distinct elements constitute a molecule of butane. Finally, *structural* formulas tell us not just the number of atoms within a molecule, but also give us their respective arrangements: for instance, propane has the molecular formula  $C_3H_8$ , and the structural formula  $CH_3CH_2CH_3$ . The problem is that as a layperson it will not always be immediately obvious which sort of formula one is faced with, and thus unclear precisely what role that formula is playing. This ambiguity between general terms that are syntactically very similar adds an additional layer of complexity to general terms such as these, that has not been acknowledged in the literature.

The distinction between types of formula also offers some insight into the *function* of what we might label broadly as chemical formulae within scientific discourse: they encode different types of information about the molecules to which they apply. The structural formula for propane tells us that there is a chain of three carbon atoms, with the hydrogen atoms arranged around these central carbon atoms, three on the left-hand carbon atom, two on the central carbon atom, and three on the right-hand carbon atom. It is clear from these simple observations that structural formulae do not merely denote, as Salmon would have us think, but connote *as they are used by the scientist*. As such it is hard to parse Salmon's claim that 'H<sub>2</sub>O' and its ilk are the general-term analogue of logically proper names, since the role of chemical formulae within chemistry is clearly intended to describe as well as denote.

In light of the role that chemical formulae play within the semantics of chemistry we should, I conclude, reject Salmon's account of the semantics of at least some general-terms, particularly those from chemistry. Salmon's account, at best, applies to a very limited class of cases, paradigmatically represented by the 'Apatosaurus is Brontosaurus' case. However, none of Kripke's original cases conform to this paradigm. Indeed, phrases like 'the element with atomic number 79' have more in common with phrases like 'Member of a species that

serves as a mascot for Chicago's NFL team' than they do with single world natural kind terms (e.g. 'Gold', 'Tungsten', 'Muon'), but on Salmon's view whilst 'the element with atomic number 79' should be rigid, the phrase 'Member of a species that serves as a mascot for Chicago's NFL team' should not be. The trivialisation problem demonstrates that this is not the case, and hence Salmon's defence of Kripke fails to carve out a special semantic category of general terms – the natural kind terms – that are semantically special.



## TAKING STOCK: THE STORY SO FAR

Given that paradigm Kripkean theoretical identities contain, on the right hand side, descriptions, I have concluded that Salmon's attempt to treat such terms as the general term analogue of proper names is highly implausible. In this respect, then, Soames' thesis, that recognises precisely this point, is preferable to Salmon's. Furthermore, Salmon's view, as we have seen, fails to effectively insulate a special class of general terms (the natural kind terms) that are rigid, whilst leaving all other general terms as non-rigid.<sup>53</sup> Soames' view seems to achieve this end, but only at the cost some *prima facie* serious metaphysical baggage. According to my earlier criticism, the only way to make a description like 'substance molecules of which contain two hydrogen atoms and one oxygen atom' figure in a necessarily true theoretical identity, whilst maintaining that descriptions like 'potable substance that falls from the sky and flows in the lakes and rivers' do not, is to say that only the former latches onto the essence of the kind. Hence some version of metaphysical essentialism is required to ground the necessity of paradigmatic theoretical identities.

In some sense, this seems like just the right thing to say. Kripke unabashedly talks about the essence of the kind, and according to Salmon's analysis offers some *a priori* arguments for essentialism. Indeed, one can view arguments like the Twin Earth thought experiment less as semantical arguments (arguments that aim to show that reference is underdetermined by description) and more as straightforward conceptual analysis, revealing

---

<sup>53</sup> Although, it should be noted, that the debate over trivialisation is ongoing, and some philosophers, for instance Martí and Martínèz-Fernandez (2010) are confident of a solution.

to us precisely what the conditions are on being such-and-such a kind. The implicit essentialism, then, should perhaps be no surprise.

On that basis, elements of Soames' view appear to be cohere better with what Kripke intended, compared to Salmon's view. However, Soames' positive thesis, recall, also required a decidedly odd view of predicate reference (natural kind predicates designate natural kinds), a false notion of 'type', and assumed that no object could be a member of two natural kinds of the same type – the special auxiliary assumption – which is also false. Hence it is not particularly satisfactory either.

In addition, there is also the concession made by Soames that if you are a Millian about semantic content, and you think that the contribution that a name or natural kind term makes to a proposition is just its referent, then co-referring terms can be substituted *salva veritate*, since they make identical contributions to the proposition. The upshot of this for proper names was that identities like 'Ehrich Weiss is Harry Houdini' were knowable *a priori*, not *a posteriori* as Kripke would have us believe. Since Salmon, like Soames, is also a Millian about proper names, he too thinks that identity sentences containing proper names, for instance 'Hesperus, if it exists, is Phosphorus', are in actual fact necessary and *a priori* (cf. Salmon 1986: 2). This has the further implication for Salmon that given his treatment of natural kind terms as names, then where 'H<sub>2</sub>O' is a name that refers to the same kind as 'Water', the theoretical identity sentence 'Water is H<sub>2</sub>O' will be as *a priori* and as necessary as 'all bachelors are unmarried men'!

Soames' own view was to endorse Millianism for simple natural kind terms, but to deny that terms like 'H<sub>2</sub>O', that is complex natural kind terms, were directly referential (i.e. they are not Millian). Instead, Soames claims that the content of such an expression is the

property that *determines* the kind; hence, reference to the kind is mediated reference in the case of complex natural kind terms (which are descriptions) like ‘H<sub>2</sub>O’, but direct in the case of simple natural kind terms like ‘Water’. Theoretical identities like ‘Water is H<sub>2</sub>O’ (the simple-complex type) are thus *a posteriori* because the two terms mean different things. But, as I said above, the distinction between kind-determining properties and non-kind-determining properties seems to rely upon metaphysical essentialism. Moreover, the move to save *a posteriority* in one special case (simple-complex theoretical identities), while conceding it in all other cases (identities and simple-simple theoretical identities are *a priori*) appears somewhat *ad hoc*.

It would seem, then, that the most plausible account, which recognises terms like ‘H<sub>2</sub>O’ as descriptions, is accompanied by the metaphysical thesis of essentialism. But since this is an independent metaphysical position, it can be separated from the causal theory of direct reference, and combined with other semantic theories. In the next chapter I will demonstrate that when essentialism is combined with descriptivism, descriptivism is not susceptible to Kripke’s original objections. Hence with no additional metaphysics (and indeed, compared to Soames’ thesis, with a good deal less), a simple version of descriptivism can be made tenable.

Finally, it is worth stressing that (*EM*) is committed to the thesis that theoretical identities are actually knowable *a priori*, and that although Kripke backed off from giving a theory of semantic content (as I noted in Chapter 3), it is plausible to suppose that his endorsement of Mill extended to an endorsement of Mill’s view of content. However, since this is controversial, I simply note the following: either Kripke is a Millian, and both identities and theoretical identities come out *a priori*, or Kripke is not a Millian, and the thesis in *Naming and Necessity* fails to show that propositions like that expressed by ‘Hesperus is

Phosphorus' or 'Water is H<sub>2</sub>O' are knowable only *a posteriori*. Hence Kripke is either mistaken about many necessary *a posteriori* truths (since they are actually *a priori*), or the lacunae that Soames identifies in *Naming and Necessity* (Kripke's lack of a theory of semantic content) is more of a gaping, cavernous pit that, somehow, the majority of proponents of *a posteriori* necessities have ignored. Either way, the prospects for the necessary *a posteriori* are at best incomplete, and at worst looking bleak.

Now, having identified the essentialism running through the plausible accounts of theoretical identity and (*CDR*), I turn (in chapter 5) to a presentation of an alternative account of the semantics of natural kind terms that employs essentialism (as Kripke's, (early) Putnam's and Soames' accounts all do), but which answers some of the key questions concerning the meaning of natural kind terms. With this in place, the remainder of the thesis addresses questions concerning the metaphysics of essence.

## CHAPTER 5

In chapter 1 I identified three types of objection to descriptivism. Following Salmon I labelled these objections (i) the modal argument, (ii) the epistemological argument, and (iii) the semantical argument. In chapter 4 I established that these arguments do not *refute* descriptivism about natural kind terms, as they perhaps do about proper names, since for all that has been said it might just be that we have thus far considered the wrong descriptions. As Salmon has intimated, what the descriptivist needs is some *general* property that is indicative of the kind, but which does not itself involve direct reference to the kind. In this chapter I will present a version of descriptivism – *Descriptivist Essentialism* – that avoids all of these objections as they apply to natural kind terms. The key to avoiding Kripke’s objections will be to combine descriptivism with metaphysical essentialism: after all, Kripke’s arguments against the descriptivist are so successful because they target descriptions of accidental or contingent properties of natural kinds e.g. ‘liquid that falls from the sky, and flows in the lakes and rivers’ for ‘water’. But it is not an essential feature of descriptivism that only descriptions of the accidental properties of natural kinds can give the meanings of natural kind terms. The point of the chapter, however, is not to endorse essentialism, but to illustrate that if all the plausible versions of (*CDR*) employ an essentialist premise, as I claimed in the previous chapter, then since essentialism comes apart from (*CDR*), descriptivism too can be made plausible using this same premise.

The plan of this chapter will be as follows: in section 2 I define a broad notion of descriptivism and remind us of Kripke’s classic objections to descriptivism. In section 3 I briefly recap Kripke’s positive thesis concerning natural kind terms and rigidity, and remind

the reader of Putnam's Twin Earth thought experiment, which neatly captures the intuition underlying Kripke's semantics. In section 4 I offer a new analysis of Kripke's thesis concerning the rigidity of natural kind terms that plays close attention to the distinction between *de jure* and *de facto* rigidity, which has been neglected in the literature. I argue that theoretical identifications are necessary because they contain *de facto* rigid designators, and that, given what *de facto* rigidity amounts to, it is Kripke's metaphysical essentialism that grounds the necessity of theoretical identifications. As such I combine Kripke's metaphysics with a descriptivist semantics and steer it around both the modal and semantical objections. In section 5 I deal with the epistemological objection by (i) endorsing a division of linguistic labour, (ii) offering a new story of how the reference of natural kind terms is fixed, and (iii) borrowing a strategy recently developed by Soames to defend Extended Millianism (*EM*). In section 6 I discuss the application of the thesis to the majority of natural kind terms and in section 7 I sum up the main achievements of the chapter.

### **5.1 Kripke vs. Description Theory**

Let me define a broadly descriptivist theory as one which states that certain terms within a language, most notably names and natural kind terms, have a *sense*, associated with the name or natural kind term by speakers. Let me narrow my concern to natural kinds terms, remaining agnostic about names, and say that the *sense* of a natural kind term is a description associated with that term (again, by speakers) such that it supplies 'a set of conditions, or properties, and the denotation, if any, is what uniquely satisfies those conditions' (Salmon 2005: 9).

Contrast this with (*CDR*) where we 'baptise' objects via ostension (or, sometimes, description) and the reference of the name is passed on causally, from speaker to speaker,

throughout the linguistic community. Names refer *directly* because their reference is not mediated by description, and, as a corollary of being nondescriptonal (cf. Soames 2002: 264), they are also *rigid designators*: ‘in every possible world [the name] designates the same object’ (Kripke 1980: 48). Now, as ‘names are always rigid designators’ (Kripke 1980: 58), and ‘[t]he old term “common name” is thus quite appropriate for... mass terms, such as “gold”, “water” and the like’ (Kripke 1980: 127), natural kind terms, as a species of name, are also nondescriptonal, and therefore always rigid designators (provided, following Salmon, that they designate at all).

The three standard arguments that (allegedly) show us why we should both reject the descriptivist theory, and endorse Kripke’s alternative (as it is not subject to the same criticisms), are as follows: take ‘gold’ as our paradigm natural kind term, and imagine that ‘the yellow, heavy, malleable, ductile metal found in the Sierra Nevada’ is the description said by the descriptivist to give the meaning of ‘gold’ and to determine the denotation of that natural kind term. If ‘gold’ *is* descriptive in this way then a descriptivist language user would assert (1):

- (1) If gold exists, then gold is the yellow, heavy, malleable, ductile metal found in the Sierra Nevada.

As the natural kind term ‘gold’ is synonymous, within the proposed descriptivist’s dialect, with the description ‘the yellow, heavy, malleable, ductile metal found in the Sierra Nevada’, (1) is synonymous with (2):

(2) If the yellow, heavy, malleable, ductile metal found in the Sierra Nevada exists, then the yellow, heavy, malleable, ductile metal found in the Sierra Nevada is the yellow, heavy, malleable, ductile metal found in the Sierra Nevada.

The nub of the modal argument is that (2) expresses a proposition which is true necessarily. Since (2) is synonymous with (1), (1) also expresses a proposition that is necessary. The objection is that we can imagine metaphysically possible situations where (1) is false: imagine a possible world,  $w_1$ , where the only metal found in the Sierra Nevada was fool's gold, which satisfies the description but which, nevertheless, is not gold (which was found elsewhere). At  $w_1$  (1) is false, while (2) is true. Thus (1) cannot express the same proposition as (2), since the truth value of (2) is distinct from that of (1) when assessed at different metaphysical possibilities. Since descriptivism is committed to saying that (1) is necessary, and the example shows that (1) is contingent (i.e. *not* necessary), descriptivism is committed to a contradiction: that (1) both is, and is not, necessary.

The epistemological argument notes that (2) is analytic, and hence its truth is knowable *a priori*. However, since (1) and (2) are synonymous, we should also be able to know (1) *a priori*. Were that true, then for the descriptivist, imagining gold *not* to have the property of being a yellow, heavy, malleable, ductile metal and found in the Sierra Nevada should be as difficult as imagining a bachelor who is not an unmarried man. But this is implausible, since the possibility that gold lacks any of these properties 'is not automatically precluded by reflection on the concepts involved' (Salmon 2005: 28). Hence (1) must be knowable only *a posteriori*. Since descriptivism is committed to saying that (1) is knowable *a*



*priori* (*mutatis mutandis* for similar sentences involving other natural kind terms and their descriptive meanings) it, again, appears to entail a contradiction.

The semantical argument is ‘concerned with the nonmodal question of denotation *simpliciter*’ (Salmon 2005: 30). So, if we (again) consider (1), descriptivism says that whatsoever satisfies the description ‘the yellow, heavy, malleable, ductile metal found in the Sierra Nevada’ *is* the denotation of the natural kind term ‘gold’. However, if, due to error, ignorance or fraud, the thing that *actually* satisfies the description is something else – so fool’s gold, but not gold, was found in the Sierra Nevada – to what do we refer with our natural kind term e.g. ‘gold’? According to descriptivism we are *committed* to referring to the fool’s gold with the term ‘gold’, because the denotation of our natural kind term is whatever satisfies the relevant description. The fact that fool’s gold *happens* to have these qualities, however, strikes Kripke as irrelevant to the reference of the natural kind term ‘gold’. That there is, or could be, fool’s gold with gold-like properties should have *no* bearing on the denotation of gold, and the fact that it does according to descriptivism shows that descriptivism is false.

## **5.2 Rigidity and Descriptiveness**

Kripke’s positive thesis, recall, avoids the objections said to beset the descriptivist by claiming that (a) natural kind terms are nondescriptiveness (they refer directly), and that (b) when we consider counterfactual scenarios we can *stipulate* that we are talking about just *this* kind, since natural kind terms, like ‘gold’, are rigid designators. So, although the denotation of ‘gold’ may have been fixed with the aid of a description such as that found in (1) above, the term ‘gold’ itself is nondescriptiveness and directly referential: its meaning just *is* the

referent. The consequence of Kripke's reference fixing story is that it does not matter whether gold is the yellow, heavy, malleable, ductile metal found in the Sierra Nevada or not. The properties that a kind of substance happens to possess have no bearing on the reference of the corresponding natural kind term. The nondescriptiveness of the natural kind term 'gold' allows us to rigidly refer to gold, and to denote samples in any possible world in which gold exists regardless of whether it has the properties we commonly associate with that kind.<sup>54</sup>

As Putnam noted, when a sample is ostensively defined, that definition comes with an 'empirical presupposition: that the body of liquid I am pointing to bears a certain sameness relation ... to most of the stuff I and other speakers in my linguistic community have... called "water" '(Putnam 1975: 225). Thus, while *introducing* a natural kind term like 'gold' requires only a passing acquaintance with a sample, it does require an empirical presupposition that anything else which is essentially similar to *this* sample can also be truly denoted using the term 'gold'. With natural kind terms behaving in this fashion we can (i) avoid the negative consequences of Kripke's three arguments, and (ii) do justice to the modal intuition we (apparently) have in Twin Earth-style cases.

One way to recast the nub of the problem with descriptivism is using the notion of rigid designation. The problem with descriptivism is that descriptivists are committed to the rigidity of the descriptions which they take to be the meanings of natural kind terms. Thus the denotation of a natural kind term across metaphysical possibilities is whatever satisfies the description that is the meaning of, and so is synonymous with, some particular natural kind term. According to Kripke (and Putnam) this raises the problem that the satisfier of the

---

<sup>54</sup> That is *not* to say that *only* terms that are nondescriptive can be rigid. The claim is rather that it is a consequence of being nondescriptive that terms such as proper names are rigid designators.

description might not be the natural kind we are attempting to denote (the conclusion of the modal argument); that the denotation of a natural kind term *at the actual world* might turn out, for whatever reason, to be something other than that which was intended (the conclusion of the semantical argument); and finally, since the natural kind term is synonymous with its associated description the truth of identity statements such as (1) are *a priori* when they should be *a posteriori* (the conclusion of the epistemological argument).

Taken together, Kripke's arguments form a *reductio ad absurdum* of the descriptivist position, at least when descriptivism uses the sorts of description found in (1). Kripke offers us, instead, the following modal theoretical identification sentence:

(3) If gold exists, then gold is the element with atomic number 79.

As we have already seen, according to Kripke, a sentence like (3) is a species of identity statement 'involving two rigid designators' (Kripke 1980: 140) which, given the necessity of identity, is a paradigmatic example 'of the necessary *a posteriori*' (*ibid.*). Schematically, if 'x' and 'y' are rigid designators (as 'gold' and 'the element with atomic number 79' are supposed to be) then they refer to the same things in all possible worlds; if the identity sentence ' $x = y$ ' is true in this world, because 'x' and 'y' have the same referent, then given that 'x' and 'y' refer to the same kind in all possible worlds, the truth of the identity statement guarantees its necessity.

I have argued, contra Salmon (2005), that theoretical identities, such as 'gold is the element with atomic number 79', represent discoveries made by empirical science during its

investigation into the properties of gold, contain a descriptive term, and form a second category of necessary *a posteriori* truth. The key passage in Kripke, recall, was:

[A] material object is (pure) gold if and only if the only element contained therein is that with atomic number 79. Here the ‘if and only if’ can be taken to be strict (necessary). In general, science attempts, by investigating basic structural traits, to find the nature, and thus the essence (in the philosophical sense) of the kind. (Kripke 1980: 138)

However, Kripke *does* claim that phrases like ‘the element with atomic number 79’ are rigid designators. But since the complex phrase ‘the element with atomic number 79’ is clearly a description of the substance gold, rather than a natural kind term, and so not obviously analogous to ‘gold’ or ‘silver’, then if ‘the element with atomic number 79’ is to be rigid despite being a description we need some account of why some, but not all, descriptions are rigid. After all, if Kripke’s objection to the descriptivist is, in part, a charge that their contingent descriptions are falsely treated as necessary, but it turns out that, by Kripke’s own admission, there are some descriptions that are rigid (and thus necessary), then there is no reason why the descriptivist cannot use those descriptions to give the meanings of natural kind terms, thereby deflecting two of Kripke’s three objections (the modal and semantical arguments). In the following section I will present what I take to be Kripke’s (severally neglected) account of rigid descriptions, and demonstrate how this gets us a new interpretation of Kripke on the necessary *a posteriori*.

### 5.3 De Facto Rigidity, Essentialism and Descriptivism.

Given that (3) contains a description, it is useful to recognise Donnellan's (original 1966) distinction between two types of *use* of definite descriptions: (i) *attributive*, and (ii) *referential*. When a description is used referentially it is 'merely one tool for doing a certain job – calling attention to a person or thing' (Donnellan 1966: 285), whereas when it is used attributively it 'states something about whoever or whatever is the so-and-so' (*ibid.*). When we consider (3) in light of Kripke's claim that science has *discovered* the essence of gold then it is clear that 'the element with atomic number 79' is being used attributively rather than referentially.<sup>55</sup> The description 'the element with atomic number 79' is *not* being used merely as a mark to pick out a certain kind of object, rather it is being used as a qualitative criterion for kind membership: anything *with* atomic number 79 *is* gold, anything *without* atomic number 79 is not gold. As Kripke says, having atomic number 79 is *the essence* of gold.

There is a clear semantic analogy between (1) and (3), since both contain a (general) name and an attributive description. But (3), unlike (1), contains two rigid designators, is necessarily true (rather than just true), avoids the epistemological argument, and does not give the *meaning* of 'gold'. But Kripke does not *explain* why a description such as that in (3) can be rigid given that the description in (1) apparently cannot. Nevertheless, we can construct an account based on other things that Kripke says. The availability of such an account is crucial to the argument in favour of descriptivism. If some descriptions can, on Kripke's view, be

---

<sup>55</sup> In the contemporary literature, and as shown in chapter 3 and 4, we can distinguish between Nathan Salmon's view and Scott Soames' view with precisely this distinction: Salmon (1982: 52 – 54) thinks 'the element with atomic number 79' is only used referentially (being completely analogous to a proper name) whereas Soames (2002: 278 – 279) thinks that 'the element with atomic number 79' *is* descriptive, and seems to intend this to be attributive.

rigid designators and avoid the modal and semantical objections, then any version of descriptivism that uses them will, *a fortiori*, also avoid these objections.

The fundamental difference between (3) and (1) is that (3) describes an *essential* property, whereas (1) describes an *accidental* property. Thus, if the descriptivist claimed that (3), rather than (1), gave the meaning of ‘gold’ they could avoid the modal argument: (3), if true, states the *essence* of the kind gold, and the essence of a kind is a property possessed by that kind in all possible worlds where that kind exists. Of course, this would still leave the descriptivist facing the epistemological and semantical arguments, which Kripke’s thesis avoids by presenting a semantic story according to which (a) the truth of (3) is *not* something we come to know *a priori* – it takes scientific investigation; and (b) no description gives the meaning of ‘gold’: the meaning of a natural kind term just is the referent. However, if (3) is making an essence claim, and descriptions of the essences of natural kinds turn out to be rigid, then the descriptivist can easily avoid the semantical argument: there is a description that is uniquely satisfied by all and only samples of a particular natural kind, in this case gold, and consequently the natural kind term gets to refer to all and only those things that satisfy the description of its essence, in this case ‘the element with 79 protons within its nucleus’. Hence, it is just implausible to suppose that a description of the essence of a kind could underdetermine reference to it. Moreover, it seems to me that Kripke provided just such an account of the rigidity of some kinds of description.

In the preface to the 1980 edition of *Naming and Necessity*, Kripke presents a distinction between two types of rigidity: *de jure* rigidity and *de facto* rigidity. According to Kripke a designator is rigid *de jure* when ‘the reference of a designator is *stipulated* to be a single object’ (Kripke 1980: 21 n21). Names like ‘Hesperus’ and natural kind terms like ‘gold’ are, for Kripke, rigid *de jure*. Designators that are *de facto* rigid, by contrast, are

descriptions that are rigid just when they happen ‘to use a predicate “*F*” that in each possible world is true of one and the same unique object’ (*ibid.*) – or, presumably, kind. This distinction has received only minimal treatment in the literature, but is central to Kripke’s account of theoretical identity.<sup>56</sup> Furthermore it also further highlights the role that Kripke’s essentialism is playing in his account of natural kind terms.

Kripke’s endorsement of essentialism has been well established in previous chapters. It is also provocatively illustrated by Kripke’s claim that ‘science attempts, by investigation basic structural traits, to find the nature, and thus the essence (in the philosophical sense) of the kind’ (Kripke 1980: 138), and his further claim that there can be ‘cases of essence and [cases] of identity between two rigid designators’ (Kripke 1980: 141). It is here that the distinction between *de jure* and *de facto* rigidity comes into play.

Kripke illustrates his notion of *de facto* rigidity with the following example. Consider the number 2, and the description ‘the smallest prime number’. According to Kripke the description ‘the smallest prime number’ rigidly designates the number 2 because *being the smallest prime number* is a property the number 2 has in all possible worlds i.e. a necessary property. Similarly LaPorte (2004: 46) argues that the description ‘the positive square root of 1600’ is a rigid designator that designates the number 40 since the number 40 *necessarily* has the property of *being the positive square root of 1600*. The obvious contrast between these two examples and (3) is that both of the examples above are knowable *a priori*, and arguably analytic. Given their apparent analyticity it is hardly surprising that the descriptions ‘the smallest prime number’ and ‘the positive square root of 60’ are rigid designators. But as

---

<sup>56</sup> More recently, the *de jure/de facto* distinction has been employed by Corrine Besson (2010) to carve out an asymmetry between natural kind terms and other general terms. However, here I use it for a completely different purpose.

Kripke so famously argued, the epistemological contrast between ‘2 is the smallest prime number’ and ‘if gold exists, then gold is the element with atomic number 79’ has no bearing on the metaphysical necessity of the identity/essence statement. Kripke grants that although ‘the contrast with mathematical cases could not be greater’ (Kripke 1980: 141) epistemologically, when we are considering cases such as (3), ‘to apply these [logical/epistemic] notions to genuine *de re* modalities is, from the present standpoint, perverse’ (Kripke 1980: 142). The key similarity between the cases is that they are both metaphysically necessary, and ‘any necessary truth, whether *a priori* or *a posteriori*, could not have turned out otherwise’ (Kripke 1980: 142).

The similarity between sentences like (3) and examples like ‘2 is the smallest prime number’ is that both are necessary. Moreover, since theoretical identities are essence claims, and typically contain descriptions of the essential properties of kinds, for instance ‘the element with atomic number 74’ describes tungsten’s essence, and given that Kripke claims they are rigid designators, there is a natural extension of the notion of *de facto* rigidity to the descriptions within theoretical identifications: after all, the descriptions in theoretical identities describe properties that the kind has in all possible worlds, and *de facto* rigid designators are rigid precisely because they refer to properties possessed by their bearers in all possible worlds.

The upshot is that the necessity of theoretical identifications is grounded in two types of rigidity. On the left-hand side of theoretical identity sentences there is a *de jure* rigid designator, that is a natural kind term like ‘gold’ that rigidly refers to the gold-kind by stipulation (proper names are rigid in just this way). On the right-hand side of many theoretical identities there is a *de facto* rigid designator, the rigidity of which is grounded in two non-semantic claims: (a) that gold is the element with atomic number 79, and (b) that



having atomic number 79 is *the essence* of gold. Clearly claim (b), as Salmon has noted on numerous occasions (see chapter 4), is a statement of an ‘irreducibly metaphysical, and philosophically controversial, theory of essentialism’ (Salmon 2005: 217). But if theoretical identity sentences like (3) are necessary because they contain a *de facto* rigid designator, and *de facto* rigidity is grounded in describing a property that is a necessary property of the kind, then, again, it is Kripke’s metaphysics which is doing the heavy lifting in his discussion of natural kinds: it is the essential (or metaphysically necessary) status of certain properties that allows descriptions to be rigid designators, and it is precisely this type of rigid designator we find in a theoretical identity sentence which is supposedly, when true, necessary.

It is worth noting here that this strategy for identifying Kripke’s essentialist commitments is quite different to Salmon’s. Whilst Salmon identifies an essentialist premise in the Kripke-Putnam derivation of necessary *a posteriori* truths (see chapter 4), identifying it as a thesis about substances and consubstantiality, here the strategy is quite different. Here I have simply observed that typically one of the terms within a theoretical identity sentence (at least the simple-complex type) is an attributive description, and that Kripke claims that such terms are rigid (although he does not identify them as descriptions). The only account available from Kripke as to why descriptions of essential properties are rigid designators was the notion of *de facto* rigidity, which we found to amount to the claim that where an object has a certain property necessarily, an attributive description of that property will refer to the bearer of that property in all possible worlds.

Again, the parallel between theoretical identities and the necessity of origin is worth drawing. Let *O* be the origin of Winston Churchill, where *O* means roughly coming from the sperm of Randolph Churchill and the ovum of Jennie Jerome, and consider the necessity of origin claim ‘Winston Churchill is the man with origin *O*’. The name ‘Winston Churchill’ is a

*de jure* rigid designator, introduced by ostension, and stipulated to apply to just the individual ostended at the original baptismal event. The description ‘man with origin *O*’, however, describes the origin of Winston Churchill, rather than naming him. However, necessity of origin claims are supposed to be necessary when true. According to Kripke, individuals necessarily have the origins they actually have; in other words, Winston Churchill exists in all and only worlds where someone with the origin *O* exists. Hence, the description ‘the man with origin *O*’ will precisely mirror the rigidity of the name ‘Winston Churchill’ except that the description is generated by the empirical investigation of Churchill’s actual origin, plus the *a priori* thesis that the origin of an individual is necessary. Thus the description ‘the man with origin *O*’ looks to be a *de facto* rigid designator, paralleling my previous treatment of theoretical identities.

#### **5.4 Descriptivist Essentialism**

The forgoing analysis, particularly the account of rigid descriptions and the commitment to essentialism, invites the following sketch of a version of descriptivism that combines essentialism with descriptivism (rather than the causal theory of direct reference): call it *Descriptivist Essentialism*. This view, I will argue, can avoid all three of Kripke’s objections outlined at the beginning of this chapter, and presented in detail in chapter 1: the modal and semantical objections can both be avoided by an application of metaphysical essentialism, and the epistemological argument can be avoided by presenting a new story about how the reference of natural kind terms is fixed, and employing a strategy developed by Soames for his Extended Millianism.

Let me very roughly define metaphysical essentialism as the view that what makes a particular object a member of a natural kind is that it possesses the relevant underlying nontrivial essence (contrast the trivial property of water, *being water* with the property *being composed of two hydrogen atoms and an oxygen atom*, which requires a theory about the essences of substances, and thus is nontrivial).<sup>57</sup> The essences of natural kinds are properties which are both necessary and sufficient for being instances of those particular natural kinds; they are ‘properties which nothing can lack and still be of the kind’ (Mellor 1977: 299). Furthermore, ‘things of a kind have [their] essential properties in all possible worlds’ (*ibid.*), which is to say that the essential properties of natural kinds hold as a matter of metaphysical necessity.<sup>58</sup>

Metaphysical essentialism is not a consequence of any semantic thesis (cf. Salmon 2005), but an independent metaphysical position. As such there is nothing *prima facie* inconsistent about the combination of descriptivism and essentialism. The descriptivist essentialist view states that the only descriptions that give the meanings of natural kind terms are those that describe the essences of those kinds. In relation to our earlier examples we should say that (3) gives the meaning of ‘gold’, whereas (1) does not. With the meaning of ‘gold’ fixed by (3), and the essence of gold being *the element with atomic number 79*, the descriptivist essentialist can quickly avoid the modal and semantical arguments.

Since descriptions of the essences of natural kinds are rigid *de facto*, anything satisfying that description can be referred to via the associated natural kind term, both in this

---

<sup>57</sup> I will define and then discuss the distinction between trivial and nontrivial essentialism in chapter 7.

<sup>58</sup> For all I have said here, it still might be that essentialism is nontrivial without also being a substantive metaphysical thesis. In chapter 7 I will analyse precisely what a substantive version of essentialism is, and distinguish between substantive and non-substantive versions of essentialism, before endorsing a non-substantive version in chapter 8.

world and in all possible worlds.<sup>59</sup> On Kripke's view (3) is (if true) necessary. The truth of (3) depends upon some relevant scientific discovery, and the necessity is a product of the name 'gold' and the phrase 'the element with atomic number 79' both being rigid designators. If 'the element with atomic number 79' is rigid *de facto* then its rigidity is a consequence of the property *being the element with atomic number 79* being the essence of gold. Kripke's criticism of the descriptivist was that if they take a description to be the meaning of a natural kind term, as in (1), they are committed the necessity of (1), given that (1) expresses the same proposition as (2). The objection was that in (1) the description associated with the natural kind term is at best contingently true of it. Thus there are possible scenarios in which (1) is false, and so the description 'the yellow, heavy, malleable, ductile metal found in the Sierra Nevada' cannot be the meaning of 'gold'. But (3) is, by Kripke's own lights, necessary when true. That necessity, I have argued, is grounded in the assumption that kinds have essences. A descriptivist essentialist endorses (3), but denies (1), as an account of the meaning of 'gold': the only descriptions that give the meanings of natural kind terms are the essentialist ones.<sup>60</sup> As the members of natural kinds have their essential properties as a matter of metaphysical necessity there will be no scenario in which (3) could be false. Thus the descriptivist's commitment to the necessity of sentences that give the meanings of natural kind terms, like (3), is no longer a vice. So the modal argument does not threaten descriptivist essentialism.

The semantical argument also fails to affect the descriptivist essentialist. The semantical argument was concerned with denotation *simpliciter*, and asked us to consider a

---

<sup>59</sup> In a recent paper Jasper Reid (2002) has argued that stability of reference is one of the key motivations for being a natural kind essentialist, although he is, at the time, endorsing a Kripkean picture of reference, rather than the descriptivist version being defended here.

<sup>60</sup> It is worth flagging that this sort of proposal is inconsistent with the sorts of suggestion made by Papineau (1996), who claims that theoretical definitions are imprecise.

situation where something else satisfies the description we associate with the natural kind term. Kripke's argument was that it is *irrelevant* to the reference of a natural kind term whether (or not) it happens to have some property associated with it. But given descriptivist essentialism the scenario we are considering is one where the description 'the element with atomic number 79' denotes something else. If the semantical argument applies to *any* description then Kripke should say that it is irrelevant to the reference of the natural kind term 'gold' what happens to have the property of *being the element with atomic number 79*. But since that property is essential, the objection looks implausible. Anything satisfying the description 'the element with atomic number 79' just *is* gold, both on Kripke's view and on the descriptivist essentialist's view. Anything with the property *being the element with atomic number 79* possesses the essence of gold, and thus is gold. The semantical argument, then, does not threaten descriptivist essentialism, as essential properties *are* relevant to the denotation of natural kind terms.

### **5.5 Locating Meaning**

This leaves the epistemological argument. If the descriptivist essentialist rejects (1) in favour of (3) as giving the meaning of 'gold', and uses their essentialism to navigate around the modal and semantical arguments, they must still deal with the counterintuitive charge that the truth of (3) is something a speaker can know *a priori*. Recent commentators (e.g. Mackie 2006) have argued that 'essentialism about natural kinds will be inconsistent with any description theory' (Mackie 2006: 176), citing the modal and epistemological arguments as the cause of that inconsistency. On the descriptivist essentialist view presented here it is only descriptions of the essences of natural kinds that can give the meaning of natural kind terms.

So, *ex hypothesi*, the description that gives the meaning of the natural kind term cannot come apart from the essence of the natural kind. But the epistemological argument can still be generated. To illustrate, if the descriptivist essentialist accepts (3), then they will undoubtedly accept (4):

- (4) If the element with atomic number 79 exists, then the element with atomic number 79 is the element with atomic number 79.

As in the initial substitution made to illustrate Kripke's three arguments against the descriptivist (see section 5.1 this chapter, and Chapter 1), if the meaning of 'gold' is 'the element with atomic number 79' then (3) is synonymous with (4). Since (4) is analytic and thus knowable *a priori*, so too is (3). The remaining challenge for the descriptivist essentialist is to accept that sentences like (3) are, in some sense, knowable *a priori* without this seeming implausible. However, it should again be noted that Soames' Extended Millianism implies that identities and some theoretical identities are *a priori*, and seemingly Salmon is also committed to something like this view. Similarly, although Kripke's stops short of giving a theory of meaning, if he endorses Extended Millianism then he too is committed to the *a priori* of identities and some theoretical identities. The first element of the descriptivist essentialist's strategy for providing just such an account is to adopt Putnam's account of the 'division of linguistic labour' (*DLL*), properly understood (Putnam 1975: 227).

The basic hypothesis of (*DLL*) is that within our linguistic community there are certain speakers who are *experts* in, say, chemistry. These experts form a sub-community that are able to recognise instances of some particular (chemical) kind, where the majority of us are

unable to do so. The ‘necessary and sufficient conditions for membership in the extension, ways of recognizing if something is in the extension (‘criteria’), etc. – are all present in the linguistic community’ (Putnam 1975: 228), but need not be known by *every* member of that community. This general division of labour engenders a further division of *linguistic* labour, where ‘[t]he features that are generally thought to be present in connection with a general name’ (Putnam 1973: 705) are known only by the expert speakers. (*DLL*) thus explains why not everyone who uses a natural kind term, and whom we consider to be a competent language user, needs to be able to reliably distinguish samples of gold from samples of some other closely resembling substance such as fool’s gold. There is, says Putnam, a ‘structured cooperation’ (Putnam 1975: 228) between ordinary and expert speakers such that for an ordinary speaker their use of a particular natural kind term *depends* upon this cooperation. Putnam takes it to be a consequence of (*DLL*) that a descriptivist theory of reference and meaning is false, as (*DLL*) shows that for the majority of speakers their use of a natural kind term is *not* determined by a concept they have, but by the experts within their community.

The division of linguistic labour does not itself warrant an inference to Putnam’s theory of reference (cf. Mellor 1977; Zemach 1976). Putnam’s intention is that the structured cooperation between non-experts and experts leads back to instances of the natural kinds themselves: for Putnam, the meaning of a natural kind term just *is* the referent, but since not all of the language users can identify the referents of their natural kind terms it looks like not all speakers know the meanings of their natural kind terms. Putnam takes (*DLL*) to provide an account of how all the language users within a community can know the meaning, in the very minimal sense that they are construed as competent users, of their natural kind terms by postulating someone within the community of which the language user is a member, who is able to distinguish samples of gold from samples of some similar substance, and knows the

conditions that are necessary and sufficient for membership into the kind. However, this is by no means the only way in which a division of linguistic labour might work. For example, it is plausible to suppose, as Mellor does, that the division of linguistic labour leads back to ‘our *experts’* beliefs’ (Mellor 1977: 304), rather than to objects within the environment that only experts can individuate. In such cases there is still a division of linguistic labour – we are still, as non-experts, deferring to what the experts say – but we defer to the experts’ *beliefs* rather than to their ability to recognise real samples of gold (although there seems to me to be no reason why the two cannot be combined).

With this in mind, consider (5):

(5) If tungsten exists, then tungsten is the element with atomic number 74.

Given *DLL* we can say that the majority of normal speakers defer to the small cadre of chemical experts in their community when they use the term ‘tungsten’. Hence Sally, our typical language user, intends to defer to the scientist when she uses ‘tungsten’. Let’s say that what the scientist has in mind, when she uses the word ‘tungsten’, is the description ‘the element with atomic number 74’. In other words, when asked what ‘tungsten’ means (or indeed what tungsten is), the expert asserts (5). Since the majority of the linguistic community defer to the experts when it comes to the meaning of ‘tungsten’ when (a) they use the word ‘tungsten’, and (b) are asked what ‘tungsten’ means, we can say that *most* speakers will not assert (5) because they lack the empirical knowledge that would allow them to do so.



For most speakers, then, sentence (5) is not known, because ‘tungsten’ means whatever the experts say it means, and a normal speaker need not know that experts identify tungsten as the element with atomic number 74. They could, of course, come to *know* what the experts say ‘tungsten’ means, by asking (or becoming) an expert: so Sally consults a scientist and she tells Sally that ‘tungsten’ means ‘the element with atomic number 74’. However, since Sally needed to ask the scientist what ‘tungsten’ meant before she could assert that (5) was true, Sally’s justification for thinking that (5) is true, which generalises to all normal speakers, is that an expert told her so. In effect, although the proposition expressed by (5) is knowable *a priori*, it is not known by the majority of the linguistic community who use natural kind terms deferentially. The application of *DLL* to descriptivist essentialism means that Kripke’s epistemological argument – that the truth of an identity sentence is knowable *a priori* – still applies. However, drawing a decisive distinction between the fact that the proposition expressed by (5) is knowable *a priori* on the one hand, and the question of whether or not Sally herself knows the proposition *a priori* on the other, allow us to avoid at least some of the force of Kripke’s argument. Although it is a consequence of descriptivist essentialism that sentences like (3) and (5) express propositions that are knowable *a priori*, this is a distinct question from whether their truth is known *a priori* by non-expert language users. An application of Putnam’s *DLL* allows us to give an affirmative answer to the former and a negative answer to the latter.

The truth of (5) for the expert, however, is both knowable and known *a priori*, according to descriptivist essentialism. The epistemological objection, then, still applies to experts as Kripke intended: an expert who *knows* the essential description of tungsten knows the truth of (5) *a priori*. But if the expert can know the truth of (5) *a priori* then reflection on the meaning of ‘tungsten’ *should* have allowed the expert to derive the description ‘the

element with atomic number 74' in much the same way as reflection on the meaning of 'bachelor' allows a competent user to derive the description 'an unmarried man'. An application of Kripke's epistemological objection commits the descriptivist to the *a priori* knowability of (5), but the thought that we do not take *not* having atomic number 74 to be ruled out *a priori* of anything to which 'tungsten' truly applies is standardly thought to show why the descriptivist thesis is obviously false. However, I shall argue that the *a priori* of propositions such as that expressed in (5) can be made both plausible and palatable if we provide a new account of how substances are named within the expert community. First, however, it is worth considering the strategy deployed by Soames (2002) to show that although Extended Millianism is committed to the *a priori* of identities and theoretical identities (just as descriptivist essentialism is!) containing co-referring names, this consequences does not constitute a *reductio* of the position. My intention is to employ an analogue of Soames' strategy here, and then proceed to offer some independent justification for it. The key point, however, is that if the *a priori* of identity sentences does not constitute a *reductio* of (EM), then nor should it constitute a *reductio* of descriptivist essentialism.

### **5.5 *A priori* knowability**

According to Soames' (EM), the meaning of a proper names just is its referent. Hence, the proposition expressed by the identity sentence 'Harry Houdini is Harry Houdini' is that same as that expressed by 'Harry Houdini is Ehrich Weiss'. Identity sentences between co-referential names, then, are 'both necessary and knowable *a priori*' (Soames 2002: 237). Thus, whilst Soames does think there is a category of necessary *a posteriori* truth 'identity

sentences involving co-referential, ordinary, linguistically simple proper names are not genuine instances of it' (*ibid.*: 328).

To rescue the Kripkean claim that many identities (and by extension theoretical identities) are *a posteriori*, Soames distinguishes sharply between the proposition expressed by an identity sentence like 'Harry Houdini is Ehrich Weiss', and the information a speaker intends to communicate when they assert it, what we might call the 'assertive utterance'.<sup>61</sup> Thus, when a speaker utters 'Harry Houdini is Ehrich Weiss' what they intend to convey is *not* the proposition expressed by 'Harry Houdini is Harry Houdini', but some contingent, *a posteriori*, descriptive information which they idiosyncratically associate with each name respectively. Imagine, for instance, that the speaker is Ehrich's neighbour, who discovers that Ehrich is in fact a world famous magician. Upon reporting this discovery to his wife he asserts 'Ehrich Weiss is Harry Houdini', and intends to communicate something like 'Ehrich next-door is a (particular) famous magician'.

Although Soames does not employ this strategy for simple-complex theoretical identities (although, presumably, he thinks that this strategy applies to the simple-simple category of theoretical identity), I propose that we borrow this strategy, and apply it to descriptive natural kind terms.<sup>62</sup> So, when a non-expert speaker utters 'tungsten is the element with atomic number 74', whilst the proposition expressed is knowable *a priori*, the information they are trying to communicate – for instance, that it is the steel-gray coloured, brittle metal which has 74 protons within the nucleus of each constituent atom – is *a*

---

<sup>61</sup> This solution, note, bears a significant resemblance to Michael Dummett's (1991) strategy for a descriptivist account of natural kind terms, where we distinguish the *assertoric content* (what Soames refers to as the proposition) from the *ingredient sense* (what Soames calls the assertive utterance).

<sup>62</sup> In chapter 3, section 4 I proposed this account on Soames' behalf for cases like the 'Woodchucks are Groundhogs' case.

*posteriori* and contingent. The falsity of ‘tungsten is the element with atomic number 79’, then, cannot be ruled out *a priori* when the content of the assertive utterance such a sentence expresses is ‘the steel-gray coloured, brittle metal is made of atoms which have 74 protons within their nuclei’. Moreover, since speakers are standardly concerned with content of the latter sort, rather than the former, our intuitions about *a posteriority* should come as no surprise. Nevertheless, if a commitment to the *a priority* of identity sentences for (EM) does not constitute a *reductio* of (EM), nor does it to the descriptivist position on natural kind terms. In other words, if one is willing to buy Soames’ story about proper names, then there is no reason why one cannot buy the descriptivist story about natural kind terms. Furthermore, either both accounts have an advantage over the initial Kripkean story insofar as they actually offer a theory of meaning, or (EM) is the account of meaning that Kripke endorses, and so everyone is in the same boat.

## 5.6 The metasemantics of natural kind terms

Once the *reductio* has been blocked, we need to make the *a priority* of theoretical identities independently plausible. The main problem, of course, is that since Kripke the standard view of essences has been that they are discovered by science, and are thus firmly in the realm of the *a posteriori*. If theoretical identities are statements of essence, the surely they are *a posteriori*?

This question, however, is misleading. Whilst the *discoveries* of science are certainly *a posteriori*, scientists’ *definitions* of their terms need not be. With this in mind, the basic tenet of my new metasemantics for natural kind terms is that the community of experts adopts the practice of naming *new* substances on the basis of discovering things that satisfy *new*

descriptions. Thus it is only when an essence description is found to have a satisfier that a name is introduced.

Consider the synthesis in 2006 of an element with 118 protons in its nucleus (Oganessian 2006: 889 - 904).<sup>63</sup> At the time of this discovery, the *New York Times* reported that the element with 118 protons in its nucleus (or, more simply, the element with atomic number 118) would be known provisionally as ‘element 118’ (for obvious reasons) until such time as it was confirmed by another independent laboratory.<sup>64</sup> The element, to the best of my knowledge, has yet to be confirmed, and thus is still referred to as ‘element 118’. In terms of the semantics, it looks like we have a description (‘the element with atomic number 118’) and a *proto*-name (‘element 118’) that has been clearly derived from the description. Since we are uncertain whether anything satisfies that description, we do not introduce a proper name.<sup>65</sup> As things stand, then, an expert might assert the following sentence:

(6) If element 118 exists, then element 118 is the element with atomic number 118.

---

<sup>63</sup> I will talk more about the synthesis of elements and natural kinds in Chapter 7, when I criticise Ellis’s metaphysical essentialism.

<sup>64</sup> I have, here, omitted some details which do not affect the argument: there are systematic element names which use a very specific set of criteria for generating names (cf. Connelly *et al* 2005). Using these criteria one can generate the name ‘Ununoctium’ for what I call ‘element 118’. However, anyone who knows the grammatical rules of the system can derive ‘element 118’ from the name ‘Ununoctium’, and hence sentences such as ‘if Ununoctium exists, then Ununoctium is the element with atomic number 118’ are knowable *a priori*.

<sup>65</sup> Presumably because our cadre of experts have yet to agree whether the evidence warrants thinking that that element 118 exists, rather than its being the product of fallacious (or fabricated) data; in effect, they are unsure whether any *actual* thing does satisfy the description.

The truth of (6) is uncontroversially knowable *a priori* (at least for anyone with basic scientific knowledge). So what changes when we substitute ‘element 118’ for some name or other? On the assumption that element 118 *will* be confirmed (that the description ‘the element with atomic number 118’ has a satisfier) the Russian scientists who discovered the element with atomic number 118 have suggested the names ‘Fyorium’ and ‘Moskovium’. Let’s say, for the sake of argument, that the name ‘Moskovium’ is chosen. We can then produce (7):

(7) If Moskovium exists, then Moskovium is the element with atomic number 118.

Now (7) certainly has the appearance of *a posteriority*, being the analogue of (3) and (5), but only, I claim, for the non-expert members of the linguistic community. The cadre of experts, when *they* name element 118 on the basis of discovering something that satisfies the description ‘the element with atomic number 118’, will know the truth of (7) *a priori*. After all, all they have done is substitute ‘Moskovium’ for ‘element 118’. For the expert at least, meanings are firmly in the head.

What should make this thesis palatable is the metasemantic story. According to descriptivist essentialism a name is not *generated* until such time as something has been found that satisfies the description ‘the element with atomic number 118’ (reflecting actual practice). Thus although the truth of the proposition expressed by (7) is knowable *a priori*, the discovery that something satisfied the description *in* (7) was *a posteriori*. The reason that the truth of (7) is knowable *a priori*, however, is that new substances are not named *until* the

relevant description is judged to have been satisfied. In effect, if it turns out that the provisional name in (6) denotes something in the actual world we introduce a new (proper) name, as in (7). Since (6) and (7) express the same proposition, (7), like (6), will be knowable *a priori*. The important point, however, is that if the description in (6) is not judged to be satisfied then (7) is never generated. So while the truth of (6) is knowable *a priori*, the naming conventions preclude the generation of (7) until something in the actual world has been found to satisfy the description in (6). Unless a substance with the property of *being the element with atomic number 118* is found, no proper name is introduced for such a substance, because no such substance is known to exist.

This account of the naming of substances does not construe the *discoveries* of science as *a priori*. It is certainly an *a posteriori* matter whether a certain description has satisfiers in the actual world. For example, it is an empirical matter whether anything satisfies the description ‘the element with atomic number 180’. However, even prior to any empirical investigation I can assert that ‘if element 180 exists, then element 180 is the element with atomic number 180’, and know the truth of my assertion *a priori*. The discovery that something satisfies the description ‘the element with atomic number 180’ is a matter for science and empirical investigation. But although this empirical investigation has a bearing on *whether* we introduce a name, on the assumption that once such a discovery is made the substitution of a proper (general) name, say ‘Birmingham’, for the placeholder or *proto*-name ‘element 180’, has no impact on the epistemic status of the proposition expressed by the sentences ‘if element 180 exists, then element 180 is the element with atomic number 180’, and ‘if Birmingham exists, then Birmingham is the element with atomic number 180’. The question of whether the description ‘the element with atomic number 180’ has denotation is *separate* from the question of whether ‘if element 180 exists, then element 180 is the element

with atomic number 180' is true. Answers to the former are *a posteriori*, but answers to the latter are *a priori*.

### 5.7 Intuitive Kinds vs. Theoretical Kinds

One obvious objection to the view defended here is that there are some clear counterexamples. Terms like 'water' or 'gold', for instance, had significant pre-scientific currency. Hence, it seems plainly false to suppose that they could have been introduced systematically as outlined above, i.e. science uncovers the essence of a new kind of substance, and introduces a name which communicates that defining information. There are two things to be said in defence of descriptivist essentialism here.

First, an observation. Whilst it may be true that some of the more pedestrian natural kind terms such 'water' have a significant pre-scientific currency, these terms are in the minority when it comes to the names for natural substance-kinds. In contrast to the names for compounds such as 'water' with which non-expert users are familiar, there are a plethora of natural kind terms for substances, such as 'fluoridoargon', 'cyanobromane' and 'hydroxidotrioxidoiodine', of which most of the linguistic community will never have heard, let alone know the meaning. These natural kind terms, as in the version of descriptivism I have expounded, are descriptive (at least according to the International Union of Pure and Applied Chemistry) in that their formulation conveys both structural and compositional information to experts in possession of the relevant grammatical rules (cf. Connelly *et al.* 2005). So, a chemist will know the truth of 'hydroxidotrioxidoiodine is  $\text{HIO}_4$ ' *a priori* because of the systematic rules governing the formulation of compound-kind names. This, however, will tell us nothing about whether or not 'hydroxidotrioxidoiodine' actually refers to



anything – in theory chemists could concoct a chemical formula and use the grammatical rules to derive a compound-kind name, and thereby formulate a theoretical identification sentence like the one above without doing any empirical investigation; in such cases it would be wrong to claim that they do not know the truth of the theoretical identity *a priori*, despite the fact that they may not know whether anything actually has the chemical composition  $\text{HIO}_4$ .

Secondly, a direct response to the objection. The descriptivist answer is to adopt elements of a strategy recently defended by LaPorte (2004, 2010), and claim that pre-scientific natural kind terms, such as ‘water’, do not mean the same thing as scientific natural kind terms. The crux of LaPorte’s thesis is that pre-scientific uses of natural kind substance terms are vague, or what he calls ‘open texture[d]’ (2004: 97), and that there are various criteria associated with the pre-scientific kind term that could figure in the determination of its extension. Those criteria, whatever they were, certainly were *not* the same as those of the modern chemist. More importantly, it is not obvious that they – pre-scientific language users, that is – were aiming at carving out *natural* kinds; as Dupré observes, typically ‘the kinds of ordinary language do not coincide with those of a scientific taxonomy’ since the kinds of ordinary language are introduced with ‘a variety of grounds for classification’ in mind (Dupré 1995: 36).

This suggests the following general strategy for dealing with terms like ‘water’ and ‘steel’, that have significant pre-scientific currency. Following LaPorte, the descriptivist essentialist claims that the pre-scientific use of words like ‘water’ were vague, and further, following Dupré, that the pre-scientific term ‘water’ may well have been introduced with a different set of classificatory interests (although these, too, might have been vague) to those of the modern chemist. Contemporary chemistry, as has already been stated, tells us that

chemical composition (the conjunction of the constituent atoms, and the molecular arrangement of those atoms) is the decisive, defining feature of substances, and as such that scientific substance terms latch onto kinds delimited by those properties.

In contrast, the classifications of ordinary language users, especially pre-scientific language users, were more likely informed by quite different interests. For instance, it *might* have been that the surface properties of objects formed the basis of a pre-scientific language user's classifications of substance kinds. LaPorte, in his (2004), invites us to draw just this sort of conclusion with regard to the kind 'jade'. According to LaPorte, despite discovering that the extension of 'jade' includes two micro-structurally distinct substances – jadeite, with the composition  $\text{NaAl}(\text{SiO}_3)_2$ , and nephrite, with the composition  $\text{Ca}(\text{Mg,Fe})_5\text{Si}_8\text{O}_{22}(\text{OH})_2$  – language users did *not* revise their use of the term 'jade' in light of this discovery, concluding that that the word 'jade' could 'apply clearly to jadeite as well as to nephrite' (LaPorte 2004: 96). Hence, the considerations of modern chemistry are not always decisive for the reference of substance terms, and the example illustrates that our classificatory interests concerning kinds sometimes diverge from the scientists'. As LaPorte sees it, the Kripke/Putnam intuition is that 'microstructure trumps superficial properties in determining reference' (*ibid.*: 98), as articulated in Twin Earth style thought experiments, are false. The jade case 'presents a historical *counterexample* to Putnam's Twin Earth lesson' (*ibid.*: 94) since speakers retained the use of 'jade' to pick out a kind of substance, despite discovering that nephrite and jadeite were compositionally distinct.

Whilst LaPorte views the jade example as to be a counterexample to Putnam's Twin Earth thought experiment, the descriptivist essentialist uses it merely to confirm the earlier hypothesis that ordinary language users are not always interested in the terms *for* the natural joints in nature: if we accept that science is interested in the joints in nature, that it has

discovered that ‘jade’ does *not* pick out such a joint, but that ‘jadeite’ and ‘nephrite’ do, and that despite this discovery ordinary language users have persisted in using the term ‘jade’ to denote a kind, then it is clear that the interests of ordinary language users and scientists *can* and *do* come apart. As Segal (2005, Ch. 2) has noted, outside of science, ordinary language users do not have a good grasp of what natural kinds are, and of which objects of their acquaintance fall within which natural kinds.

Zemach’s (1976: 119-20) illuminating discussion of Putnam’s account of substance terms, for example, distinguishes between the considerations of the normal language user and the expert within the community, pointing out that the normal language users’ use of ‘substance words is not determined by chemical considerations’ (Zemach 1976: 119). This, says Zemach, means that the stuff denoted with an English substance-word, say ‘water’, may not have a common nature (indeed, he claims that ‘[a]s a matter of fact, “water” *is* used in this way’ (Zemach 1976: 120), although this seems more controversial), being unified by the kinds of consideration LaPorte outlines in the jade example. But if *bona fide* natural kind terms are those that pick out natural kinds demarcated by their essences, and it turns out that there is no essence unifying the stuff we have been denoting with a general term, then it looks like we should say that that general term was never a natural kind term (although we might have thought that it was).

Take the kind denoted by the term ‘water’. As Zemach notes, we have eighteen isotopic varieties of a single water molecule given the three isotopes of hydrogen, protium (H), deuterium (D), and tritium (T), and the three stable isotopes of oxygen O<sup>16</sup>, O<sup>17</sup>, and O<sup>18</sup>. These variations constitute different kinds of water (in some sense), and since they could all occur in a single sample of water in various proportions, there looks to be ‘no chemical constitution common to all bodies of water’ (1976: 120). The chemist, unlike a typical

language user, does distinguish between  $H_2O$ ,  $D_2O$  and  $T_2O$  in the laboratory, referring to them (respectively) as ‘diprotium oxide’, ‘dideuterium oxide’ and ‘ditritium oxide’, rather than as ‘water’.<sup>66</sup>

Since we are interested in natural kind terms, rather than non-natural kind terms or practical kind terms, and examples like ‘jade’, and perhaps even ‘water’, provide good grounds for thinking that the taxonomy of ordinary and pre-scientific language users are informed by a variety of interests, we should *not* be interested in the kind terms of the ordinary language user. Rather, since we are interested in the terms that designate natural kinds, and are confident that it is the scientists who discover nature’s joints, it is the semantics of *those* terms with which we should be concerned.

So, given the vagueness of pre-scientific kind terms, the often different interests of ordinary language users, and the additional confidence we have in modern science (that it has discovered at least some of nature’s joints), the focus of this discussion appropriately targets chemical nomenclature – the branch of chemistry dedicated to the naming and classification of substances. The question we need to ask is: what does chemistry have to say about the use of what we might call ‘local names’ – that is, names with significant pre-scientific currency – for classifying substances? According to the most recent guidelines from the International Union of Pure and Applied Chemistry (IUPAC):

---

<sup>66</sup> Although presumably the chemist’s usage will depend on what she is doing: she may ask for dideuterium oxide (or heavy water) if she is working with a nuclear reactor, where deuterium absorbs less neutrons, allowing the use of unenriched uranium; however she may just ask for water if she is making a cup of tea.

Local names do not necessarily convey structural and compositional information to a wider audience. To be widely useful, a nomenclature system must be recognisable, unambiguous, and general; the unnecessary use of local names and abbreviations in formal scientific language should therefore be discouraged. (Connelly *et al* 2005: 3)

Despite discouragement, scientists and ordinary language users alike persist in their use of local names – terms like ‘water’ – which do not convey the structural and compositional information indicative of the relevant kind. But if the goal is an unambiguous system of classification, why persist in using these local names?

The explanation is that certain kind terms have significant natural language currency, and that scientists interested in naming are sensitive to the standard usage of certain general terms by the linguistic community. As IUPAC’s most recent guidelines on nomenclature state, the ‘public needs and common usage must also be borne in mind’ (Connelly *et al.* 2005: 3) when naming substances. Since terms like ‘water’ and ‘steel’ have such significant natural language currency, scientists have decided to retain these terms, rather than introduce new systematically generated names in light of discoveries concerning the defining characteristics of such substances.

Nevertheless, and vitally for the version of descriptivism defended here, scientists also stipulate that when local names are retained, experts must come to understand, and associate with the relevant term, ‘the devices used for [the] identification’ (*ibid.*) of the substance in question. When the term is systematically introduced, as, for instance, ‘hydroxidotrioxidoiodine’ was, these conditions of application are associated with the name *a priori*, using the grammatical rules of the nomenclature system. However, since terms like

‘water’ do not conform to this grammar, it is not obvious that those conditions are associated with the term *a priori*. But, since the scientist is required to associate the conditions for the identification of the substance with that term *as if* it conformed to the proper grammar, the suggestion here is that the scientist’s natural kind term ‘water’ has come to mean something substantially different from the pre-scientific language user’s term ‘water’.

The key claim is that scientific natural kind terms like ‘water’ are homonyms – that is, terms with the same spelling but different meanings – of their pre-scientific counterparts. The name ‘water’, as it is used by the scientist, has the same meaning as ‘dihydrogen oxide’, and since the theoretical identity ‘dihydrogen oxide is H<sub>2</sub>O’ is something the scientists can know to be true *a priori*, so too is ‘water is H<sub>2</sub>O’ knowable *a priori*. This, of course, does not imply that scientists discovered *a priori* that there was a substance with the chemical composition H<sub>2</sub>O. The key move, remember, was to separate the question of whether a natural kind *has* denotation from the question of *how* we come to know the truth of theoretical identifications. Successfully isolated, an *a posteriori* answer to the former need not imply an *a posteriori* answer to the latter. However, the descriptivist essentialist’s metasemantics ensures that a natural kind term is not introduced *until* the scientist is certain that a substance has been discovered. In cases where the substance was ostensibly discovered *before* science, the term used to denote it is re-coined (if it has significant natural language currency) where, and when, modern science confirms that there is a unique natural kind to be denoted.

Of course, this does not constitute an account of what normal speakers mean when they use natural kind terms. In a sense, this question is irrelevant to the thesis articulated here: it has already been assumed that it is science that uncovers the real natural kinds, and that as such it is the semantics of scientific terms we should be interested in. However, putting this to one side, the descriptivist essentialist can, following Mellor (1977), simply adopt Putnam’s

socio-linguistic hypothesis, the ‘*division of linguistic labour*’ (1973: 704) in order to give the account wider application.

The basic idea of (*DLL*), as discussed earlier, is that there is a structured cooperation between members of a linguistic community and an expert sub-community, where it is only that sub-community that know what the criteria associated with a particular kind term are. In the Putnam story, this expertise amounts to the ability to successfully discriminate between samples of the kind, and samples of some other substance that closely resembles the kind, for instance gold and fool’s gold. To employ this general strategy, all the descriptivist essentialist need do is maintain that although there is a division of labour, it is a ‘Fregean labour for all that’ (Mellor 1977: 304), since the division leads back to a description that the expert speaker has in mind, and associates *a priori* with the relevant term. Thus the extension of a natural kinds term is still a ‘Fregean function of our *experts*’ beliefs’ (1977: 304), rather than being fixed via the standard Kripkean metasemantic story, and the practical expertise of the scientist, provided the expert knows the description that uniquely identifies the kind *a priori*. And this is precisely what the metasemantic story presented here allows the descriptivist essentialist to say: natural kind terms are *not* introduced until (or, in cases like ‘water’, are re-coined when) the essence of a hypothesised substance is confirmed. When it is confirmed (that such a substance exists) the name introduced (or re-coined) adheres to particular set of grammatical rules that ensure the name communicates information about the kind’s essence; hence, experts who are competent with the grammar know the essence of the substance *a priori* when they come to understand the meaning of the term. Normal language users, by entering into a structured cooperation with the experts, qualify as competent users of the term *because* the expert knows how to use the term. Since what it is for an expert to know the meaning of, say, ‘hydroxidotrioxidoiodine’ is to know that hydroxidotrioxidoiodine is  $\text{HIO}_4$  *a*

*priori*, the same is knowable (although not necessarily *known*) *a priori* for normal language users.

Admittedly the sketch I have offered here certainly leaves work to be done. So far I have said nothing about biological kinds or mineralogical kinds, for instance, and work would certainly be required to adapt the account to such cases. Nevertheless, the account I have sketched does work for a large class of natural kind terms. It endorses Soames' analysis of complex natural kind terms, claiming that they are indeed descriptions, and employs the same essentialism required to make Soames' account work. It employs LaPorte's notion of 'open texture', agreeing that pre-scientific terms are vague, and that they undergo a process of precisification via redefinition. Like Extended Millianism, it offers a story concerning the *a priori* knowability of theoretical identities, borrowing Putnam's division of linguistic labour to make it more plausible. Finally, descriptivist essentialism produces intuitively correct results where Kripke's account will not. For instance, on Kripke's account if 'element 118' is a name, and a *de jure* rigid designator, then even if the element we were investigating, thinking it was element 118, had turned out to have atomic number 119, the name 'element 118' would rigidly designate it. Thus we are committed to saying that 'if element 118 exists, then element 118 is the element with atomic number 119', but this seems obviously wrong. What we should say is that element 118 turned out not to exist, and that the *proto*-name 'element 118' was *mistakenly* applied to what we should have called 'element 119'. Indeed had we known that the element we were investigating had atomic number 119 we would have derived the name 'element 119'.

## 5.8 Conclusion



In this chapter I have offered a novel version of descriptivism that avoids Kripke's modal and semantical arguments by endorsing the same brand of essentialism found in the important versions of (*CDR*). Then, rather than attempting to avoid the epistemological argument, I endorsed Putnam's division of linguistic labour (*DLL*), arguing that only a small sub-community of experts know the truth of theoretical identifications like (3), (5) and (7) *a priori*. The key point was that although theoretical identifications are *knowable a priori* for the entire community, it is only the experts that *know* them *a priori*: the majority of normal language users just don't know them, thereby maintaining the important intuition that Kripke's objection originally highlighted. Meanings, then, are in the head for experts, but not for the normal language user.

To make the consequence that theoretical identifications are both knowable and known *a priori* by experts palatable, I first pointed out that (*EM*) is also committed to this consequence across a range of important cases, and that if *a priori* does not constitute an objection to that thesis, then neither does it to the descriptivist thesis. I then presented a novel metasemantic story, where we start with descriptions, investigate the world to see whether that description has any attributive denotation, and then, when (and if) it does, generate a name for that natural kind. The key move is to separate the question of whether a natural kind term *has* denotation, from *how* we come to know the truth of theoretical identifications such as have been presented here. Successfully isolated, an *a posteriori* answer to the former does not imply an *a posteriori* answer to the latter.

The version of descriptivism defended here works by wedding descriptivism with metaphysical essentialism, but if my argument that Kripke's own account relies on essentialism to get the necessity of theoretical identification sentences off the ground is correct, then the account is no worse for that. Moreover, the point of this chapter was not to

defend or endorse any version of essentialism *per se*. Rather, the point was to demonstrate the work that *some* version of essentialism can do, when combined with a descriptivist theory of reference.

My concern now, then, is to look at precisely what essentialism is, since it has been employed by Kripke, Putnam, Soames and myself, without much in the way of explication. From the analysis in previous chapters we have established that there were two brands of necessary *a posteriori* truth within Kripke from the outset, and that one of them concerns the essences of individuals and natural kinds. Hence we can expect the essences of objects to be knowable only *a posteriori*. Furthermore, Kripke explicitly differentiates his notion of ‘essence’ from Peter Geach’s (1957: S16) ‘nominal essence’, claiming that Geach’s ‘nominal essence should be understood in terms of *a prioricity*, not necessity, and this is quite different from the kind of essence advocated here’ (Kripke 1980: 116 n. 58). Presumably the distinction Kripke intends is the fairly standard distinction attributed, although I shall claim falsely attributed, in the following chapter, to Locke: namely, that there is a dichotomy between ‘real’ and ‘nominal’ essences, that real essences concern the underlying properties of objects that explain why they are the way they are, and that nominal essences are only to do with our concepts. Hence the obvious place to start in attempting to articulate just what this notion of essence is, is with John Locke. It is to this I now turn in chapter 6, where I shall claim that the whilst those working in the Kripke-Putnam tradition might appeal to the Lockean notion of ‘real’ essence to capture this notion of an *a posteriori* non-definitional essence, they have simply misunderstood Locke, and are actually appealing to something more like the Aristotelian notion of essence.









































































































































































































