REFERENCE AND MODALITY:

A THEORY OF INTENSIONS

Alik Pelman

PhD Thesis 2007

University College London

I, Alik Pelman, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

The study of reference often leads to addressing fundamental issues in semantics, metaphysics and epistemology; this suggests that reference is closely linked to the three realms. The overall purpose of this study is to elucidate the structure of some of these links, through a close examination of the "mechanism" of reference. As in many other enquiries, considering the possible (i.e., the modal,) in addition to the actual proves very helpful in clarifying and explicating insights. The reference of a term with respect to possible worlds is commonly called "intension"; so this is a study of intensions. The main contribution of the study is an outline for a "calculator" of intensions. It is argued that the intension of a term is a function of three variables: (a) the way in which the term "picks out" its referent in different possible worlds (semantics); (b) criteria of identity (metaphysics); and (c) the actual state of affairs (actuality). While considering different possible values for these variables, it is demonstrated how the variables combine to generate the term's intension. In other words, the result is a calculator that when provided with the required values, yields the reference of the term in different possible worlds. By taking into account the possible gap between what we take the values of these variables to be and what they may in fact be, we also gain important insights into the epistemic aspect of reference. In addition, since a rigid designator is a term with constant intension, the proposed thesis provides an elaborate account of rigidity.

The first chapter is devoted to the development of the calculator of intensions. Each of the following three chapters elaborates on one aspect of intensions, namely, the semantic, metaphysical and epistemic aspects. In the course of these chapters, various familiar puzzles pertaining to the respective philosophical realms are addressed (many of these puzzles are discussed in Kripke's *Naming and Necessity* – a work that considerably inspired this study). In the fifth and last chapter the analysis of intensions is applied to two case-studies from relatively recent philosophical literature: the Kripke-Lewis debate over the identity theory of mind, and the debate over the significance of Donnellan's referential/attributive distinction. The novel accounts that these applications generate purport to illustrate the importance and originality of the proposed thesis.

In order to say what a meaning is, we may first ask what a meaning does, and then find something that does that

David Lewis, "General Semantics"

Contents

Acknowledgements

INTRODUCTION

i. Motivation and Overview

This dissertation is about reference, i.e., the relation between language and reality; between words and world. Moreover, it is about reference not only in actual states of affairs, but also in possible states of affairs, or, as it is commonly put, in *possible worlds*. Rudolf Carnap called the reference of a term with respect to different possible worlds "intension" (spelt with "s").[1] So this dissertation is about intensions.

Why reference? The study of reference often leads to addressing fundamental issues in semantics, metaphysics and epistemology. A familiar example is Saul Kripke's seminal work *Naming and Necessity*.[2] (Indeed, this dissertation is considerably inspired by Kripke's work and by the subsequent vast body of literature it generated). This suggests that reference is closely linked to the three realms. It is these links that are the main interest of the present dissertation. The overall purpose of this study is thus to elucidate the structure of some of these links. This is achieved through a close examination of the "mechanism" of reference.

Why intensions? In many enquiries, considering possible worlds (i.e., modality,) in addition to the actual world proves very helpful in clarifying and explicating insights (more on this below). The case of reference is no different in this respect: examining reference not only in the actual world but also in other possible worlds significantly improves our understanding of reference. Specifically, this study examines how different *factors* affect the reference of a term with respect to different possible worlds, i.e., how different factors affect intensions. (However, given the extensive analogy between the modal and the temporal, much of what is said here about reference in different worlds holds for reference in different *times* as well. Incidentally, the latter may be of more direct relevance to everyday concerns, such as identity over time. Nevertheless, this study focuses almost exclusively on reference in different possible worlds rather than in different times.)

This process results in the main contribution of the dissertation, which is an outline for a "calculator" of intensions. It is argued that the intension of a term is a function of three variables: (a) the way in which the term "picks out" its referent in different possible worlds (semantics); (b) criteria of identity (metaphysics); and (c) the actual state of affairs (actuality).[3] While considering different possible values of these variables, it is demonstrated how the variables *combine* to generate the term's intension. In other words, the result is a calculator that when provided with the required values, yields the intension of a term. By taking into account the possible gap between what we take the values of these variables to be and what they may in fact be, we also gain important insights into the epistemic aspect of reference.

In addition, since a *rigid designator* is a term with constant intension, the proposed analysis provides an elaborate account of rigidity.

It should not come as a surprise then that this study of reference addresses various familiar puzzles in semantics, metaphysics and epistemology (many of which are raised in Kripke's above work). The following list of puzzles gives a flavour of the types of matters with which we will be concerned throughout this dissertation. (Indeed, many of the puzzles belong in more than one realm, which implies the links between them):

Semantics: Does the term "water" refer to the watery stuff on Twin Earth, which is not H_2O but rather XYZ? Does it refer to the pink solid H_2O on another possible world? "Hesperus" and "Phosphorus" are two names for the same planet, so in a sense Hesperus might not have been distinct from Phosphorus, but in another sense we feel that it might have been; that they may have turned out to be distinct (e.g., for all the ancient astronomers knew). How shall we reconcile these seemingly conflicting intuitions? We see a man holding a martini glass, and thus refer to him as, "the man drinking a martini". It turns out that he is drinking water; did we succeed in referring to him nevertheless? "The police know concerning Smith's murderer, whoever he is, that he committed the murder, but they're not saying who he is"; is this a singular proposition about a particular person, or is it a general one concerning whoever murdered Smith?

Metaphysics: Could *this* table come into existence from a different hunk of wood than the one it was originally made of? Can the clay statue of Goliath survive a complete part-replacement? Can it survive being smashed into a ball? Is there such an entity, called "Prez", which is composed now (2007) by George W. Bush, was composed in 1790 by George Washington, and in 1863 by Abraham Lincoln? Can there be a possible world in

which Adam and Noah "switched roles"; that is, a world in which the guy who is absolutely similar to our Noah in *all* respects is in fact Adam, and vice versa? The essence of gold is said to be, having atomic number 79; what is the essence of being stuff with atomic number 79? If it is true that once we have created a physical duplicate of Fido the dog there is nothing left to duplicate (i.e., all other aspects of Fido were thereby also duplicated), does it follow that Fido's criteria of identity are physical? Can pain not be painful? Can something that feels painful not be pain?

Epistemology: Could we discover that Queen Elizabeth was not the daughter of George VI, i.e., is it Queen Elizabeth that we would be talking about in this case? Is it the job of science to discover the essence of things? What happened when we discovered that whales are mammals and not fish; was there a change in the concept 'fish'? How should we react if we were to discover that all things to which we refer to as "cat" are in fact robots controlled from Mars; should we say that cats are nothing but robots, or that we thought that there were cats but there aren't any?

The overall structure of the dissertation is the following: the first chapter is mainly devoted to the development of the calculator of intensions, and to the consequent account of rigidity. Each of the following three chapters discusses one aspect of intensions, namely, the semantic, metaphysical and epistemic aspects. Each of these chapters offers some more specific arguments and analyses relevant to the aspect of intensions that it explores. In addition, in the course of these chapters the above puzzles pertaining to the respective philosophical realms are addressed. In the fifth and last chapter we apply our intensions formula to two case-studies. The first is the Kripke-Lewis debate over the identity theory of mind as entailed by functionalism. Here our analysis is based on revealing the intension of "pain" as entailed by the competing views. The second case-study is Donnellan's distinction between the referential and the attributive uses of definite descriptions. Our analysis in this case is based on revealing the intension of "Smith's murderer" as entailed by different interpretations of Donnellan's distinction. It is hoped that the novel accounts that these applications generate illustrate the original contribution of this dissertation.

Due to the central role that intensions play in this study, it is worthwhile to briefly summarise the history of the notion of intension as reference across possible worlds. This is what we do next.

ii. A brief history of intensions

The history leading to the formation of the notion of intension as reference across possible worlds begins with the following well-known problem. In many contexts, substituting an expression with another expression that refers to the same object preserves the truth of the sentence. Thus, given that the number of the planets is nine,[4] we may substitute "nine" in

(1) Nine is greater than seven

with "the number of the planets", to get

(2) The number of the planets is greater than seven.

In other words, in such contexts, all that matters is the referent, also called the *extension*, of the terms involved (in our case, the number nine). Such contexts are thus called *extensional contexts*. Classical first-order logic is designed to deal with such contexts; it is extensional in nature. However, in some contexts, such substitutions fail to preserve truth. For instance, the ancient astronomers believed that there were six planets. Thus, although it is true that

(3) The ancient astronomers believed that nine is greater than seven

it is nonetheless false that,

(4) The ancient astronomers believed that the number of the planets is greater than seven;

even though "nine" and "the number of the planets" pick out the same referent, namely, the number nine.[5] Such belief context is called *propositional attitude* context. Other propositional attitude contexts create a similar problem. E.g., "S *knows* that…", "S *hopes* that", "S *regrets* that…" etc. But there are contexts other than propositional attitudes that similarly do not permit substitution *salva veritate* (i.e., in a truth preserving way) of expressions that pick out the same referent. One such context is the *modal* context; for example, although it is true that

(5) Nine is *necessarily* greater than seven

it is nonetheless false that

(6) The number of the planets is *necessarily* greater than seven[6]

(at least on one plausible reading of the sentence). Another such context is the *temporal*; e.g., although it is true that

(7) Britain's prime minister *used to be* a woman

(having Thatcher in mind), nevertheless,

(8) Tony Blair *used to be* a woman

is false, despite the fact that the terms "Britain's prime minister" and "Tony Blair" refer to the same individual. A similar problem applies to obligation contexts, like "it ought to be the case that…".

Furthermore, in addition to violating the principle of substitution, such contexts also violate a family of other extensional principles.[7]

So we have a range of contexts that behave in a way that violates extensional principles, and consequently cannot be dealt with by classical extensional logic. In Quine's view, this inability to be accommodated within classical logic does not indicate a flaw in classical logic, but rather a flaw in those contexts.[8] In his view, instead of amending logic to fit those constructions, we should rather amend the defiant constructions. Others thought otherwise. In contrast to Quine, they tried to "save the phenomena", i.e., they chose to take the linguistic phenomena as a given and to explicate them and, in some cases, even to develop a logic that would accommodate them.

Frege, who famously raised a powerful version of the above problem, suggested the following solution.[9] First he introduced a distinction between two aspects of expressions – their *bedeutung* and their *sinn*, commonly translated as *reference*, or denotation, and *sense*, or meaning, respectively.[10] (Frege's account of sense is as that which accounts for the *cognitive significance* of an expression.) Based on this distinction, Frege claimed that although expressions like "the number of the planets" and "nine" have the same reference, namely, the number nine, they nonetheless have different senses. More specifically, the two terms express different ways of conceiving the referent, and hence have different cognitive significance. Now, according to Frege, in extensional contexts, expressions stand for their reference. Hence such contexts allow for substitution of co-referring terms *salva veritate*. However, what happens in non-extensional contexts is that expressions no longer stand for their referents, but instead they stand for their senses. Now sameness in reference does not guarantee sameness in sense (think for instance of the celebrated example of "creature with heart" and "creature with kidneys" – apparently,

the two terms happen to refer to the same group of organisms, yet the terms clearly differ in their cognitive significance). Consequently, in such contexts we may not substitute terms that merely co-refer, as they may have different senses. Carnap's system (to be introduced shortly) used *intension* and *extension* instead of Frege's *sense* and *reference* respectively.[11] Following this distinction, contexts that do not permit substitution of co-referring expressions became known as *intensional contexts*.

The two aspects of expressions, sense and reference, apply not only to singular terms, such as "the number of the planets" or "nine" but also to predicates and whole sentences. In a letter to Husserl, Frege conveniently draws a chart specifying the reference and sense of each of these types of expressions (i.e., in extensional contexts. As mentioned above, according to Frege, in intensional contexts expressions refer to their senses). The following table is based on Frege's chart:[12]

	Singular term	**Predicate**	**Sentence**
Reference	Object	Concept/set[13]	Truth-value
Sense	Individual concept[14]	Sense of the predicate	Thought[15]

Table I.1

The problem with Frege's solution is that it is too "qualitative". That is, since senses remain somewhat obscure entities, such an account does not furnish a disciplined enough way to talk about intensional phenomena. In particular, Frege's solution does not propose any *formal* way of introducing senses into logic. In order for logic to deal with intensional phenomena we need to bring senses into logic. We need an explicit symbolic distinction between senses and references, one that will permit rigorous analysis.

Such an attempt at developing an intensional logic was pioneered by Alonzo Church, building on Russell's theory of types:[16] Russell discovered the famous paradoxes (that carry his name) entailed by Frege's system, and developed his theory of types as a solution.[17] Church took Russell's theory of types on board, and developed and formalised it into an intensional logic.

Church's pioneering attempt proposed a logical account of what senses *do*, but not so much of what senses in fact *were*. Such a system was developed by Rudolph Carnap in roughly the same period.[18] (Apparently Carnap and Church's ideas had significant mutual influence, and both philosophers refer to one another.)[19] Carnap's core idea is ingeniously simple. Although "the number of the planets" and "nine" have *actually* the same extension, they *might not* have. I.e., there is a *possible state of affairs* in which the

number of the planets is not nine; say, a state in which there are merely six planets. (Similarly, although the group of creatures with heart and the group of creatures with kidneys happen to coincide, they might not have.) In other words, the difference in the behaviour of the two terms is captured not by their *actual* extensions, but rather by their *possible* extensions.

Thus, on Carnap's view, in intensional contexts, we may substitute only such expressions that have the same extension, not only in the actual state, but rather in *all* possible states. For instance, since "the number of the planets" and "nine" have different extensions in some states, they may not be substituted in intensional contexts. Thus, substituting "nine" in (5) above with "the number of the planets" generates (6), which does not preserve the truth of (5), and hence such substitution is not permissible. However, since "nine" and, say, "the positive square root of eighty-one" do have the same extension in all possible states, substituting "nine" in (5) with "the positive square root of eighty-one" yields:

(9) The positive square root of eighty-one is *necessarily* greater than seven

which does preserve truth.

Technically, Carnap formulated his idea in terms of *state-descriptions*. A state-description is a "representation" of a possible state.[20] Each expression picks out a certain extension in the actual state as well as in possible states. Thus, Carnap defined *intension* as a function that assigns to each state-description the extension of the expression in that state-description.[21] For example, the intension of "the number of the planets" assigns to the actual state-description the number nine; to another possible state-description the number six; and so forth, it assigns to each state-description the number of the planets in that state.

Since different types of expressions – singular terms, predicates, sentences – have different types of extensions, they will also have different types of Carnapian intensions. In detail, since the extension of a singular term is an individual, the intension of a singular term is a function from state-descriptions to individuals; likewise, as the extension of a predicate is a set, the intension of a predicate is a function from state-descriptions to sets; and if the extension of a sentence is a truth-value (as Frege thought), then the intension of a sentence is a function from state-descriptions to truth-values (or, simply, a set of state-descriptions).

We may compare Frege's and Carnap's views in the following way. Frege suggested that in intensional contexts we may not substitute expressions that merely have the same reference, but rather only expressions that have the same meaning. Frege proposed to understand the meaning of the expression in terms of sense, i.e., that which accounts for the expression's cognitive significance. Carnap endorsed Frege's idea that in intensional contexts we may substitute only expressions that have the same meaning, and not merely the same reference. However, Carnap proposed to understand meanings in terms of *intensions*, i.e., functions from state-descriptions to extensions. (Accordingly, he used *extension* instead of *reference*, but in this case the two terms are largely equivalent.) Juxtaposing Carnap's intensions with Frege's senses (see Table I.2 below), Carnap's system suggests a clear account of what Frege's somewhat obscure senses are. In detail, individual concepts are understood as functions from state-descriptions to individuals; properties, or concepts, are understood as functions from state descriptions to sets, and thoughts, or propositions, are understood as functions from state-descriptions to truth-values (or simply as sets of state-descriptions):

	Singular term	**Predicate**	**Sentence**
Reference/extension	Object	Concept/set	Truth-value
(Frege's) sense	Individual concept	Sense of the predicate	Thought/proposition
(Carnap's) intension	A function from state-description to objects	A function from state-descriptions to sets	A function from state-descriptions to truth-values (or, simply, a set of state-descriptions)

Table I.2

Thus in Carnap's system, intensions are in fact defined in terms of extensions. Consequently, as opposed to Frege's theory which involves two qualitatively different aspects of expressions – their reference and their sense – Carnap's method of extension and intension involves, fundamentally, only one element, which is extension. It may be said then that Carnap's system "extensionalises" meanings. Thus understood, intensions become model-theoretic entities that may be thereby analysed within the framework of set-theory.

Carnap's ideas laid the foundation for the possible-worlds approach in intensional logic. Carnap defines his "state-descriptions" as follows:

> A class of sentences in S_I which contains for every atomic sentence either this sentence or its negation, but not both, and no other sentence, is called a state description in S_I, because it obviously gives a complete description of a possible state of the universe of individuals with respect to all properties and relations expressed by predicates of the system. Thus the state descriptions represent Leibniz's possible worlds or Wittgenstein's possible states of affairs.[22]

He credits Wittgenstein's ideas in the *Tractatus*[23] for being "the starting-point for the development of [his] method". Indeed, as the above quotation indicates, Carnap's state-descriptions are *representation of* Wittgenstein's possible states of affairs, or Leibniz's possible worlds. Now intensions for Carnap are functions from state-descriptions to extensions. If we replace Carnap's state-descriptions, i.e., the representations of possibilia, with the possibilia themselves, we get the basics of what became known as the application of *possible-worlds semantics* (PWS) to intensional logic. On this version of intensional logic, the intension of an expression is simply a function from *possible worlds* to extensions.

Following Carnap, Bressan[24] and Montague[25] both developed the possible-worlds approach in intensional logic. As Bressan's motivation was mainly in the realm of physics, it was largely Montague's work that contributed to the application of intensional logic in philosophy.[26]

However, although a possible-worlds approach in intensional logic manages to cope well with intensional contexts such as modality, tense, obligation, and indexicals, nonetheless propositional attitude contexts pose some serious problems to the possible-worlds semantics approach. Let me demonstrate this.

To recall, in many contexts, having the same extension is sufficient for substitution *salva veritate* (such contexts are called "extensional"). In intensional contexts however, having the same extension is not enough. Such contexts require not only the same extension but also the same meaning to allow for substitution *salve veritate*. Possible-worlds semantics interprets meanings as functions from possible worlds to extension, i.e., as intensions. Thus, on this interpretation, substitution in intensional contexts requires co-extension across *all* worlds. E.g., as the above example shows, although we may *not* substitute "nine" in the modal context (5) with "the number of the planets", we may nevertheless substituted "nine" with "the positive square root of eighty-one"; such substitution generates (9) which preserves the truth of (5). However, what seems to work in modal contexts appears not to work in propositional attitude contexts. For example, although it may be true that

(10)   Jones *believes that* nine is greater than seven

nevertheless, if Jones fails to realise that the positive square root of eighty one is nine, it might be false that

(11)   Jones *believes that* the positive square root of eighty-one is greater than seven.

Jones may simply have the one belief without having the other. So it seems that in the case of propositional attitude contexts, even having the same extension in *all* possible words, i.e., the same intension, cannot guarantee truth-preserving substitution.

Put differently, we may say that Carnapian intensions cannot provide the full story about what meaning is. The meaning of "nine" is different from the meaning of "the positive square root of eighty-one", and this difference cannot be captured by the extension in possible worlds; both terms have exactly the same extension in all possible worlds, i.e., they have the same intension.[27] (Similarly, think of the well-known example of "triangular" and "trilateral").

Such problems led to the development of some alternative intensional logics (Notably the systems suggested by Creswell[28], Anderson[29] and Zalta[30]).

However, despite ultimately failing to satisfactorily address problems such as the above, possible worlds semantics proved to be extremely fertile, and is in extensive use in many areas of philosophy, as well as in some neighbouring fields of study. As the following passage from Perry indicates:

> [Possible-worlds semantics] has been applied to a number of intensional phenomena in addition to necessity and possibility, including conditionals, tense and temporal adverbs, obligation and reports of informational and cognitive content. PWS spurred the development of philosophical logic and led to new applications of logic in computer science and artificial intelligence. It revolutionized the study of the semantics of natural languages. PWS has inspired analyses of many concepts of philosophical importance, and the concept of a possible world has been at the heart of important philosophical systems.[31]

To summarise, intensions are functions from possible worlds to extensions. In particular, the intension of an expression is a function that assigns to each possible world the extension of that expression in that world. Carnap's original aim in developing this notion of intension was to provide an account of meaning. However, as an account of meaning, intensions were shown to face a crucial problem: although they manage to capture many differences in meanings, nonetheless, there are some cases in which intensions fail to capture such differences. Nevertheless, despite this flaw, intensions, and

possible worlds semantics in general, proved very helpful for many other purposes, regardless of their original purpose. In fact, I believe that the use of possible worlds semantics in general, and of intensions in particular, is a mere formal expression of a much broader effective approach that underlies many of our practices, namely, gaining insights about what *is* the case by contemplating what *might have been* (and what *may be*) the case. Accordingly, this dissertation is largely guided by this approach in its exploration of reference.

Lastly, a brief comment about notes: to avoid frequent obstructions to the flow of the argument, I chose to use endnotes rather than footnotes in this dissertation. The drawback is of course the relative inconvenience in consulting endnotes. However, it is hoped that on the whole, the advantage of this decision in terms of convenience of following the main thread of thought overrides this drawback.

Endnotes

---

[1] Carnap (1947)

[2] Kripke (1972/1980)

[3] In fact, the initial formulation of the formula in Chapter 1 is slightly different, and we shall only reach this particular formulation at the end of Chapter 2 (See Section 2.4).

[4] Ignoring some recent astronomical controversies.

[5] The discovery of the puzzle is commonly attributed to Frege, but at least some version of it probably dates back to the ancient Greeks in the form of *The Hooded Man* paradox: "You say that you know your brother. Yet that man who just came in with his head covered is your brother and you did not know him" (Hyde, 2005); the paradox was later widely discussed by medieval logicians. (The paradox is attributed to the Megarian logician Eubulides of Miletus – a contemporary and critic of Aristotle – as one of the "Seven Megarian Paradoxes" which include the famous Sorites paradox and the Liar paradox.)

[6] This influential example is due to Quine (1963a)

[7] For example, existential generalisation. It is a principle of extensional logic that the fact that a has the property P entails that there exists something that has the property P (i.e., $Pa \supset \exists x Px$). However, from "Jones likes Santa Claus" (paraphrased as "Santa Claus has the property of being liked by Jones") it does not follow that Santa Claus exists.

[8] In particular, Quine (1956, 1961) thought that in such contexts expressions do not refer at all; he called them *referentially opaque*.

[9] Frege (1892)

[10] Something like that distinction was of course around much before Frege, and goes back to at least the medieval distinction between *signification* and *supposition* (William of Sherwood, 1240's, tr. N. Kretzmann; 1966). Closer to Frege's times, Mill (1867) distinguished the *denotation* of a term from its *connotation*.

[11] Carnap (1947)

[12] The letter is dated 24 May 1891. The original (translated) chart (as it appears in Wiggins (1984)), is:

Sentence	Singular term	Concept-word	
sense of the sentence (thought)	sense of the singular term	sense of the concept-word	
reference of the sentence (truth value)	reference of the singular term (object)	(reference of the concept-word (concept))	object falling under the concept

[13] Extensional semantics, developed following Frege's ideas, standardly treats the reference of predicates as sets; specifically, the reference of an *n*-place predicate is the set of *n*-tuples of objects to which the predicate applies. As Frege's letter to Husserl indicates (see Endnote 12), his own view was along these lines.

[14] I.e. an identifying condition for individuals. Many (notably Kripke) understood such a condition to be an abbreviation of a definite description (see sections 1.1.1 and 2.2.1.1.2 for some discussion) However, such an interpretation is controversial and there are other important alternatives (Cf. Marti, 1998).

[15] Commonly referred to more technically as *proposition*.

[16] Church (1951). Although an abstract carrying the same title appeared already in his (1946), and an informal discussion was presented in the introduction to his (1944).

[17] Russell's discovery of the paradoxes was made as early as 1901; first introduced in his (1903) and further developed in his (1908) and in Russell and Whitehead (1910, 1912, 1913). The paradoxes are related to the set of all sets that are not members of themselves. Such a set is a member of itself if and only if it is not a member of itself.

[18] Carnap (1947). This observation is due to Fitting (2006)

[19] Fitting (2006)

[20] Carnap (1947, p. 9); see quote on p. 22.

[21] In the context of logic, the use of the term *intension* dates back to at least the early 19th century, and stands for something like the meaning of the expression:

> The Internal Quality of a notion,- its Intension or Comprehension is made up of .. the various characters connected by the concept itself into a single whole in thought.
>
> (Sir W. Hamilton (1836-60), *Logic viii*, as cited in the OED).

Clearly, Carnap's use of the term is much more technical.

[22] Carnap (1947, p. 9)

[23] Wittgenstein (1921). It can be derived from Wittgenstein that there are many possible states of affairs, one of which is the actual state of affairs. Thus Wittgenstein's ideas are sometimes seen as anticipating possible-worlds semantics (e.g., Fitting, 2006).

[24] Bressan (1972)

[25] Montague (1960; 1970)

[26] Recently, Copland (2006) has suggested the following genealogy of possible-worlds semantics. He proposes three "threads": first the analysis of modalities in terms of quantification over possible worlds. This idea he attributes to Carnap (1946). The second thread is the relation of accessibility between worlds. This relation is absent in Carnap's system. Although the relation was introduced and discussed in the 1940's, it was only in the mid 50's that it was applied explicitly to possible worlds by Meredith and Prior (1956). The third thread, according to Copeland, includes completeness proofs. Such proofs were first provided by Smiley (1957). Other proofs were provided by Bayart (1958, 1959); Kripke (1959); and, Hintikka (in his informal talks as a research fellow in Harvard, around 1958-9).

[27] It should be noted however, that although ultimately sameness in intension does not guarantee sameness in meaning, still, intensions and meanings do bear an "almost" one-one correlation to one another, and therefore sameness in intension, for the overwhelming majority of cases, does go hand in hand with sameness in meaning. For some more discussion see Section 3.3.5.

[28] Creswell (1985)

[29] Anderson (1984)

[30] Zalta (1988)

[31] Perry (1998)

1 A BLUEPRINT FOR A CALCULATOR OF INTENSIONS

Overview

Our prime interest is in reference. More specifically, reference in actual and possible states of affairs. In other words, we are interested in (Carnapian) intensions. Formally, intensions are defined as functions from possible worlds to extensions: the intension of a term is a function that assigns to each possible world the extension of the term in that world.

First, we locate the different factors upon which intensions depend. (These factors are found through insights that we collect from different philosophical debates, at least some of which are traditionally unrelated to one another.) Specifically, in the first stage these factors are found to be; (a) the semantic function of the term, i.e., the way in which the term is designed to "pick out" its referent with respect to different possible worlds (e.g., as a Russellian definite description, or as a Millian proper name, etc.); (b) the criteria of identity (e.g., whether objects are primarily material objects that merely happen to have their manifest properties, or whether they are primarily manifest objects that merely happen to have their material constitution); and (c) the actual referent. Changing the values of any of these variables may result in a change in the reference of the term with respect to different possible worlds, viz., it may result in a change in the intension of the term.

However, learning that each of these factors affects intensions, is still far from being enough for determining how the three work *together* to fix intensions; the whole in this case is greater than the sum of its parts. Thus, we next explore how the reference of a term across counterfactual worlds is fixed by the three factors together. (This will make the main original contribution of this chapter.) To avoid presuppositions, we construct a thought experiment which takes place on Mars. Our Martian colleagues introduce us to some Martian stuff referred to as "T", and ask us to help them identify T on other possible worlds. Or, technically speaking, we are asked to determine the *intension* of "T", i.e., what the term designates with respect to different possible worlds. Following a short

series of (thought-) experiments on the planet, we come up with a general "calculator" of intensions, i.e., a "formula" that, when provided with some required data, enables the determination of the intension of a term in question. We then draw some morals about rigidity.

Having developed our intensions calculator, we turn to use it for some critical discussions. First, we return to examine the traditional debates from which we inherited the factors in our formula. We learn that the intensions that are implicit or explicit in those debates are merely partially justified, as they take only one factor into account and not all three. Finally, we use our insights about rigidity to critically discuss Kripke's essentialism. We conclude that although Kripke is justified in holding that rigidity entails *some* modal properties, Kripke is nonetheless not justified in using rigidity to derive *specific* essential properties.

As this is the first chapter, terms and ideas are brought into discussion for the first time, and hence require frequent qualifications and commenting, as well as addressing worries that the reader may have in mind before more has been said as the argument unfolds. However, to facilitate the flow of the argument, it is advisable to read the endnotes to this chapter after reading the main text, unless there is a required specification or a particular worry that seeks response.

1.1   The Dependence of Intension on Other Factors – Three Cases

1.1.1   The dependence of intension on the term's semantic function

(Or: what we learned about intension from the New Theory of Reference)

A common way of dealing with problems regarding the semantics of terms is by invoking thought experiments that involve counterfactual worlds[1,2] (also known as "twin-earth" thought experiments). A paradigm is the case of the term "water". Consider a counterfactual world in which the watery stuff, i.e., the colourless odourless stuff that falls from the sky and flows in rivers etc., is not H_2O but rather XYZ. Does it deserve to be called "water"?[3] This thought experiment can be extended: consider also a counterfactual world in which the H_2O stuff is non-watery, but is say, pink, opaque and solid. Does it deserve to be called "water"?[4]

According to the so-called Frege[5]-Russell[6] view, (or at least according to the way that Kripke understands their view,[7]) terms like "water" have a descriptive content, and they designate[8], with respect to every possible world, that which fits their descriptive content.[9] Thus, "water" – whose descriptive content is, presumably, 'the watery stuff' – designates, with respect to every possible world, that which is watery in that world. That will include the watery XYZ, but not the non-watery, i.e., pink solid H_2O. This theory of reference is commonly called "descriptive".[10] Accordingly, let us call terms that have such a semantic function "descriptive terms".

By contrast, according to Kripke's view,[11] terms like "water" are designed to designate, with respect to every possible world, the same referent that they designate in the actual world.[12,13,14] Thus, says Kripke, since in the actual world "water" designates H_2O, "water" designates H_2O in every possible world. That will exclude the watery XYZ, but will include the pink-solid H_2O.[15] This theory of reference is commonly called "referential" or simply, the "New Theory" of reference. Let us call terms that designate according to this theory "nondescriptive".[16]

There are many arguments against, as well as in favour of, each view. However, as in this study we are interested mainly in intensions, we shall refrain from taking sides in the debate, and instead concentrate our attention on what we can learn from the debate about reference in possible worlds, i.e., about intensions. And, as we shall see, the above semantic debate can teach us something about intensions.

We can divide counterfactual worlds into four relevant types. W_1: containing (like the actual world) watery H_2O (H_2O+W); W_2: containing watery stuff which is not H_2O, e.g., XYZ ($\neg H_2O+W$); W_3: containing non-watery, e.g., pink-solid, H_2O ($H_2O+\neg W$); and W_4: containing only stuffs which are neither H_2O nor watery ($\neg H_2O+\neg W$). (For the sake of simplicity, we suppose homogeneity within a world, i.e., we ignore here worlds that contain more than one of these types, e.g., worlds in which there is both watery XYZ *and* non-watery H_2O.[17]) Thus, according to the Frege-Russell theory, "water" is descriptive and therefore designates the watery stuffs on worlds W_1 and W_2, but not the non-watery stuffs on worlds W_3 and W_4. According to Kripke's view, by contrast, "water" is a nondescriptive term, and hence it designates, with respect to every counterfactual world, the same stuff that it designates in the actual world, namely, H_2O. A nondescriptive "water" thus designates the H_2O stuffs on W_1 and W_3, but not the non-H_2O stuffs on W_2

and W_4 (note that the two views are compatible with our use of the term in the actual world, i.e., of type *W1*). The resulting intensions of "water" are thus:

	W_1 $H_2O + W$	W_2 $\neg H_2O + W$	W_3 $H_2O + \neg W$	W_4 $\neg H_2O + \neg W$
"Water" is descriptive	+	+	-	-
"Water" nondescriptive	+	-	+	-

Table 1.1

What Table 1.1 shows is that the intension of "water", i.e., the reference of "water" with respect to different counterfactual worlds, is dependent on whether the term "water" is descriptive or not. And in general, the intension of a term depends, at least in part, upon its semantic function. In other words, changing the semantic function of a term may result in a change in the intension of that term.

1.1.2   The dependence of intension on criteria of identity

(Or: what we learned about intension from discussions about identity and change)

It is customary to engage with the metaphysical problem of identity by invoking thought experiments that involve manipulations of objects. Specifically, these manipulations commonly involve changes to the material constitution of the object, hence the more restricted title, "the problem of material constitution".[18]

Following the ancient Greeks, many versions of the problem make reference to a statue. Consider a clay statue of David. Or, let us describe it more neutrally: consider an object which is made out of clay and has a David shape. Call it "David".[19] Now consider a scenario in which David is smashed into a ball. Is the clay ball David? Let us extend the thought experiment. Consider another scenario in which David undergoes a gradual part-replacement, until it is entirely made out of bronze.[20] Is this David?

Mereological essentialism has it that parts are essential to their wholes (commonly summarised by the slogan "composition is identity"); hence objects cannot undergo part-replacement and remain the same.[21] In other words, on this view, the *criterion of identity*[22,23] for objects is: an object remains the same as long as it keeps its material parts. Thus, on this view, the clay ball is identical to David but the bronze statue is not.

By contrast, Burke (1994) suggested another criterion of identity, which he discusses in terms of *sortal* (a sortal is what gives the criterion of identity to an object[24]): "of the sortals satisfied by an object, the one that tells the object's sort is the one whose satisfaction entails possession of the widest range of properties."[25] Based on this criterion, Burke selects, for example, "tree" over "hunk of cells", and "statue" over a "piece of copper" as sortal properties.[26] Accordingly, being the same object as David is not being the same lump of clay, but rather, being the same statue (which also allows for part-replacement[27]). Thus, according to this principle, the above bronze statue is David, but the clay ball is not.[28]

In Wiggins's terminology,[29] we can say that on the first view, being a lump of clay is what David *is*, while being a statue is merely what David *is like*, whereas on the second view being a statue is what David *is*, and being a lump of clay is what David *is like*. Another terminology uses the distinction between *things* and *properties of things*. Thus on this terminology, according to the first view, lumps of clay and lumps of bronze belong to the group of *things* that inhabit the world, whereas being statue-shaped and being vase-shaped belong to the group of *properties of such things* (and others) in the world. By contrast, Burke's view has it that it is statues and vases that belong to the group of *things* in the world, while being constituted by a lump of clay and being constituted by a lump of bronze are but *properties of such things* (and others).

Admittedly, there are other views, e.g., that there are two objects in the same place at the same time – a statue *and* a lump of clay;[30] or, that what the original object primarily is, is relative to the way it is described/intended/thought of,[31] i.e., the clay ball is identical to David relative to the description "lump of clay" but not relative to the description "statue"; or, that the original object is a *stage* (i.e., a temporal part of a four-dimensionally extended object), shared by two different collections of stages – a statue-collection and a lump-collection;[32] or that identity is governed by Leibniz's law (i.e., two objects are identical if and only if they share all their properties[33]) and hence that neither the clay ball nor the bronze statue are David;[34] and more.[35] However, for the purpose of the argument advanced here, it suffices that there is more than one metaphysical view.

Again, there are strong arguments for and against each view, but being concerned with intensions, we shall not engage here in choosing sides. Interestingly, this debate can also teach us something about intensions. Let me demonstrate this.

We can translate the above scenarios into a modal version by supposing that each occurs in some counterfactual world. Accordingly, we can divide all counterfactual worlds into types according to the end-state of the manipulation. W_1: (David remains as is) a clay statue (C+S); W_2: the same lump of clay, different shape (C+¬S); W_3: the same shape, different matter (¬C+S); and W_4: different matter and different shape (¬C+¬S).[36]

The request to determine which object is David, is in fact the request to determine, with respect to each counterfactual world, whether the relevant object deserves the name "David". In other words, the request is simply to determine the intension of "David".[37] Specifically, on the mereological essentialist view, "David" designates the lump of clay, whatever shape it has, thus "David" designates the objects on W_1 and on W_3 but not those on W_2 and on W_4; on Burke's view, by contrast, "David" designates the same statue, whatever it is made of, i.e., "David" designates the objects on W_1 and on W_2, but not those on W_3 and on W_4. The intension of "David" is thus:

	W_1 C + S	W_2 ¬C + S	W_3 C + ¬S	W_4 ¬C + ¬S
Mereological essentialism	+	-	+	-
Burke's view	+	+	-	-

Table 1.2

It turns out that different criteria of identity result in different designations with respect to counterfactual worlds. Thus, one interesting lesson from metaphysical discussions of identity through change is that the intension of a term may depend, at least partially, on criteria of identity. In other words, different criteria of identity of an object may result in different intensions of the term that designates this object.

1.1.3   The dependence of intension on the actual referent of the term

(Or: what we learned about intension from two-dimensional semantics)

The type of dependence discussed in this section is mainly manifest in the two-dimensional semantics literature.[38]

Consider the Leibniz-Newton controversy over who invented the Calculus first. Surely, one of them was the first to invent it, but it was a matter of acrimonious controversy who actually did. This controversy can also teach us something about intensions.

Take the singular term "the actual inventor of the Calculus".[39] Like the simple definite description "the inventor of the Calculus", it designates in the actual world the person who fits the description, i.e., the person who actually invented the Calculus. However, with respect to counterfactual worlds, whereas the simple definite description (arguably) designates, with respect to every counterfactual world, the individual who invented the Calculus *in that world*, the actualised description "the *actual* inventor of the Calculus" by contrast designates, with respect to each counterfactual world, the individual who invented the Calculus *in the actual world*, i.e., regardless of who invented it in that counterfactual world. Or so at least we shall postulate here.[40]

Now, consider a counterfactual world, in which it is clear that Newton was the first to invent the Calculus. Is he the *actual* inventor of the Calculus in that world? Similarly, in a counterfactual world in which Leibniz originally invented the Calculus: is it him who is the *actual* inventor of the Calculus in that world?

The answers depend on the inventor in the *actual* world. Leibniz claimed to have conceived the invention first. Hence, according to Leibniz, "the actual inventor of the Calculus" designates himself in every counterfactual world, regardless of who first conceived the invention in that world. However, according to Newton and Hook (it was the latter who conducted this bitter debate with Leibniz), Newton preceded Leibniz, and hence the term designates Newton in every counterfactual world, regardless of who first conceived the invention in that world.

Thus, the designation of the singular term "the actual inventor of the Calculus" (or, indeed, of any term which abbreviates this actualised description) with respect to the different counterfactual worlds in each case is:

Counterfactual worlds ⟍ The actual world	W_1 The inventor of the Calculus is Leibniz	W_2 The inventor of the Calculus is Newton	W_3 The inventor of the Calculus is neither Newton nor Leibniz
The inventor of the Calculus is Leibniz	Leibniz	Leibniz	Leibniz
The inventor of the Calculus is Newton	Newton	Newton	Newton

Table 1.3

It turns out that different actual referents of a certain term may result in different intensions of that term.

History is but one realm in which we deliberate different possibilities of the actual world. In fact, any realm of *a posteriori* exploration, like searching for the gravitational force of Mars, or for the unit of inheritance in organisms, aims at finding what the actual state of affairs is like in a certain respect. Different findings about the actual state of affairs may result in different intensions. In particular, different findings with respect to the actual referent of a term may result in different intensions of that term.

Note that Table 1.3 posits two sets of possible worlds: a set of counterfactual worlds (horizontal), and a set of actual worlds (vertical).[41] The two sets of worlds correspond to two types of possibilities: possibilities of the actual vs. possibilities of the counterfactual. Thus, in one sense, it is possible that the actual inventor of the Calculus was Newton, and similarly, it is possible that the actual inventor of the Calculus was Leibniz. This sense of possibility corresponds to possibilities of the actual world: the term "the actual inventor of the Calculus" designates different referents in different *actual* worlds. However, given what the actual world is like, there is another sense in which it is *not* possible that the *actual* inventor of the Calculus is different from who it in fact was; this sense corresponds to possibilities of the counterfactual: the term "the actual inventor of the Calculus" designates the same referent in all *counterfactual* worlds.[42]

Thus the division into two sets of possible worlds is a good way to account for conflicting intuitions that we sometimes have: e.g., if the actual inventor was Newton, then on the one hand the actual inventor might have been Leibniz, whereas on the other hand the actual inventor might not have been Leibniz.[43]

The intension of a term was defined as a function that assigns to each possible world the extension of the term in that world. Given the two sets of worlds – actual worlds and counterfactual worlds – we can thus define two types of such functions: one from *counterfactual* worlds to extensions, the other from *actual* worlds to extensions.[44] Chalmers calls these *secondary intension* and *primary intension*, respectively;[45] Jackson calls them *C-intension* and *A-intension* respectively ("C" for "counterfactual"; "A" for "actual").[46] Throughout this dissertation, we shall simply use "intension" to indicate the first function, i.e., from *counterfactual* worlds to extensions.

1.1.4   An epistemological note

In each of the above three cases there was one factor for which we have considered different options. Our reason for considering the different options in each of the cases was a certain state of *uncertainty*; some epistemic ignorance. The reason for the uncertainty in each case was that the facts, at least as far as we could tell, seemed to underdetermine the definite value of the variable in question. In particular, in the first case, the way people normally use the term "water", i.e., in the actual world, seems to be perfectly compatible with both the term being descriptive and with the term being nondescriptive. In the second case, the way the world looks and behaves seems to be compatible with both a material criterion of identity for David and with a manifest criterion of identity. (Generally, a world with material criteria of identity and a world with manifest ones could look and behave exactly the same.) Lastly, in the case of the actual inventor of the Calculus, given all the evidence we have, it is not clear to date whether it was Newton or Leibniz who first arrived at the Calculus, if any. Overall then, we may say that in each case the information available to us underdetermined the values of the variables in question. (Note that this has no entailment as to facts of the matter; in each case there may or may not be facts of the matter.)

Now, we have also witnessed how, in each case, different options regarding the variable in question may result in different intensions. Clearly then, being in a certain state of ignorance about a certain variable *highlights* the dependence of the intension on that variable. However, such a state of ignorance is *not* essential to the dependency. In other words, although the different possibilities with regard to the three variables *can* account for corresponding epistemic states of ignorance, they *need* not; the possibilities stand independent of our epistemic state. What the above cases show is that *were* the value of a certain variable changed, the intension *would* change accordingly. And this would remain true even if we knew beyond any doubt the correct values! For example, had we simply known, beyond any doubt, that "water" is a nondescriptive term, it would still remain true that *were* it not such a term, but rather a descriptive one, its intension would change accordingly. Similarly, had we known beyond any doubt that the criterion for being identical to David is material, it would still be true that *were* the criterion of identity manifest, the intension of "David" would change accordingly. And, lastly, had we known beyond doubt that the actual inventor of the Calculus was Newton, it would still be true that *were* the actual inventor Leibniz, the intension of "the actual inventor of the

Calculus" would change accordingly.[47] And these are enough to indicate the relevant dependencies.

The conclusion to be drawn here is that the dependence of intension on each of the three factors by no means relies on epistemic doubts, although such doubts clearly emphasise it.

1.2   The Intension of a Term as a Function of *all* Three Factors

So far, we have established that the intension of a term may depend, at least partially, on each of the following three factors: (a) the semantic function of the term; (b) the criteria of identity;[48] and (c) the actual referent of the term.[49]

Yet apparently in each of the above cases we were able to determine the intension of our term on the basis of just *one* factor. Specifically, in the first case of water, we've determined the intension of "water" on the basis of the semantics of the term "water" alone; in the case of the clay statue David, we've determined the intension of "David" on the basis of the criteria of identity alone; and in the last case of the Leibniz-Newton controversy, we've determined the intension of "the actual inventor of the Calculus" on the basis of the actual referent alone. So, this may suggest that we may need but *one* of the three factors, if at all, and not *all* three of them, in order to determine intensions.

This, however, is a mistake. The reason for our ability to determine intensions on the basis of just one factor in each case was that the values of the other two factors were either provided, or speculated by us. Let me demonstrate this.

In the case of water, although we had no presuppositions regarding the semantic function of "water", nonetheless, we've assumed certain criteria of identity, namely, that being the same as the actual watery H_2O is simply being H_2O, and not, for instance, being watery. If the latter were the case, the intension would change accordingly (as was illustrated in the statue/lump case). In addition, we were told that the stuff we called "water" in the actual world, was indeed watery H_2O. This is largely based on our current scientific reports. However, science is an *a posteriori* endeavour, and as such is fallible, i.e., it may be the case that current science is just wrong about it, and that the material constitution of the actual watery stuff is something other than H_2O. Again, if the latter

were the case, the intension would alter accordingly (as was illustrated in the Leibniz-Newton case).

In the second case, of David, we had no metaphysical presuppositions. Yet, we have assumed that the proper name "David" is not a descriptive one, i.e., we have assumed that it is not an abbreviation of some definite description, say, "having a statue-shape of David", but rather (as the New Theory of reference dictates) a nondescriptive term. Yet if the former is the case, it will modify the intension of "David" accordingly (as we have seen in the first example). In addition, we have been told what the actual referent was, namely, a clay statue. However, if we were really to be at the scene, then for all we know, the actual statue could be a clay-like plaster statue. Or even a hologram of a clay statue. Again, this would change the intension of "David" as well (as the Leibniz-Newton case shows).

Lastly, in the case of the actual inventor of the Calculus, although we made no assumptions about the actual referent, we've assumed no problem with re-identifying Leibniz and Newton in other possible worlds, i.e., we've assumed clear criteria of identity for persons, which is a heavy metaphysical presupposition. Different criteria of identity for people entail different referents in counterfactual worlds, viz., different intensions.[50] Also, the semantic function of the term was given to us by clearly stating that it was an actualised description. However, a term with a less clear semantic function, e.g., a term for which the semantic function is under dispute, or simply a term in a foreign language, would not enable us to determine the intension in this case on the basis of the actual referent alone.

To summarise, out of the three factors upon which intension depends, in each of the three cases there was one factor that was not presupposed, while the other two factors were either given or speculated. The important point is that, regardless of whether the presuppositions made were right or wrong, were the values of *any* of the three factors changed, the intension could change accordingly. (We shall later see how exactly in detail). Hence, it follows that the intension of a term is always dependent on *all* three factors.[51]

We may look at it this way: the formula that determines the intension (to be constructed soon) consists of three variables. The exact intension of a certain term is fixed by the values of these three variables. In each of the above cases the values of two of the variables were fixed, leaving the value of the third one open. This has generated a

manipulation which isolated one of the variables, thus enabling us to learn the contribution of this variable to the overall intension, separately for each variable.

The next stage is to see how the three factors work *together* to fix the overall intension. To illustrate the combined contribution, let us consider the following science-fiction scenario.

1.2.1 The way the three factors *combine* to fix intensions – a Martian case-study[52]

Suppose our Martian fellow-philosophers introduce us to some Martian, jellylike, purple stuff called "T", and ask us to help them identify T on other possible worlds to which they frequently travel.[53] (Needless to say, Mars, as well as Earth, both belong to the actual world). In other words, we are asked to determine what would deserve to be called "T" in counterfactual situations. Or, more technically, we are simply asked to determine the *intension* of the term "T".[54] We, of course, are happy to assist. So happy, that we postpone our doubts about the possibility of travelling to other possible worlds.

So we start working. If "T" is a *descriptive* term,[55] then "T" has a descriptive content, for instance, 'a jellylike purple stuff'. In such a case, the Martian stuff is designated by "T" by virtue of fitting this descriptive content of "T", and likewise "T" designates, with respect to every counterfactual world, whatever fits this descriptive content, i.e., whatever is jellylike purple stuff, in that world. As there might be different stuffs that fit this descriptive content, T on some counterfactual worlds may be distinct from the actual Martian referent.

Yet "T" may not be such a term; it may be a *nondescriptive* term, i.e., a term that is designed to designate, with respect to every counterfactual world, the same stuff that it designates in the actual world.[56] If "T" is indeed such a nondescriptive term, then "T" designates, with respect to every counterfactual world, just this very same stuff that we were introduced to a short while ago on Mars, in that world.

So if "T" is a nondescriptive term, then in order to identify T on counterfactual worlds we need to be able to tell whether a certain stuff that we might encounter on some such world is or is not identical to our actual stuff. Thus before travelling to counterfactual worlds, we need to get a pretty good idea of what this stuff that "T" actually designates consists in. So we go and have a closer examination of that actual

referent of "T". We note its manifest properties – that it is purple, jellylike, etc., and upon further examination we determine its physical constitution to be P. We learn from the Martians that they in fact conceive the collection of manifest properties as one property, M. We easily manage to adapt and conclude that the actual referent of "T" has the properties P and M. So "T", if nondescriptive, designates, with respect to every counterfactual world, that which is identical to this P+M stuff. Had this actual referent been different, the reference of "T" with respect to different possible worlds, i.e., the intension of "T", would accordingly be different.

With respect to these two properties, P and M, there are four kinds of stuffs that may be found on counterfactual worlds: (a) stuff that is P and M (P+M); (b) stuff that although shares the manifest property M has nonetheless a different physical constitution, say, Q (Q+M); (c) stuff that has the same constitution P but a different collection of manifest properties, say, N (P+N);[57] and finally, (d) stuff that is neither P nor M, say, some Q and N stuff (Q+N). Each of these stuffs in fact stands for a type of a counterfactual world:[58] W_1 containing (like the actual world) P+M stuff; W_2 containing Q+M stuff; W_3 containing P+N stuff; and W_4 containing Q+N stuff. (We are aware, that if there is more than one alternative to P, or more than one alternative to M, there will be accordingly more candidates, and thus more types of counterfactual worlds. Our list of counterfactual worlds may thus not be exhaustive. This fact, however, need not currently worry us, for the argument as it stands by no means depends on a complete list of counterfactual worlds).

For a start, we consider the candidate that has the same manifest property M as the actual stuff, but different material constitution, namely, Q+M, which is the stuff on counterfactual worlds of type W_2. Is it identical to the actual stuff or not?

Being heavily indoctrinated by the presuppositions of our Earthean science, we tend to believe that the world is such that it contains various materials, such as P and Q, that merely happen to have manifest properties like M and N. In other words, we presuppose that the criterion of identity for stuffs is their material constitution. So it seems that despite the sameness of the manifest property, the stuff on counterfactual W_2 (Q+M) is different from the actual stuff (P+M). It is, rather, the stuff on counterfactual W_3, P+N, which is identical to the actual stuff P+M, despite the difference in their manifest property. Yet, to our surprise, we soon learn that the Martians have a different view on this matter; they take the world to be such that it primarily contains manifest

entities, such as M and N, that merely happen to have some material constitution, like P or Q. In other words, they believe that the criterion of identity for stuffs is their manifest property. Thus in their view, the stuff on counterfactual W_2, Q+M, and the actual stuff, P+M, are identical since they share the same manifest property M. After the initial, mutual puzzlement about these conflicting views, we soon realise how dogmatic our beliefs are, and thank the fortunate encounter that allowed us to broaden our outlook and to see both options. We both admit, however, that we have no way of determining what the world is really like in this respect. Indeed, for all we know, there may be no fact of the matter here at all.

Furthermore, we now appreciate that with regard to any property Ω of our stuff, it is possible, contrary to both what we or the Martians believe, that Ω is the criterion of identity of that stuff, i.e., the world is such that it primarily contains Ωs, that merely happen to have other properties, such as material constitution like P or Q, and manifest properties like M or N (such a property Ω may, for instance, be the referent's function). So in principle, there are many more possible metaphysical alternatives.[59]

Turning back to our job, we conclude so far that the intension of the term "T", i.e., "T"'s reference with respect to different counterfactual worlds, is dependent upon three different factors: (a) the semantic function of "T"; (b) the referent of "T" in the actual world; and (c) the criteria of identity of the actual referent, namely, what the referent is vs. what it is merely like (despite our strong, yet utterly dogmatic, conviction that the manifest properties are merely what this material object is like).

But what then *is* the intension of "T"? What does in fact deserve to be called "T"?

Well, surely, to determine the intension of "T" we need the *values* of each of the above three variables. We have already come to suspect that we do not know for sure the value of the third variable – the criteria of identity. But as if this worry was not enough, the Martians go on to tell us that they are not even sure about the values of the other two variables either – the semantic function of "T" and the actual referent of "T" (more precisely, the properties of the actual referent). All they are willing to commit to is the purple jellylike stuff to which the term "T" actually applies on Mars. Firstly, regarding the true properties of this actual referent; although their current science tells them that it has the properties P and M, they are nonetheless well aware of how fallible their science has been, and therefore are careful not to assume that it isn't failing this time as well. And

secondly, as to the semantic function of "T", some of them think that "T" designated that stuff by virtue of simply naming it; whereas others, by contrast, contend that "T" designated the stuff by virtue of that stuff fitting the descriptive content of the term "T", namely, by being jellylike, purple etc. (i.e., having the property M). Furthermore, they think that other views on this matter are also possible. And there is nothing in the actual application of the term on Mars to tell them which of the views is right. Pondering this, we begin to suspect that our epistemic state may not be very different from theirs. Firstly, the one certain thing about our science is that it is fallible, and hence it may well be that the actual stuff is not P+M as our investigation tells us. Secondly, concerning the semantic function of "T", for all we know, "T" could refer to its stuff by virtue of different semantic functions. And thirdly, as for the criteria of identity, we have already established that it was likewise underdetermined by the facts.

(To be sure, the intension of "T" does not depend upon the Martians' views about the three factors; nor upon ours. The intension of "T" depends upon the real values of these factors (if there are facts of the matter with regard to these). When there is a disagreement about these values, whether between us and the Martians, or between the Martians and themselves as well as amongst ourselves, it is about the *real* values of these variables.)

The problem with this epistemic state is that without the values of the required variables, we simply cannot carry out the job. It is just impossible to determine the intension of "T" without this data, since different data, i.e., different values of the three variables, will result in different intensions. So, what do we do now? Is there a way out?

Well, of course there is! We could provide the Martians with some well-defined procedure, which would enable them to calculate the intension of "T" *once* the values of the variables are determined. We could come up with some sort of a *formula*, that when filled in with the appropriate data would simply generate the exact intension. Such a general formula carries some important advantages. Firstly, with an appropriate re-interpretation of its variables, such a formula can be used to calculate the intension of *any* term. Secondly, such a general formula would provide an understanding of the *general regularity* behind individual intensions of specific terms. After all, isn't it what drives great scientists in their pursuit of theories? Admittedly, unlike us, Newton had methods for measuring weights, speeds, forces, and other values for the variables in his formulas, whereas we, regrettably, do not seem to possess such philosophical yardsticks.

Nonetheless, what Newton was really after was not so much the specific outcomes of his formulas, but rather the general *rule* that would expose the *regularity* connecting such values. And in *this* respect our formula should enjoy the same merit. And lastly, our difficulty in conclusively determining the exact values presupposes realism, i.e., the view that there is a fact of the matter regarding the above three variables, and that our mission is to *discover* these facts. Yet there seems to be nothing to prevent the possibility that there are no such facts at all (at least as far as the semantic function and the criteria of identity are concerned. Such antirealism is less likely to apply to the third variable – the properties of the actual referent – in relation to which realistic intuitions are rather strong). Under such antirealist assumptions the search for the values of these variables would be plainly absurd. As antirealism has it, we do not *discover* whatever it is that we are antirealist about; rather, we *stipulate* it. But if indeed such antirealisms hold, then applying the formula becomes a rather straightforward act; all one needs to do is input the stipulated values and get the resulting intension immediately.

So it now seems that even if we knew beyond any doubt the exact values of the three variables, it would still be highly beneficial to deliberately assume a "veil of ignorance" and to consider the hypothetical values of each variable in order to find the general formula. Encouraged by these revelations about the importance of a general formula, we enthusiastically undertake the task.

The main values that we considered are the following: (a) the values of the semantic function: either descriptive or nondescriptive; (b) the values of the actual referent (corresponding to our apparent P+M actual referent of "T"): P+M, or Q+M, or P+N, or Q+N;[60] and (c) the values of the criteria of identity: either material or manifest.

(We are aware that there are possibilities not covered by the ones mentioned – at least as far as the (properties of) the actual referent and the criteria of identity are concerned. For the sake of simplicity however, we shall construct our formula by first considering these options. After arriving at the principal formula, it shouldn't be too difficult to extend it to include other options as well.)

Apparently the logical space of possibilities here amounts to two (possibilities of the semantic function) times four (possibilities of the actual referent) times two (possibilities of the criteria of identity), which is sixteen. Fortunately, the options are considerably less than that. This is due to the fact that if the semantic function of "T" is descriptive, the values of the other variables play no part in fixing the intension. "T" will

simply designate, with respect to every counterfactual world, that which fits "T"'s descriptive content (allegedly, 'being an M') in that world – regardless of the criteria of identity, or the nature of the actual referent. Thus all eight combinations linked with a descriptive semantic function for "T" merge into one.

Overall then, we end up with the following nine possible combinations:

	Semantic function	The actual referent	Criterion of identity	W_1 P+M	W_2 Q+M	W_3 P+N	W_4 Q+N
1	Descriptive	(Whatever)	(Whatever)				
2	Nondescriptive	P+M	Material				
3			Manifest				
4		Q+M	Material				
5			Manifest				
6		P+N	Material				
7			Manifest				
8		Q+N	Material				
9			Manifest				

Table 1.4

All we need to do now is to calculate the intension of "T", i.e., "T"'s reference with respect to the different counterfactual worlds, for each of the nine rows. This is a fairly easy task, which we can quickly perform.

In Row 1, "T" is descriptive and is linked with the descriptive content 'have the (manifest) property M'; "T" thus designates, with respect to each type of counterfactual worlds, that which M's (i.e., that which has the property M) in that type of world. Hence, the intension of "T" in that case is:

	Semantic function	The actual referent	Criterion of identity	W_1 P+M	W_2 Q+M	W_3 P+N	W_4 Q+N
1	Descriptive	(Whatever)	(Whatever)	+	+	-	-

Table 1.5

In all subsequent cases, "T" is nondescriptive, and therefore designates, with respect to every counterfactual world, that which it designates in the actual world. So in order to determine the reference of "T" in the given types of worlds, we need to determine first what is the nature of the actual referent of "T". We can do that on the

basis of the other two variables: the criterion of identity, and the (properties of) the actual referent. Let us demonstrate this.

In Row 2, "T" designates in the actual world stuff that is P+M (as we actually believe to be the case), and the criterion of identity is material, i.e., the world is inhabited, among other things, by things like P's and Q's, some of which happen to have manifest properties like M and N. Hence our actual referent is such a P that happens to have the manifest property M. So in this case, "T" designates, with respect to every counterfactual world, that which is P in that world. The intension is thus:

	Semantic function	Actual referent	Criterion of identity	W_1 P+M	W_2 Q+M	W_3 P+N	W_4 Q+N
2	Nondescriptive	P+M	Material	+	-	+	-

Table 1.6

Row 3 is like Row 2, only that now we have a manifest criterion of identity. So in that case, the same actual referent, P+M, is in fact an M that merely happens to have the material constitution P. In this case, "T" designates, with respect to every counterfactual world, that which is M in that world. The intension of "T" is therefore:

	Semantic function	Actual referent	Criterion of identity	W_1 P+M	W_2 Q+M	W_3 P+N	W_4 Q+N
3	Nondescriptive	P+M	Manifest	+	+	-	-

Table 1.7

In Rows 4 and 5, the actual referent is not P+M but rather Q+M. (I.e., in these rows we consider the possibility that our actual referent has different properties than we believe it to have.) Now if the criterion of identity is material, then the referent, Q+M, is primarily a Q that merely happens to have the manifest property M. So "T" will designate, with respect to every counterfactual world, that which Q's (i.e., has the property Q) in that world. If, however, the criterion of identity is manifest, this actual referent is an M, and "T" will designate, with respect to every counterfactual world, that which M's in that world:

	Semantic function	Actual referent	Criterion of identity	W_1 P+M	W_2 Q+M	W_3 P+N	W_4 Q+N
4	Nondescriptive	Q+M	Material	-	+	-	+
5	Nondescriptive	Q+M	Manifest	+	+	-	-

Table 1.8

Similar considerations will determine the intension for the remaining four rows.

Overall, we end up with the complete Table 1.9, which lists "T"'s intension relative to each combination of values of our three variables:

	Semantic function	The actual referent	Criterion of identity	W_1 P+M	W_2 Q+M	W_3 P+N	W_4 Q+N
1	Descriptive	(Whatever)	(Whatever)	+	+	-	-
2	Nondescriptive	P+M	Material	+	-	+	-
3			Manifest	+	+	-	-
4		Q+M	Material	-	+	-	+
5			Manifest	+	+	-	-
6		P+N	Material	+	-	+	-
7			Manifest	-	-	+	+
8		Q+N	Material	-	+	-	+
9			Manifest	-	-	+	+

Table 1.9

And so we're done. Mission complete. (Well, at least an outline of our task is complete. For our account to be truly complete we would still need to include more possible options for each variable.)

One last comment is that although the values of the three variables determine the intension of a term, the converse does not hold; i.e., we cannot determine the values of the variables on the basis of some given intension. This is due to the fact that each type of intension recurs more than once in the table. In particular, our formula contains four different types of intensions, that we can call a, b, c and d:

	Semantic function	The actual referent	Criterion of identity	W_1 P+M	W_2 Q+M	W_3 P+N	W_4 Q+N	
1	Descriptive	(Whatever)	(Whatever)	+	+	-	-	**a**
2	Nondescriptive	P+M	Material	+	-	+	-	**b**
3			Manifest	+	+	-	-	**a**
4		Q+M	Material	-	+	-	+	**c**
5			Manifest	+	+	-	-	**a**
6		P+N	Material	+	-	+	-	**b**
7			Manifest	-	-	+	+	**d**
8		Q+N	Material	-	+	-	+	**c**
9			Manifest	-	-	+	+	**d**

Table 1.10

Thus if we are told, for instance, that the intension of "M" is such that it designates the stuffs on *W1* and *W2*, this fits Rows 1, 3 and 5, and hence this information alone underdetermines which of the three rows represents the true facts. In other words, it may be that "M" is descriptive; alternatively, it may be that "M" is nondescriptive and that the actual referent is P+M and the criterion of identity is manifest; but it may also be that "M" is still nondescriptive and the criterion of identity is still manifest, yet the actual referent is Q+M.

Let us summarise. We were looking for the intension of "T", i.e., the reference of "T" with respect to different counterfactual worlds, or simply, what would deserve to be called "T". We have found the three factors upon which intensions depend. We have further considered possible values for each of these variables. We have listed the different possible combination of these values. Finally, we have calculated the resulting intension for each combination. Thus, using the corresponding values with regard to the term "T", our Martian colleagues can now determine the intension of "T". In fact, they can do more than that. By reinterpreting "P", "M", "Q" and "N", our hosts are now in a position to determine the intension of *any* term, since the intensions given in Table 1.9 remain fixed. Finally, they can also appreciate the common general rule underlying all such intensions.

And, frankly, so can we.

1.3   Rigidity

We can now use our newly devised formula to gain some important insights about rigidity.

1.3.1   Rigid/nonrigid vs. rigidified/non-rigidified

A *rigid term* is defined as a term that designates the same referent in all counterfactual worlds, and nothing but that referent.[61] A nondescriptive term was defined above as a term that is designed to designate, with respect to every counterfactual world, the same referent that it designates in the actual world. Thus it follows that a nondescriptive term is *by definition* a rigid term. A descriptive term, by contrast, carries no such entailment: recall that a descriptive term was defined as a term that is linked with a descriptive

content, and that designates, with respect to every counterfactual world, that which fits its descriptive content in that world. Thus, a descriptive term might be nonrigid – e.g., "the fourth planet from the sun" designates Mars in the actual world, Venus in another counterfactual world, etc. Yet a descriptive term might also be rigid – e.g., "the sum of 2+2" designates 4 in every counterfactual world, and nothing else; likewise, "stuff with atomic number 79" (arguably) designates gold in every counterfactual world and nothing else.

Thus, whereas a nondescriptive term is *rigidified* by the very semantic function that is attached to it, a descriptive term (which may or may not be rigid) is not thus rigidified. In Kripke's terms, the difference between the two kinds of rigidity is that a nondescriptive term is rigid *de jure*, i.e., by stipulating a rule, whereas a descriptive terms is rigid, when it is rigid, *de facto*, i.e., as it happens in practice (without stipulation).[62]

It follows that the distinction between rigid and nonrigid is not equivalent to the distinction between rigidified and not rigidified: while all rigidified terms are rigid, some non-rigidified (i.e., descriptive) terms are rigid and some are not. It is the distinction between rigidified and non-rigidified – rather than between rigid and nonrigid – which is the more important from a semantic point of view. The following quote from Kaplan expresses this idea.

> For me, the intuitive idea is not that of an expression which *turns out* to designate the same object in all possible circumstances [i.e., a rigid descriptive term], but an expression whose semantic *rules* provide *directly* that the referent in all possible circumstances is fixed to be the actual referent [i.e., a rigid nondescriptive term].[63]

Nondescriptive terms are thus rigidified terms. Descriptive terms are not rigidified: they may turn out rigid or not rigid. We can thus replace the title "nondescriptive" with "rigidified". This should be welcomed as some types of "nondescriptive" terms, as we shall see in Chapter 2, have nonetheless some descriptive element to them, and hence the term "nondescriptive" is somewhat misleading. Thus, we shall henceforth use "descriptive" vs. "rigidified" (rather than "nondescriptive") to express the two types of semantic functions with which we were concerned so far.

1.3.2   Rigidity and Metaphysics

Thus all rigidified terms are rigid, whereas only some of the descriptive terms are rigid. However, what is it that makes a descriptive term rigid?

A rigid term is one that designates the same referent in all counterfactual worlds, and only that referent. Our intensions formula enables us to determine the referent of a term with respect to each counterfactual world; but in order to determine rigidity, we must also know whether these referents are identical or not. For instance, the formula entails that the intension of "T" when "T" is descriptive (and has the descriptive content 'being M') is as follows:

	Semantic function	Actual referent	Criterion of identity	W_1 P+M	W_2 Q+M	W_3 P+N	W_4 Q+N
1	Descriptive	(Whatever)	(Whatever)	+	+	-	-

Table 1.11

I.e., the referents of a descriptive "T" across counterfactual worlds are P+M and Q+M. Yet in order to determine whether "T" is rigid or not, we have to further establish whether these referents, namely, P+M and Q+M, are identical or not.

Now this depends on the criteria of identity (which is our metaphysical factor). The dependence works as follows: if the criteria of identity are material, i.e., two stuffs are the same iff they share the same material constitution, then the two stuffs P+M and Q+M, one being P (that happens to have the manifest property M) and the other being Q (that happens to have the same manifest property M) are in effect distinct. If, by contrast, the criteria of identity are manifest, i.e., two stuffs are the same iff they share the same manifest properties, then the two stuffs P+M and Q+M, being both M, are the same (despite having different material constitutions). As a consequence, material criteria of identity will entail that a descriptive "T" (with the descriptive content 'being M') is nonrigid, whereas manifest criteria of identity will entail that such a descriptive "T" is rigid; and this is despite the identical intension in both cases:

	Semantic function	Actual referent	Criteria of identity	W_1 P+M	W_2 Q+M	W_3 P+N	W_4 Q+N	*Rigidity*
1a	Descriptive	(Whatever)	Material	+	+	-	-	Nonrigid
1b		(Whatever)	Manifest	+	+	-	-	Rigid

Table 1.12

It turns out that the same term, with the same descriptive semantic function, may be rigid or nonrigid, depending on the criteria of identity, which is a metaphysical issue. Rigidity in general is thus not an exclusively semantic concept; it is no less a metaphysical concept.[64]

To generalise, what makes a descriptive term like "T" rigid, is that its descriptive content – in our case 'being M' – is linked with criteria of identity, since such metaphysics guarantees that all that fits that descriptive content is the same stuff. Whereas when the descriptive content is not so linked, then the term is rendered nonrigid, since there are different things that would fit this descriptive content.

Given that all options in which "T" is rigidified render "T" rigid, it follows that there is but a single case in which "T" is rendered nonrigid. This happens only when "T" is descriptive *and* its descriptive content is not linked with criteria of identity. In all other cases, "T" comes out rigid (despite the difference in intensions):

	Semantic function	**Actual referent**	**Criteria of identity**	W_1 P+M	W_2 Q+M	W_3 P+N	W_4 Q+N	*Rigidity*
1a	Descriptive	(Whatever)	Material	+	+	-	-	*Nonrigid*
1b		(Whatever)	Manifest	+	+	-	-	*Rigid (De facto)*
2	Rigidified	P+M	Material	+	-	+	-	*Rigid (De jure)*
3			Manifest	+	+	-	-	
4		Q+M	Material	-	+	-	+	
5			Manifest	+	+	-	-	
6		P+N	Material	+	-	+	-	
7			Manifest	-	-	+	+	
8		Q+N	Material	-	+	-	+	
9			Manifest	-	-	+	+	

Table 1.13

One important fact that Table 1.13 highlights is that the mere rigidity of a rigidified term is independent of criteria of identity, i.e., a rigidified rigid term is compatible with quite different criteria of identity. Hence, the mere rigidity of a rigidified term entails no *specific* criteria of identity. However, the mere rigidity of a rigidified term does entail that there are *some* criteria of identity.[65]

Consequently, the mere rigidity of a rigidified term "T" entails no *particular* intension. A rigidified "T" is designed to designate the same referent in all counterfactual worlds. Being the same is determined by criteria of identity. Thus, different criteria of identity will result in different intensions of a rigidified term. So the fact that a certain term "T" is rigid is compatible with various, quite distinct, intensions that this term may have.

To summarise, the only nonrigid terms are descriptive terms whose descriptive content is not linked with criteria of identity. All other terms, namely, all rigidified terms, and all descriptive terms whose descriptive content is linked with criteria of identity, are rigid. Rigidified terms are rigid *de jure*, i.e., their rigidity follows from their semantic function, while descriptive terms, when rigid, are rigid *de facto*, i.e., it merely *turns out* that they are rigid. It follows that the rigid/nonrigid distinction and the rigidified/non-rigidified distinction are not identical: whereas all rigidified terms are rigid, not all rigid terms are rigidified: some rigid terms are not-rigidified, i.e., they are merely *de facto* rigid.

Finally, one way to compare descriptive and rigidified terms is the following. As far as rigidified terms are concerned, their rigidity is guaranteed, regardless of criteria of identity or the actual referent. However, their intension is determined by the latter two, i.e., the mere rigidity of a rigidified term does not determine a specific intension, as that can only be determined by the criteria of identity and the actual referent. By contrast, when it comes to descriptive terms, their intension is determined irrespective of the metaphysical criteria of identity or the actual referent, yet their rigidity is dependent on the metaphysical criteria of identity. This is clearly demonstrated in Table 1.13.

1.4   Applying the Intensions Calculator to the Initial Three Cases

We return now to the initial three cases that were presented at the outset: the water case, the clay statue case, and the inventor of the Calculus case. Having developed our intensions formula, and the consequent account of rigidity, we can now apply those to the three cases.

1.4.1   The "water" case revisited

The generality of our intensions formula is generated due to the ability to reinterpret the letters in the formula. One such reinterpretation is the following: "T" stands for "water"; "P" for "H_2O"; "M" for "watery"; "Q" for "XYZ"; and "N" for "non-watery". Such reinterpretation generates a calculator for the intensions of "water". This calculator shows that the intension of "water" (like the original "T") has many more possibilities than those considered in the traditional debate over the semantics of "water" (as was detailed in Table 1.1):

	Semantic function	Actual referent	Criteria of identity	W_1 H₂O+W	W_2 XYZ+W	W_3 H₂O+¬W	W_4 XYZ+¬W	*Rigidity*
1a	Descriptive	(Whatever)	Material	+	+	-	-	*Nonrigid*
1b		(Whatever)	Manifest	+	+	-	-	*Rigid (De facto)*
2	Rigidified	H₂O+W	Material	+	-	+	-	*Rigid (De jure)*
3			Manifest	+	+	-	-	
4		XYZ+W	Material	-	+	-	+	
5			Manifest	+	+	-	-	
6		H₂O+¬W	Material	+	-	+	-	
7			Manifest	-	-	+	+	
8		XYZ+¬W	Material	-	+	-	+	
9			Manifest	-	-	+	+	

Table 1.14

We will use the above calculator of the intensions of "water" to reject two common New-Theorist claims.

The first New Theorist claim is that since "water" is rigid (or, more precisely, as we now understand, *rigidified*), i.e., since "water" designates in all counterfactual worlds that which it designates in the actual world, it follows that it designates H₂O in all counterfactual worlds.[66] Another New Theorist claim is that on the Frege-Russell view, since "water" designates, with respect to every counterfactual world, that which is watery, the term is thereby nonrigid on that view, i.e., "water" designates different stuffs in different counterfactual worlds. Both these claims need not be true, as they are based on unfounded metaphysical speculations. Let me explain.

Consider the first claim; that if "water" is rigidified, then it designates H₂O in every counterfactual world. This, indeed, is true when the criterion of identity of the actual referent of "water", i.e., the watery H₂O stuff, is material (Row 2 in Table 1.14). But what if the criterion of identity is not material but rather manifest? Namely, what if some stuff is identical to the watery H₂O iff it is watery, i.e., if and only if it has the same manifest property (just as Burke would hold that David remains the same as long as it maintains its shape)?[67] As the Martian case-study shows, in principle there is nothing in our knowledge – *a priori* or *a posteriori* (i.e., the way the world looks and behaves) – to rule out manifest criteria of identity. And if the criterion of identity is indeed manifest (Row 3), then a *rigidified* "water" designates, with respect to every counterfactual world, that which is watery in that world. In fact the intension of a rigidified "water" under manifest

criteria of identity (Row 3) turns out identical to the intension of a descriptive "water" (Row 1); it designates watery XYZ.

Thus, the New Theorists' assumption that a rigidified "water" designates H_2O in every counterfactual world, whether watery or not (Row 2), need not be true: under different metaphysical assumptions, the same rigidified "water" designates watery stuff in every counterfactual, whether H_2O or not.

Furthermore, even if the criterion of identity is indeed material, a rigidified "water" need not designate H_2O in every counterfactual world. It may be that our scientists just got things wrong about the actual referent of "water". As Chalmers exemplifies:

> Given that *this* world is actual, it turns out that 'water' refers to H_2O, and its Kripkean intension [i.e., function from counterfactual (rather than actual) worlds to extensions] picks out H_2O in all possible [i.e., counterfactual] worlds. But if another world had been actual (e.g. Putnam's Twin Earth world in which XYZ is the clear liquid in the oceans), 'water' might have referred to something quite different (e.g. XYZ), and it might have had an entirely different Kripkean intension (e.g. one that picks out XYZ in all worlds). [68]

Thus, despite our scientists' best judgement, it may be the case that our actual watery stuff is not H_2O at all, but rather XYZ.[69] In that case, a rigidified "water", under material criteria of identity (which Chalmers, like Kripke, takes to be the case, with no absolute justification[70]), will designate XYZ in every counterfactual world, just like Row 4 in Table 1.14 shows.

It should be noted however, that Kripke formulates his claim in a conditional manner: *if* (by which I believe he means 'if and only if') "water" actually designates watery H_2O, *then* "water" designates (all and only) H_2O in every counterfactual world.[71] Thus, with respect to the last point, Kripke seems to be aware of the fact that if "water" in fact turns out to designate watery XYZ, then (given material criteria of identity) a rigidified "water" designates XYZ in every counterfactual world.

Finally, the mistake can be on both fronts, i.e., it may be the case that the criterion of identity is manifest, *and* that the watery stuff in the actual world is other than what we take it to be, say, it is XYZ. As Row 5 shows, "water" in that case does not designate H_2O in all counterfactual worlds. (In fact the intension of "water" in that case would be the same as that of a descriptive "water".)

(The other four rows in the table relate to cases in which the actual referent is not watery. As it is highly unlikely that we should be mistaken here, these rows, although correct, are quite irrelevant here.)

Thus overall, the claim that "water", if rigidified, designates H_2O with respect to every counterfactual world is far from being obviously true. As our formula of intensions shows, there are many cases in which this is simply not the case.

The second claim of the New Theory was that on the Frege-Russell view, "water" is nonrigid. Specifically, since "water" designates, with respect to every counterfactual world, that which is watery, the term designates different stuffs in different counterfactual worlds. This, also, need not be true. A descriptive "water" indeed designates, with respect to every counterfactual world, that which is watery in that world. But how do we determine whether all those watery referents are identical or not? This can only be done with the aid of criteria of identity. Suppose that the criteria of identity of stuffs are manifest, i.e., that two stuffs are the same if and only if they share the same manifest properties. In that case (as Row 1b demonstrates), all the referents of a descriptive "water", by virtue of being watery, are rendered identical. In other words, in this case "water" turns out a rigid (albeit *de facto*) designator. It is only relative to material criteria of identity that these referents are distinct. Hence, the New Theorists' claim for the nonrigidity of "water" on the Frege-Russell view need not be true as well.

1.4.2   The statue/lump case revisited

The statue/lump case taught us that intension is dependent upon metaphysical criteria of identity. Specifically, relative to a manifest criterion of identity, "David" designates the same statue, whether it is constituted by clay or by bronze, whereas relative to a material criterion of identity, "David" designates the same lump of clay, whether it is statue-shaped or not. Yet, we later understood that this was true only given the following tacit assumptions: (i) that "David" is a rigidified term, i.e., that "David" designates in all counterfactual worlds that which it designates in the actual world; and (ii) that the actual referent indeed has the properties of being constituted by a lump of clay (and not, say, by a clay-like plaster), and of being statue-shaped. Again, a complete account of the intension of "David" should take all alternatives to these assumptions into consideration. In other words, it should consider all possible values of these variables. An application of

our intensions-formula to the statue/lump case yields the following possible intensions of "David":

	Semantic function	The actual referent	Criterion of identity	W_1 Clay statue	W_2 Bronze statue	W_3 Clay ball	W_4 Bronze ball	...
1	Descriptive	(Whatever)	(Whatever)	+	+	-	-	
2	Rigidified	Clay statue	Material	+	-	+	-	
3			Manifest	+	+	-	-	
4		Bronze statue	Material	-	+	-	+	
5			Manifest	+	+	-	-	
6		Clay ball	Material	+	-	+	-	
7			Manifest	-	-	+	+	
8		Bronze ball	Material	-	+	-	+	
9			Manifest	-	-	+	+	

Table 1.15

(As in the water case, the above intensions are based on the general formula for "T"; they are arrived at by simply inserting the values relevant to the David case into that formula).

The original discussion concentrated merely on the metaphysical aspect, while *presupposing* that the term "David" is rigidified, and that the actual referent was indeed a statue-shaped lump of clay. By this, the original discussion considered merely two options, namely, those represented by Row 2 - a material criterion of identity, and by Row 3 – a manifest criterion of identity. Yet, again, other options are also possible in this case. E.g., if "David" is descriptive rather than rigidified term, its intension would be the one entailed by Row 1 (which, as it happens, entails the same intension as the one implied by Row 3). Also, if the actual referent was not a clay statue, but rather, say, a bronze statue, the intension would change as well. In other words, in order to determine the intension of "David", in addition to considering the metaphysical issue, there is also a semantic issue to be considered, namely, whether "David" is a descriptive term or not, as well as actuality issues, namely, whether the properties of our actual object are really being a statue and being constituted by a lump of clay. These factors will also affect the intension.

1.4.3 "The actual inventor of the Calculus" case revisited

Finally, the case of the actual inventor of the Calculus taught us that intension is dependent upon the actual referent of the term. Specifically, this case revealed that it makes a difference to the designation of "the actual inventor of the Calculus" in counterfactual worlds whether the actual referent was Newton or Leibniz. Yet we later appreciated that the intensions determined in that discussion were true only relative to two other crucial variables, namely, (i) that the semantic function of the actualised description "the actual inventor of the Calculus" is to designate, with respect to every counterfactual worlds, the individual who is the inventor *in the actual world* (and hence that the term is a rigidified one); and (ii) that the criterion of identity for people is not problematic. However, again, these need not be true.

Firstly, in the original case, we were told that the term was "the actual inventor of the Calculus". So we knew it was a rigidified term. However, the situation could be slightly modified in a way that would conceal this fact. Suppose, for instance, that we just heard someone referring to Newton in a foreign language using an expression Φ. Now this expression may be a rigidified one (like, "the *actual* inventor of the Calculus", or, arguably, like "Newton"), or not rigidified – i.e., it may be descriptive (e.g., an abbreviation of the non-actualised description "the inventor of the Calculus"). This, we know well by now, may affect the intensions. In particular, with respect to Table 1.16, a *descriptive* Φ with the descriptive content "the inventor of the Calculus", will designate Newton in counterfactual world W_1, yet it will designate Leibniz in counterfactual W_2, (regardless of criteria of identity, and, indeed, of the actual referent) – as specified in Row 1. The intension of a *rigidified* Φ, by contrast, will be as specified in Rows 2 and 3.

	Semantic function	The actual inventor of the Calculus	W_1 the inventor of the Calculus is Newton	W_2 The inventor of the Calculus is Leibniz	…
1	Descriptive	(Whatever)	Newton	Leibniz	…
2	Rigidified	Newton	Newton	Newton	Newton
3		Leibniz	Leibniz	Leibniz	Leibniz

Table 1.16

Secondly, the criteria for the identification of Newton and Leibniz in counterfactual worlds also play a crucial role in determining the intension of "the actual inventor of the Calculus", i.e., the intension of the term is dependent upon this

metaphysical variable. Discussions of personal identity commonly involve thought-experiments similar to ones like the statue/lump thought experiment, only that the shape-aspect is replaced with a psychological aspect. Such "body swapping" or "mind swapping" thought experiments date back at least to Locke's much-discussed scenario:

> … should the soul of a prince, carrying with it the consciousness of the prince's past life, enter and inform the body of a cobbler, as soon as deserted by his own soul …[72]

A similar thing could happen to the "soul", i.e., the mental features, of the cobbler. And it could also happen to the mental features of Newton and Leibniz. Thus, in some counterfactual worlds it is possible to transfer mental features from one body to another. As a result, on some such "body-snatchers" world, lives, for example, someone with Newton's body, but with someone else's mental features, say, Leibniz's. Similarly, on some other world lives someone with Leibniz's body, but now with someone else's mental features, say, Newton's. So, among the people that we may encounter on counterfactual worlds, there are the following four. Newton's actual mental features in Newton's actual body (B_N+M_N); Newton's actual mental features in Leibniz's body (B_L+M_N); Leibniz's mental features in Newton's body (B_N+M_L); and Leibniz's mental features in Leibniz's body (B_L+M_L) (there are, of course, many more types of worlds in this respect):

W_1	W_2	W_3	W_4	…
B_N+M_N	B_L+M_N	B_N+M_L	B_L+M_L	

Table 1.17

Now suppose we know for certain that in the actual world it was Newton, i.e., a guy with Newton's mental features and Newton's body ($B_N + M_N$), who invented the Calculus. Thus, the term "the actual inventor of the Calculus" designates, with respect to every counterfactual world, just that guy, Newton. But now, how should we re-identify Newton on counterfactual worlds as described above? Is Newton identical to the guy who shares actual Newton's mental features, but Leibniz's body (B_L+M_N), or with the guy who shares actual Newton's body but Leibniz's mental features (B_N+M_L), or none? The same problem, of course, applies to the case in which the actual inventor was Leibniz; is he identical with the first, or second, or none?

To determine whether a certain individual is Newton or not we need to have the criteria of identity of Newton, and generally, the criteria of identity of people – traditionally referred to as criteria of *personal identity*. Many philosophers advocate some

version of the so-called *psychological criterion*, according to which, roughly, two individuals are the same if and only if[73] they share the same mental features.[74] Others – admittedly far fewer - promote some version of the so-called *somatic view*, according to which two individuals are the same if and only if they share the same body[75] (or are the same animal,[76] or something like that).[77] Thus, according to the psychological view, Newton is identical to the individuals on worlds W_1 and W_2 and not to the ones on W_3 and W_4 (in Table 1.17 above); whereas according to the somatic view, Newton is identical to the individuals on W_1 and W_3, but not to the ones on W_2 and W_4. In other words, the intensions of "the actual inventor if the Calculus", in case Newton was the one are:

Semantic function	The actual referent's properties	Criterion of identity	W_1 B_N+M_N	W_2 B_L+M_N	W_3 B_N+M_L	W_4 B_L+M_L
Rigidified	Newton	Somatic	+	-	+	-
		Psychological	+	+	-	-

Table 1.18

On the other hand, Leibniz, according to the psychological view is identical to the guys in W_3 and W_4, who share his mental features, whereas according to the somatic view, he is identical to the guys on W_2 and W_4, who share his body. Thus, the intensions of "the actual inventor of the Calculus", if Leibniz was the one, are:

Semantic function	The actual referent's properties	Criterion of identity	W_1 B_N+M_N	W_2 B_L+M_N	W_3 B_N+M_L	W_4 B_L+M_L
Rigidified	Leibniz	Somatic	-	+	-	+
		Psychological	-	-	+	+

Table 1.19

However, unbeknownst to us, the actual world may have body-snatchers as well. Thus, the actual inventor of the Calculus might be one of our four types of individuals, namely, a guy with Newton's bodily and mental features; a guy with Newton's body but with someone else's mental features (say, Leibniz's); a guy with Leibniz's body but with someone else's mental features (say, Newton's); or a guy with Leibniz's bodily and mental features. Table 1.18 and Table 1.19 specify the intensions of "the actual inventor of the Calculus" in case the actual inventor was either the former or latter individual, respectively. However, the actual inventor may be one of the other two guys as well. In which cases the intensions will change accordingly. Thus, the complete list of intensions is:

	Semantic function	The actual inventor of the Calculus is	Criterion of identity	W_1 B_N+M_N	W_2 B_L+M_N	W_3 B_N+M_L	W_4 B_L+M_L	...
1	Descriptive	(Whatever)	(Whatever)	+	+	-	-	
2	Rigidified	B_N+M_N (Newton)	Somatic	+	-	+	-	
3			Psychological	+	+	-	-	
4		B_L+M_N	Somatic	-	+	-	+	
5			Psychological	+	+	-	-	
6		B_N+M_L	Somatic	+	-	+	-	
7			Psychological	-	-	+	+	
8		B_L+M_L (Leibniz)	Somatic	-	+	-	+	
9			Psychological	-	-	+	+	

Table 1.20

And this, again, is just our general formula adapted to apply to the case of "the actual inventor of the Calculus".

1.5 Discussion: Challenging Kripke's Essentialism

Our analysis of intensions, and the subsequent account of rigidity, is at odds with three essentialist theses of Kripke's. In what follows, I shall present arguments against these essentialist theses.

Table 1.13 showed the following links between rigidity and criteria of identity: (a) a rigid term designates the same referent in all counterfactual worlds. Being the same is dependent upon criteria of identity. Hence rigidity presupposes that there are criteria of identity; (b) the rigidity of a rigidified term is independent of any *specific* criteria of identity; a rigidified term is compatible with any coherent criteria of identity. It follows that the mere rigidity of a certain term entails no specific criteria of identity. E.g., the mere fact that "water" is rigid is compatible with both material criteria of identity, according to which the criterion for being identical to water is being constituted by H_2O; as well as with manifest criteria of identity, according to which the criterion for being identical to water is being watery.

This analysis seems to conflict with three, closely related, theses in Kripke's *Naming and Necessity*. The three theses are, (i) the *intuition thesis*: modal properties of a

referent can be derived from our modal intuitions. E.g., the intuition that Nixon might have lost the election implies that winning the election is a contingent property of Nixon; (ii) the *necessity of origin thesis*: objects have their origin necessarily. E.g., a table is necessarily originally made out of the hunk of wood from which it was actually made; and (iii) the *stipulation thesis*: possible worlds are "stipulated" rather than "discovered"; such stipulation is not in purely qualitative terms. E.g., we may stipulate a world in which *Nixon* lost the election; not just a world in which *a guy with such and such properties* lost the election. I believe Kripke is ultimately wrong in holding these three theses, although there is nonetheless something to say in these theses' favour. The following sections aim to establish this.

1.5.1 Kripke's intuition thesis

In accordance with Kripke's view, we may intuitively stipulate a world in which Humphrey, who actually lost the 1968 election, wins. However, we cannot intuitively stipulate a world in which Humphrey is a dog. Thus, our intuitions teach us that Humphrey is contingently a loser of the election, yet he is necessarily, i.e., essentially, a person. In other words, our intuitions teach us that losing the election is not part of the criteria of identity of Humphrey, whereas being a person is. Thus in general, we can learn about an object's essential and contingent properties – and thereby about its criteria of identity – from our intuitions about counterfactual worlds.[78] This is Kripke's intuition claim.

However, as our Martian story reveals, there may be different intuitions regarding criteria of identity, and hence about the modal properties of things; e.g., we took the stuff to be necessarily P and only contingently M, yet our Martian colleagues intuit that the stuff T is necessarily M, and not necessarily P as we took it to be. Thus, what our intuitions teach us is merely what our (inter-)subjective view about criteria of identity is, which need not necessarily reflect *reality* (if there is a fact of the matter here at all). In fact, as Della Rocca (2002) ingeniously shows, some of Kripke's own arguments seem to undermine his above intuition thesis. Della Rocca points to Kripke's discussion of our intuitions about some contingent identities. For instance, we have an intuition that in 1750, instead of finding out that water was H_2O we might have found out that water was something else, say XYZ. I.e., that there is a counterfactual world in which water is not H_2O, but rather XYZ, and hence that "water = H_2O" states a mere contingent truth. Yet

in such a case, says Kripke, our intuitions fail us; the counterfactual world we are talking about is *not* one in which *water* is XYZ, but rather a world in which something that *resembles* water, i.e., something which is *epistemically indistinguishable* from water, namely, a watery stuff, is XYZ. Thus, it is "the watery stuff is H_2O" which is contingent, and not "water = H_2O".[79] Della Rocca charges as follows: what makes the intuition that Humphrey might have won, different from the intuition that water might have been XYZ? Why in the water case are we not talking about water but rather about something merely similar to water, whereas in the case of Humphrey, we *are* talking about Humphrey, and not about someone merely similar to Humphrey?:

> Perhaps the Humphrey intuition, the intuition that Humphrey might have won, also involves a similar understandable mistake. Perhaps it is the case that this intuition, too, is really about a person similar to Humphrey, and thus not particularly about Humphrey himself. [80]

We may add: a Martian may intuit that a criterion for being identical to Humphrey is losing the election, and hence that there is no counterfactual world in which Humphrey won the election, since that will not be Humphrey but rather someone merely similar to him.[81] Likewise, they may intuit that water might have been XYZ; i.e., that there is a counterfactual world in which that same thing, i.e., an essentially watery thing, happens to be contingently constituted by XYZ rather than by H_2O. We may therefore conclude, with Della Rocca, that the mere rigidity of a term together with our intuitions cannot determine criteria of identity and essential properties. There is something "arbitrary", even "invidious",[82] in such intuitions.

1.5.2 Kripke's necessity of origin thesis

As just pointed out, Kripke claims that some of our modal intuitions fail us. One such interesting case applies to the origin of objects. E.g., we may have the intuition that the Queen of England might not have been the daughter of George VI, or that this table might have been originally made out of ice. These, says Kripke, are false intuitions. The reason they are false is to do with the necessity of origin thesis. According to that thesis, objects have their origin necessarily, e.g., it is necessary for Kripke to be the product of the union of a particular sperm and a particular egg.[83] Anyone whose origins are in a different gamete cannot be Kripke. Indeed, it seems not only necessary, but also sufficient. Anyone originating from this gamete is by that mere fact Kripke; no-one else could originate from this gamete. Similarly, it is necessary for a table to be originally

made from the hunk of wood from which it was made. Something is *this* table iff it is a table originally made out of that hunk of wood. (Although the table does seem to be allowed to undergo gradual part-replacement afterwards.) As in the case of water, we have some intuition that the table might have originated from a different hunk of wood, or even from ice, but, as we have seen above, Kripke dismisses this intuition as misleading: it is not *this* table that we are contemplating in this case but rather a qualitatively similar, i.e., an epistemically indistinguishable, table. Surely Della Rocca's above objection applies to this case as well. However, this particular thesis of the necessity of origin came under a different attack much earlier by Salmon in his "How *Not* to Derive Essentialism from the Theory of Reference".[84] Drawing on an argument of Chandler's,[85] Salmon raises a Sorites-style counterexample to Kripke's necessity of origin thesis:

Chandler asks us to consider a bicycle. We can agree with Kripke that *that* same bicycle might not have originally come into existence from entirely different parts. Yet, surely, it might have come into existence from the same parts, save for just one spoke, which might have been a different one. I.e., there is a counterfactual world W_1 in which *this* bicycle came into existence with one different spoke. But the bicycle on W_1 might have been originally made out of the same parts as it (i.e., the bicycle on W_1) was, save for *another* different spoke. i.e., there is a counterfactual world W_2 in which the bicycle on W_1 (which is, arguably, identical to the actual bicycle,) came into existence with another different spoke. If we go on like this, we end up with a counterfactual world W_n with a bicycle that came into existence from entirely different parts than our actual bicycle in W_0. Assuming the transitivity of identity, this seems to violate Kripke's principle of the necessity of origin.[86]

1.5.3 Kripke's stipulation thesis

Another closely related issue is the problem of transworld identity,[87] and Kripke's stipulation claim about counterfactual worlds.

On Kripke's view, we may simply stipulate a counterfactual situation in which Nixon lost the 1968 election. This presupposes that the guy in the counterfactual world under consideration is identical to actual Nixon. We, so to speak, "build it into" the counterfactual world. By contrast, Lewis's view of modality involves no such presuppositions. According to Lewis's view, rather than talking about a counterfactual

world in which *Nixon* lost the election, we should talk about a counterfactual world in which *some guy*, who is blond etc., lost the election; this requires to further establish whether that guy is Nixon or not. These two views represent two different paradigms of thinking about counterfactual situations. We can put the difference as follows. On Kripke's view, we may discuss an individual in a counterfactual situation by specifying both the individual's name and its properties.[88] On Lewis's view, by contrast, counterfactual worlds ought to be described in purely qualitative terms, i.e., we may only discuss individuals in counterfactual situations by specifying their properties, but we may not specify their names, unless we also provide criteria of (transworld) identity.[89] (This view of Lewis should not be confused with his other claims about possible worlds, namely his modal realism, and his counterpart theory.[90])

Kripke and Lewis's views came under mutual attack. As Kripke notes,[91] Lewisians object to his thesis by claiming that before referring to a guy in counterfactual situations as "Nixon", one needs first to establish that that guy is indeed Nixon, i.e., one needs to know Nixon's criteria of (transworld) identity. Kripke, in turn, charges that Lewis's view involves a misguided conception of counterfactual worlds, which treats such worlds as if they were "distant planets" at which we look through a telescope.

But what is entailed in this respect by our own proposed view?

If we stipulate a counterfactual world in which *Nixon* lost the election, as Kripke does, we thereby speculate that winning the election is contingent to Nixon, and thus that winning the election is not part of Nixon's criteria of identity.[92] Similarly, holding it impossible that Nixon might not have been a person, as Kripke does, speculates that being a person is essential to Nixon. (Recall that the intuition that there is such a possibility is, for Kripke, a false intuition, which in fact involves someone who is merely qualitatively identical to, yet ultimately distinct from, Nixon.) However, as we have just seen above, such speculations are ultimately unfounded. There is nothing to prevent, in principle, the converse claim from being true, namely, that winning the election *is* part of Nixon's criteria of identity, and that it is *not* part of his criteria of identity that he is a person. It is perfectly compatible both with our *a priori* knowledge, as well as with our *a posteriori* knowledge about the way the world looks and behaves. Thus, relative to strict metaphysical standards, our stipulations of counterfactual worlds ought not to involve both an object's name and the object's properties. E.g., we ought not to stipulate a counterfactual world in which *Nixon* lost the election. However, stipulations which

involve *only* names make no such metaphysical speculations about modal properties or about criteria of identity. Nor do stipulations that involve only properties. For instance, no such metaphysical speculations are involved in simply talking about Nixon in some counterfactual situation while mentioning no properties, or in talking about a blond guy losing the election in some counterfactual situation, without mentioning his name. As such, these latter two types of stipulations are perfectly valid, and should not be prohibited.

Hence, if we wish to avoid making metaphysical speculations about modal properties in general, and about criteria of identity in particular, our view prescribes as follows. We may discuss objects in counterfactual worlds in purely qualitative terms, i.e., without specifying their names. We may also discuss objects in counterfactual worlds whilst only specifying their names, i.e., without specifying their properties or criteria of identity (contrary to Lewis). What we should refrain from if we wish to avoid metaphysical speculations, is discussing objects in counterfactual worlds whilst specifying *both* their names and their properties (contrary to what Kripke in fact does).

We can now use these observations to address the mutual objections of the two views, starting with the Lewisians' objection to Kripke's view, and proceeding to Kripke's objection to Lewis's view.

Adherents of Lewis's notion of modality object to Kripke's talk about Nixon in counterfactual situations on the grounds that in order to talk about Nixon in a counterfactual situation we first need to be able to identify Nixon in such a world, i.e., we need to know Nixon's criteria of (transworld) identity. This criticism, I believe, is based on a confusion. Indeed, talk about Nixon in counterfactual situations presupposes *some* criteria of transworld identity for Nixon; yet it does not involve any such *specific* criteria of identity. When we talk about Nixon in a counterfactual situation, we talk about the guy who meets the criteria for being identical to Nixon, *whatever these criteria may be*. We need not *know* these criteria in order to say that. This, in fact, is one of the great advantages of using rigidified terms like "Nixon"; they allow us to talk about things before we have fully determined their criteria of identity (see Section 4.1 for discussion). We may talk freely about water before 1750, or about the clay statue David, and we may even *ask whether* they might or might not have such and such properties; by using "water", "David", as well as "Nixon", we still talk about water, David and Nixon in those counterfactual situations. As Kripke rightly notes:

> Don't ask: how can I identify this table in another possible world, except by its properties? I have the table in my hands, I can point to it, and when I ask whether it might have been in another room, I am talking, by definition, about it. I don't have to identify it after seeing it through a telescope.[93]

However, *pace* Kripke, the thing we are not entitled to is make *assertions* about these counterfactual situations in a way that involves both names *and* properties, e.g., that water might have been constituted by XYZ (or, alternatively, be non-watery); or that David might have been smashed into a ball (or, alternatively, come into existence from another piece of matter), or that Nixon might have lost the election (or might not have been an automaton). These assertions involve speculative, and hence ultimately illegitimate, presuppositions about these objects' criteria of identity. Thus Kripke's mere talk about Nixon in counterfactual situations is vindicated, although his talk about Nixon's losing the election in counterfactual situations is not.

Kripke's objection to Lewis's view is based on the claim that this picture treats counterfactual worlds as "foreign countries", which we observe through a "telescope".[94] Through this telescope, indeed, we can only describe the counterfactual world in purely qualitative terms: we can see a guy losing the election, but we cannot *see* that the guy is Nixon. It's not part of what one can observe. And this picture, Kripke says, is misguided:

> Intuitively speaking, it seems to me not to be the right way of thinking about possible worlds. A possible world isn't a distant country that we are coming across, or viewing through a telescope. Generally speaking, a possible world is too far away. Even if we travelled faster than light, we won't get to it.[95]

I believe we should agree with Kripke that the "foreign country" way of thinking about counterfactual situations is misguided. If only for the problem of the epistemic access to spatiotemporally distinct worlds, raised in the above quotation.[96] However, I also believe that Kripke is not right to identify Lewis's qualitative way of thinking with this "foreign country" way of thinking. We may accept Kripke's contention that counterfactual worlds should be stipulated rather than discovered, and yet maintain that one way of making such stipulations is by means of purely qualitative descriptions.[97] So the debate is not between stipulative vs. "foreign country" models, but rather between different ways of stipulating counterfactual worlds. Specifically, one way of stipulating counterfactual worlds would be purely qualitative; another way would be purely with names; and yet another involves both. As we have seen, the first and second ways are legitimate; the latter, strictly speaking, is not.

To recap, Lewis's view is right to suggest that counterfactual worlds may be introduced in purely qualitative terms. It is nonetheless wrong to ban altogether talk of counterfactual worlds that involves names. Kripke is right to suggest that discussion of counterfactual worlds may involve names. He is wrong to suggest that such discussions may *in addition* involve properties. He is also wrong to claim that a purely qualitative description of counterfactual worlds is misguided.

Finally, I would like to suggest some sense in which there is nevertheless some legitimacy, although not in a strict philosophical sense, to stipulate counterfactual worlds that involve objects with both names and specific properties (as Kripke suggests), despite the unfounded speculations about modal properties that such stipulation is bound to involve. In other words, I believe that Kripke is onto something when he insists that one can simply stipulate that "Nixon might have lost the election", without having to go into issues of criteria of identity. An analogy might help to make my point here. One may ask whether a certain war is morally just. This is a perfectly legitimate question. However, a no less legitimate question is whether wars exist at all, i.e., whether wars are a genuine part of the fabric of the world, as opposed to, say, an abstract construct that we merely project onto collections of some empirical facts. This latter question is simply on a more fundamental level of philosophical doubt. Surely, in order to answer whether wars are just or not, one need not necessarily *prove* first that wars exist; in the context of the question of moral justification, it seems legitimate to simply *presuppose* the existence of wars. Every discussion has to start somewhere, and thereby make some assumptions. However, the fact that in the context of the moral question one need not provide a proof for the existence of wars does not entail that the existence of wars cannot be independently questioned on another level.

Something similar, I believe, applies to discussion of counterfactuals. For some purposes, it makes perfect sense to discuss objects in counterfactual worlds by specifying *both* their properties and their names, e.g., a world in which *Nixon* lost the election, and not just some guy with such and such properties, who lost the election. However, this does not in itself rule out questions about cross-world identities and modal properties on another level. The latter questions are simply on a more fundamental level of discussion in the sense that we make fewer assumptions. E.g., we do not simply assume that the guy who lost the election is Nixon, i.e., that Nixon's victory is contingent to him. On that level of doubt we simply raise the *question* whether Nixon might have lost or not.

Endnotes

---

[1] We shall refrain in this dissertation from addressing the question of the existential status of counterfactual worlds, i.e., whether possible worlds are like other universes, spatiotemporally separated from one another (Lewis, 1986), or whether possible worlds are (*maximal*, i.e. complete) possible states of affairs, or, "ways the world might have been" (Plantinga, 1974). In other words, we shall not take sides in the modal realism vs. modal fictionalism debate.

[2] To address two-dimensionalist worries that might arise at this early stage, the possible worlds we discuss here are possible worlds considered as *counterfactual*, not as actual. In Jackson's terminology, we discuss here the C-extensions and C-intensions, not A-extensions and A-intensions (1998a, pp. 48-52); in Chalmers's terminology, we discuss here secondary intensions and not primary intensions (1996, p.57). See Section 1.1.3 for discussion.

[3] Putnam (1975)

[4] The example is based on Steward (1990). In her discussion, the pink solid H_2O has a different proton from the watery H_2O. The purpose of this qualification is apparently to address worries of supervenience. The worry is that physicalism requires that the manifest supervenes upon the physical and hence that no change in the manifest is possible without a change in the physical. Non-watery H_2O entails a change – although a modal one – in the manifest (from watery to non-watery) without a change in the physical (both are H_2O). Thus the proton difference guarantees that physicalism is not violated, while keeping both stuffs being H_2O. However, I believe that as we are considering a counterfactual world, this qualification is not necessary in our case; it is commonly accepted that, for example, there *might have been* immaterial souls etc., but there *actually* aren't. (However, as the possible world that Steward is considering is explicitly an *epistemic* one, "in the sense that, *for all we know* it could be the actual world" (p. 389, my *italics*), it does make sense to suppose that such a world should indeed conform to supervenience conditions, even if a *metaphysically* possible world, i.e., a counterfactual world, need not. For a discussion of this distinction see Section 1.1.3.)

[5] Frege (1892)

[6] Russell (1905)

[7] Overall, Kripke (1972/1980) takes the two views to be in essence the same. Although in one footnote he does acknowledge some difference between the two views (1980, p.27, n.4). For a discussion of the differences between the two views see Bach (online essay, http://userwww.sfsu.edu/~kbach/FregeRus.html). See also Endnotes 13 and 14 in Chapter 2.

8 Throughout this dissertation, I shall be using "refer" and "designate" entirely interchangeably. Specifically, I take the designatum/reference to always be an object(s)/stuff, and never a property.

9 Although Frege and Russell were more concerned with proper names like "Aristotle" than with natural-kind terms like "water", Kripke thought it was natural to extend their descriptive view to natural-kind terms. See Section 2.2.4.

10 By naming this theory of reference "descriptive", we diverge from various other uses of this title, which apply the term to views that take singular terms to be abbreviation of *rigidified* definite descriptions, e.g., widescopism (Dummett, 1973) or actualised descriptivism (Plantinga 1985, pp.82-87). These views will be expounded in detail in Chapter 2. In fact such views, according to the definitions provided here, are closer to the view described in the next paragraph.

11 Along with Kripke (1971; 1972/1980), Barcan-Marcus (1961), Putnam (1975), and Kaplan (1977/1989) advanced a similar view.

12 We can avoid for the time being the question of whether terms like "water" on Kripke's view are Millian, i.e., directly-referring, or rigidified description (this issue will be addressed in Chapter 2). The above description of Kripke's view is shared by both views, and at any rate, it is this shared aspect that seems to be the important one for him.

13 There is some controversy over whether singular terms designate their actual referent in worlds in which that referent does not exist (a view favoured, for instance, by Plantinga (1985, p. 84)), or not. (Following a suggestion made by Salmon (1981, p. 34), Kaplan (1989, p. 571) calls a term of the latter kind *obstinately rigid designator*, while Kripke calls it *strong rigid designator*). Kripke (1980, p.21, n.21) himself seems to be reluctant to commit to either view, and leaves the matter unsettled. I tend to side with Brock (2004, p. 285 n.13) on this matter, in thinking that nothing too important bears on this. At the least, I believe this matter to be of little concern to our present interests, and hence will ignore this sensitivity here.

14 Apparently, cross-world identity is problematic as it seems to be committed to modally extended individuals. For that reason, some regard the cross-world relation a *counterpart* relation rather than an identity relation (Notably, Lewis, 1971. More recent versions can be found in Stalnaker (1986); Sider (1996, 1999); and Forbes (1985; chs. 3 and 7)). In this dissertation, I shall ignore this sensitivity.

15 In his discussion, Kripke does not consider non-watery H_2O.

16 Admittedly, not the best choice of title, since this category includes rigidified descriptions too (to be explained in detail in Chapter 2). Salmon (1981) uses *nondescriptional*, but acknowledges that

such nondescriptional terms need not lack *any* descriptive content: While discussing terms that he takes to be nondescriptional, like proper names and indexical singular terms, he notes:

> The theory that … these singular terms are entirely nondescriptional, should not be understood as involving the thesis that no descriptional concepts or properties, other than haecceities, are ever semantically associated with names or indexicals. (p. 31)

An alternative title for such nondescriptive terms in the literature is *de jure rigid* terms (Kripke 1972/1980, p. 21 n. 21). E.g., Sidelle (1992) draws a distinction between *de jure* rigid terms that refer in other worlds "through identity" – this corresponds to our *nondescriptive* terms – and terms that refer in other worlds "through sense" – this corresponds to our *descriptive* terms. However, since this title requires quite a bit of explanation, I will spare it for a later stage (Section 1.3.1). For want of a better title at this stage, we shall stick to "nondescriptive" for a while.

[17] The way to accommodate such worlds would be to replace the +/- signs in the table, that apply to the whole world, with a specification of what is designated by "water" in that world.

[18] Rea (1997). For some good surveys of the problems, as well as analyses of the relations between them, see Geach (1962); Wiggins (1967); Lowe (1989); Rea (1997); and Noonan (2006)

[19] I use "David" rather than the commonly used "Goliath" in this context – following Gibbard's (1975) famous thought experiment – in order to distinguish our example from Gibbard's. The main difference is that in Gibbard, "Goliath" names only the statue, while "Lumpl" names only the lump of clay; in our discussion, by contrast, "David" names that thing, whatever it is, that has a statue-shape and a clay material constitution.

[20] The extension is based on the notorious puzzle of The Ship of Theseus (See Plutarch, *Life of Theseus*) – the wooden-planked ship that was maintained by replacing the worn planks with new ones until it underwent a complete part replacement. Hobbes's (1839) version of the original puzzle adds the reassembling of the discarded planks to form another ship, so that we end up with two ships.

[21] E.g., Chisholm (1976)

[22] It is assumed throughout this dissertation that identity relations require criteria of identity. Sometimes philosophers say that there may be *facts about* identity without there being *criteria of* identity. E.g., those who believe in *bare particulars* may claim that the clay statue, and, say, the clay ball, are identical in virtue of sharing the same bare particular, regardless of any properties that they may happen to have. I believe that this view may be easily paraphrased in terms of criteria of identity: "the criterion for being identical to A is having the same bare particular as A."

[23] "Criteria of identity" is sometimes used as an epistemic notion – i.e., the criteria for *recognising* whether two entities are identical or not – and sometimes as a metaphysical notion – i.e., as

necessary and sufficient conditions for the truth of an identity statement (Mackie, P., 2006). It is the latter sense that we shall be concerned with in this dissertation.

[24] The relation between sortals and criteria of identity is discussed in Chapter 3. For the time being, we can use the intuitive relation given here.

[25] Burke (1994/1997, p. 252)

[26] Burke (1994/1977, p. 253)

[27] Note, however, that this view does not violate Kripke's thesis of the essentiality of origin, according to which an object has its (material) origin necessarily (whether that thesis is true or not). All it allows is that David undergoes a part replacement; not that it is *originally* made of different parts. (Kripke's thesis is discussed in Section 1.5.2.)

[28] Note that this criterion of identity need not be in conflict with supervenience. For the mere fact (if it is a fact) that the manifest supervenes upon the material, does not by itself entail that the criteria of identity are material; it seems perfectly consistent to assume that the world is inhabited by, e.g., animals, that merely happen to have the material constitution they have – i.e., some specific collection of molecules – and yet at the same time that these animals supervene upon their material constitution. (See Section 3.4 for discussion)

[29] Wiggins (1980)

[30] Wiggins (1980; 2001)

[31] Quine (1961) (see also Endnote 3 in Chapter 3); Della Rocca (2002). Geach (1962) is talking about *sortal relativity* in this context.

[32] Quine (1963b); Lewis (1971; 1986: Ch. 4). This view can be seen as a four-dimensionalist version of the sortal-relativity view: there are different ways to unite stages into a whole; different descriptions/sortals/intentions will determine different such unities.

[33] Formally, $(x=y) \equiv \forall \Phi (\Phi x \leftrightarrow \Phi y)$

[34] Stevenson (1972, pp. 155-158). According to Stevenson's view, the relation between the clay statue and the other two objects is merely that of *equivalence*, and not of *identity*, as the latter relation must obey Leibniz's Law.

[35] For a detailed survey of proposed solutions see Rea (1997)

[36] For the sake of simplicity, let us ignore cases of partial change.

[37] Indeed, the modal version faces problems peculiar to trans-world identity which do not apply to identity over time – e.g., re-identifying the *same* lump of clay in a counterfactual world – but we

shall ignore this for the time being. We shall however elaborate on the dis-analogy between the temporal and the modal in Section 3.2.1.2.

[38] For some classical discussions, see Evans (1977); Stalnaker (1978); Kaplan (1977/1989) Jackson (1998a); and Chalmers (2002).

[39] The following is a variation on Evans's (1977) "Julius" example, where "Julius" is the name of whoever invented the zip. Evans called such names *descriptive names*.

[40] We shall ignore here uses of "actual" in natural language that do not have this semantic function. E.g., using "actually" merely to stress something, as in "I actually mean it".

[41] Some prefer talk about "possible worlds *considered as* counterfactual" vs. "possible worlds *considered as* actual" (Davies and Humberstone, 1981) rather than about counterfactual worlds vs. actual worlds; thereby suggesting that it is the same set of worlds considered differently, and not two separate sets of worlds. As our use of "possible worlds" is agnostic with respect to the question of realism, talking about "two sets of worlds" amounts to neither two distinct sets of worlds, nor to one such set considered in two ways. Therefore, for brevity, we shall stick to the latter terminology, interpreted neutrally.

[42] The two types of possibilities are commonly referred to in terms of the conceivable vs. the possible, or in terms of epistemic possibilities vs. metaphysical possibilities (e.g., Chalmers, 2002. Although Kripke clearly distinguishes epistemic from metaphysical possibilities, it is unclear whether his distinction indeed amounts to the one between possibilities of the actual vs. possibilities of the counterfactual). Another interpretation is in terms of *contexts of utterance* vs. *circumstances of evaluation* (Kaplan, 1977/1989). These interpretations of the two sets of worlds are quite contentious. However, the distinction between the two sets of worlds stands independently of these interpretations, and hence need not be committed to any of them. We shall thus use the neutral distinction, without the further epistemic vs. metaphysical load. See also Endnote 47.

[43] Similar conflicting intuitions that Kripke (1972/1980) raises are: that Nixon might have been an automaton, but also that he might not have (p. 47); that Queen Elizabeth might have been the daughter of someone other than George VI but also that she might not have (p. 112, p. 141); that heat might have been distinct from molecular motion, but also that it might not have (p. 140); that gold might have been an element with an atomic number other than 79, but also that it might not have (p. 140); and, that Hesperus might have been distinct from phosphorus, but also that it might not have (p. 141). However, Kripke himself accounts for the difference in terms of *epistemic* possibilities as opposed to *metaphysical* possibilities. Yet, as noted in Endnote 42, it is far from obvious that this distinction and the distinction between actual and counterfactual worlds amount to the same thing (*pace* Chalmers's strong urge).

[44] The function from counterfactual worlds to extensions is represented in the table by the horizontal rows. Supposing that the "real" actual world is the top left one, this intension is represented by the top row. The function from actual worlds to extensions is represented in the table by the diagonal.

[45] E.g., Chalmers (2002)

[46] Jackson (1998a)

[47] This last claim appears to conflict with Kripke's (1972/1980) view, (as well as Chalmers's (2002; 2006) view), according to which possibilities such as that the actual inventor might have been Leibniz (or Newton), *are* (merely) epistemic. We can respond to this as follows. There are different ways to understand "epistemically possible" here. One is as what is possible for *all* we know, i.e., for all we know *a priori and a posteriori*. On this interpretation, the space of possibilities diminishes as we make more *a posteriori* discoveries. Some of what Kripke says seems to support this interpretation. E.g., (Kripke, 1972/1980, p. 47) :

> Suppose Nixon actually turned out to be an automaton. That might happen. We might need evidence whether Nixon is a human being or an automaton. But that is a question about our knowledge. The question whether Nixon might have not been a human being, given that he is one, is not a question about knowledge … It's a question about, even though such and such things are the case, what might have been the case otherwise.

Thus, on this interpretation, the above passage suggests that for *all* we know – *a priori* and a *posteriori* – Nixon may turn out to be an automaton, but, if we discover that Nixon was in fact a person, then it is no longer a possibility that he actually wasn't. With respect to this conception of epistemic possibility, our above analysis stands in conflict; our analysis shows that even if we know beyond doubt that Nixon is actually human, it still makes perfect sense to consider a case in which Nixon was actually an automaton.

Another interpretation of knowing what is possible is as only *a priori* knowledge; on this interpretation, for something to be epistemically possible is (arguably) simply for it to involve no contradiction. In other words, *a posteriori* knowledge that we gain makes no difference to the space of possibilities. Others things that Kripke says seem to support this interpretation (1972/1980, p. 112):

> Could the Queen – could this woman herself – have been born of different parents from the parents from who she actually came? Could she, let's say, have been the daughter of Mr. and Mrs. Truman? There would have been no contradiction, of course, in an announcement that … she was indeed the daughter of Mr and Mrs Truman.

Likewise, it makes no contradiction to say that Nixon was an automaton. But it would still make no contradiction to say it even if we discovered that he was a person. Thus *this* latter interpretation is in no conflict with our above analysis.

[48] Note: criteria of identity of stuffs like water, and of individual objects like David, appear to have some important differences (see Section 3.2.1.2). However, it remains true that the intension of terms – whether of mass-nouns that designate stuffs like water, or of singular terms that designate individual objects like David – is dependent, at least partially, on the criteria of identity of their actual referent. Whatever those criteria of identity are, i.e., regardless of any such differences.

[49] These factors are slightly different from what was presented in the introduction. We shall arrive at the formulation as presented in the introduction at the end of Chapter 2 (see Section 2.4).

[50] Some may be already worried at this stage about Kripke's stipulation claim – i.e., that a counterfactual world in which Newton invented the Calculus first is *stipulated*, and hence that there is no worry about criteria of identity. However, we shall have to postpone discussion of this worry to Section 1.5.3.

[51] Indeed, in the case of descriptive terms, the intension is dependent on neither of the other two criteria (we shall soon see why exactly, in Section 1.2.1). However, I believe that we are nonetheless right to assert the general claim, namely, that intentions do depend on the three factors, based on the following analogy. A formula of the form $Z=XY$ expresses the dependence of Z on X and Y. However, surely when X equals 0, Z is no longer dependent on Y; whatever Y is, Z will turn out 0. Yet this *special* case does not seem to undermine the *general* dependency of Z on Y. Similarly, it is true that when the value of the factor "the semantic function of the term" is "descriptive" the intension is independent of the other two factors. But this does not seem to undermine the general dependency of the intension on those factors.

[52] A revised version of this section was published as Pelman (2006)

[53] True, the task of re-identification of entities in other possible worlds is at odds with Kripke's famous contention that such identities are *stipulated* rather than discovered (1972/1980, p. 46). Although I agree with Kripke that possible worlds *may* be thus stipulated, I am nonetheless not convinced that an alternative, purely *qualitative* introduction of possible worlds – i.e., in which no such stipulation of identity is involved – is not as permissible. However, as Kripke's claim is related to his specific essentialist assumptions, we shall postpone the discussion of this matter to Section 1.5.3, after we've discussed these essentialist assumptions of his.

[54] Again, to avoid confusion: our mission is to determine the intension of "T" on other possible worlds *taken as counterfactual*, and not as actual (see Section 1.1.3).

[55] As defined in Section 1.1.1.

[56] As defined in Section 1.1.1.

[57] As in the pink-solid H_2O case discussed in Section 1.1.1, one may object that this option violates physicalism. Our response to the objection applies in this case as well (see Endnote 4).

[58] For the sake of simplicity, we suppose homogeneity within a world, i.e., we ignore here worlds that contain more than one of these stuffs.

[59] There are various other metaphysical alternatives in another sense as well (some of which were listed in Section 1.1.2): E.g., that the world contains both types of stuffs, namely P's (and Q's) as well as M's (and N's), that simply coincide; or, that what the stuff is is sortal-relative; and more. However, I believe that such alternatives should not be too difficult to incorporate in the final analysis, once such an analysis is attained.

[60] Indeed the last two possibilities are not very probable, since they suggest a different manifest property from M, and given that we seem to perceive such properties quite directly we are most likely not to be mistaken about them. However, in order for our formula to be general, we'd do better to consider those options as well, as it will make the formula applicable to cases in which the property M is not manifest and directly perceived, but rather has to be discovered.

[61] Kripke, who introduces the term, actually uses only the first part of the above definition, namely, that a rigid term designates the same referent in all possible worlds; yet much of what he says indicates the other part as well, i.e., that a rigid term designates nothing but that referent (Cf. Stanley (1997)).

[62] Kripke (1980, p.21, n.21)

[63] Kaplan (1977/1989), p. 493. Others who stressed the distinction between rigidified (i.e., *de jure* rigid) and non-rigidified designation instead of the one between rigid and nonrigid designation are Smith (1984), and Sidelle (1992).

[64] Sidelle (1992) makes a similar point. He introduces a descriptive term "Prez", whose descriptive content is "The President of the U.S.", and a corresponding entity Prez, that

> is [in 1992] composed by George Bush, was composed in 1790 by George Washington, and in 1863 by Abraham Lincoln. In some other worlds, Prez is [in 1992] composed by Michael Dukakis, and in others by Elizabeth Dole.

(p. 417)

In other words, the criterion of identity of Prez is occupying the role (i.e., the office). By contrast, the criterion of identity of an occupant of the presidential role, say, of Lincoln, is that of a person, i.e., a criterion of personal identity. Sidelle's conclusion about rigidity is the following:

> On a permissive ontology [i.e., ontology that accepts unusual entities like Prez], descriptions used to refer through sense [i.e., descriptive terms like "the President of the U.S."], without

> further specification, are indeterminate between rigidly denoting unusual entities [like Prez], and nonrigidly denoting the particulars that constituted them at a given time [like Lincoln].

(p. 418)

There is, however, a small, yet important, difference between our discussion and Sidelle's. Sidelle's conclusion is based on *one* "permissive" ontology, that admits *both* "usual" entities, like Bush, *and* unusual entities, like Prez. Both these entities simply happen to Spatio-temporally overlap in the actual world, and hence the descriptive term "the President of the U.S." may pick out either: if it picks out Bush in the actual world, Dukakis in another, etc., it turns out nonrigid; if, however, it picks out Prez in the actual world, Prez in another counterfactual world, etc., it turns out rigid (*de facto*). In the above discussion, by contrast, we assume no such spatio-temporal overlap; rather, we consider *different* ontologies: the same descriptive term will come-out rigid on some such ontologies and nonrigid on others.

[65] The difference between the two is very important, and I believe it to be at the heart of much of the debate between essentialists and anti-essentialists. See Section 1.5.1 for some related discussion.

[66] E.g., Kripke (1972/1980, p. 128-9)

[67] See Section 1.1.2.

[68] Chalmers (2006)

[69] For a more realistic mistake consider the following: there may be a more fundamental nature to the liquid that we call "water", say, some FN. Thus, a stuff is identical to our liquid iff it is FN. Now, although in actuality all H_2O is FN and vice versa, still, in principle, it is possible for something to be FN and not H_2O and vice versa, viz., some counterfactual worlds contain FN which is not H_2O, and some others contain H_2O which is not FN. Consequently, a rigidified "water", under material criteria of identity, will not designate non-FN H_2O, and thus will not designate H_2O in all counterfactual worlds.

[70] Shortly after the above quotation, Chalmers writes:

> if the H_2O-world turns out to be actual (as it has), then 'water' will have a Kripkean intension that picks out H_2O in all worlds; but if the XYZ-world turns out to be actual (as it has not), then "water" will have a Kripkean intension that picks out XYZ in all worlds.

Thus Chalmers, like Kripke, presupposes material criteria of identity.

[71] Kripke (1972/1980, p. 143)

[72] Locke (1689, II xxvii 15)

[73] Different versions of this view are divided on this relation, namely, whether it is just 'if', or 'only if', or, indeed, 'if and only if'.

---

[74] Locke (1689), who introduced the prince/cobbler thought experiment, in fact used it to argue for such a psychological criterion. Among contemporary defenders of some version of that view, Olson (2002) lists Garrett (1998), Johnston (1987), Lewis (1976), Nagel (1986, p. 40), Noonan (1989), Nozick (1981), Parfit (1971; 1984, p. 207), Perry (1972), Shoemaker (1970; 1984, p. 90; 1997; 1999), and Unger (1990, ch. 5; 2000). As popular as it may seem, it has it's strong opponents. Among such opponents, Olson lists Ayers (1990, pp. 278-292), Carter (1989), Feldman (1992), Mackie D. (1999), Olson (1997), van Inwagen (1997) and Williams B. (1956-7; 1970).

[75] Williams B. (1956-7) is a known representative of such a view. Olson (2002) mentions Thomson (1997), and Ayer (1936, p. 194) as advocates of some versions of that view.

[76] E.g., Snowdon (1990)

[77] As in the case of David, there are, of course, quite a few other metaphysical views, e.g., a four-dimensionalist view according to which there is a temporal stage which is shared by two different collections of stages – a psychological-collection and a bodily-collection (Hudson 2001, ch. 4) and more. However, again, we are not so much concerned here with taking sides in this debate; rather, we look for the lessons that can be drawn from the debate to the determination of intensions.

[78] Kripke (1972/1980, p. 42)

[79] Kripke's discussion includes a long list of such false intuitions (Kripke, 1972/1980, p. 142-4). See also Endnote 43.

To recall, in Section 1.1.3 we have shown how handy the distinction between actual worlds and counterfactual worlds comes in reconciling such seemingly conflicting intuitions: E.g., the intuition that water *might* have been XYZ can be simply accounted for by the existence of a possible world considered as *actual* in which water is XYZ, whereas the intuition that water *might not* have been XYZ can be accounted for by the absence of a *counterfactual* world in which water is XYZ, since in all *counterfactual* worlds (that correspond to the "real" actual world) water is H_2O. However, as noted above (Footnote 42) it is not clear that Kripke's distinction between epistemic and metaphysical possibilities is equivalent to the one between actual worlds and counterfactual worlds.

[80] Della Rocca (2002, p. 239). Della Rocca's positive claim, however, is that criteria of identity are determined by the way the term is described, or thought of, or referred to. As our discussion shows, this need not be the case. It may be that there are mind-independent criteria of identity. In which case, these are independent of the way the object is thought of, referred to, or

described. Thus, one may be right in the way they think about the object, if the criteria of identity they ascribe to the object fit reality, or wrong if they don't. If, however, there is no fact of the matter, i.e., there are no mind-independent criteria of identity out there, then Della Rocca is of course right. But I can see no way in which we, humans, can prove this. It simply appears to be beyond our ken. So our discussion joins Della Rocca only as far as claiming that *specific* criteria of identity (whether mind-independent or not) should not be derived from rigidity and intuitions.

[81] Sidelle's (1992) Prez is in fact a version of such an object. (See Endnote 64)

[82] As Della Rocca notes (2002, p. 230) a version of this "arbitrariness" worry goes back to Quine (1961/1976, p. 184).

[83] Interestingly, contemporary biologists indeed delimit kinds historically, i.e., according to evolutionary history. In other words, they take the historical origin to be necessary (and, indeed, sufficient) to the kind (LaPorte, 2004). This seems to fall in line with Kripke's intuition about individuals. However, when it comes to natural-kinds, like tigers, Kripke seems to be more in favour of *internal structure* (presumably DNA) as the essential feature, rather than the historical origin.

[84] Salmon (1979)

[85] Chandler (1976). Indeed, Chandler's main concern is epistemological (in particular, he wishes to claim that the accessibility relation between possible worlds is not transitive). However, Salmon (1979, pp. 722-5) uses Chandler's argument to dismiss Kripke's metaphysical thesis of the necessity of origin.

[86] Salmon's positive conclusion is, however, that rigidity does not entail essentialism. As our formula shows, this is one step too far. We can agree with Salmon that the necessity of origin does not follow from rigidity. Moreover, we can even accept that no other *specific* necessary properties can be derived from rigidity. But this is as far as we should go. We can still hold that rigidity does entail that referents have *some* essential properties, without committing to what these essential properties are (whether we know them or not; and whether they are mind-independent or not). As mentioned earlier, a rigid designator, by definition designates the *same* referent in all counterfactual worlds; the notion of "the same" involves criteria of identity; and criteria of identity involve essential properties. Hence, the very notion of rigid designation does presuppose some (although not specific) essential properties.

[87] For a discussion about the special problems of transworld identity of individuals (as opposed to transworld identity of stuffs and kinds) see Section 3.2.1.2.

[88] Kripke (1972/1980, p. 44)

[89] E.g., Lewis (1986)

[90] Even though some mutual support between these three Lewisian theses can be shown, the three stand independently of one another. E.g., counterfactual worlds described qualitatively, can have identity relations between their inhabitants – for instance, between the brown haired guy who actually won the election, and the blond guy who won the election in that world; or counterpart relations – between the same two guys. The same holds for worlds described with name-tags: actual Nixon and counterfactual Nixon may be held identical or counterparts.

[91] Kripke (1972/1980, p. 42)

[92] For a more detailed discussion about the relation between criteria of identity and modal properties, see Chapter 3. Note however that the above metaphysical commitment is the *only* one made by such stipulation. For instance, such stipulation is not committed to the fact that Nixon is necessarily a person.

[93] Kripke (1972/1980, pp. 52-3)

[94] Similarly, Kaplan (1967/1979, p. 93) mentions "Jules Verne-O-scope".

[95] Kripke (1972/1980, pp. 43-4)

[96] See also Plantinga (1974, p. 94)

[97] Attributing the objection to such stipulation to Kripke is further supported by Kripke's discussion of probability in his introduction to *Naming and Necessity* (1980, pp. 15-20). In his discussion, Kripke points to the following disadvantage of introducing counterfactual worlds in purely qualitative terms. With respect to two dice, there are 36 different possibilities as regards the numbers they show on top. In other words, there are 36 different possible worlds in this respect. However, on the qualitative view, the following two possible worlds are qualitatively identical: (1) die A: 6; die B: 5; and, (2) die A: 5; die B: 6. It thus follows that, on the purely-qualitative paradigm, the two worlds are in fact the same world. And the same applies to 2/4 vs. 4/2 etc. Kripke concludes as follows:

> The 'possibilities' simply are not given purely qualitatively (as in: one die, 6, the other, 5). If they had been, there would have been just twenty-one distinct possibilities, not thirty six. And the states are not phantom dice-pairs, viewed from afar, about which we can raise epistemically meaningful questions of the form 'Which die is that?'

2 INTENSIONS: SOME SEMANTICAL ASPECTS

Overview

Our general formula of intensions includes three variables: (a) the semantic function of the term; (b) the actual referent; and (c) criteria of identity. This chapter focuses on the first of these three. In what follows, we shall examine more closely the relation between intensions and the semantic function of terms. The main thesis of this chapter is a proposed classification of semantic functions, which provides a more fine-grained insight into the mechanism of reference across possible worlds. This classification is based on two distinctions: (a) designation in the actual world vs. designation in counterfactual worlds; and (b) designation via fitting a descriptive content vs. designation regardless of such content. These two distinctions give rise to a fourfold classification of semantic functions. Based on this classification, we define three basic semantic functions (the fourth turns out to be un-instantiated).

The proposed classification is put to work in two ways.

First, the classification is being used to compare and contrast competing views about the semantics of *natural-language term*s. Specifically, we will be looking at: proper names, definite descriptions, complex demonstratives and natural-kind terms. Different views ascribe different semantic functions to these types of term. Thus, altogether, there appear to be many semantic functions on offer. However, it will be argued that all these different semantic functions boil down to merely three basic semantic functions, which were defined by our proposed classification.[1] I.e., it is argued that every proposed semantic function for a certain type of term is, ultimately, equivalent to one of these three basic functions. As this part involves presentation of the different views, it will naturally have some literature-review air to it.

Next, we illustrate the unique pattern of intensions that characterises each of the three basic semantic functions. These unique patters are used to clearly illustrate the differences between the proposed formal semantic functions, and thereby, between the natural-language terms.

The second use of the proposed classification is for fine-graining the intensions formula. We shall see that our initial dealing with semantic functions, whereby we distinguished descriptive from rigidified terms, was exclusively concerned with designation in *counterfactual* worlds. The classification offered in this chapter adds the aspect of taking into account designation in the *actual* world. Surprisingly, this added input will prove to be closely linked to the second factor in the formula – the actual referent. In particular, it will be shown that the factor of the actual referent can be replaced by two other factors that determine it: designation in the actual world and the state of the actual world. In other words, a more fine-grained version of our formula includes four variables.

2.1 A Fourfold Classification of Semantic Functions

So far, we have been dealing with two semantic functions: descriptive and rigidified. A descriptive term was defined as a term that has a descriptive content and that designates, with respect to every counterfactual world, that which fits that content. A rigidified term was defined as a term that designates, with respect to every counterfactual world, the same referent that it designates in the actual world. Hence the distinction between rigidified and descriptive terms is concerned with designation in counterfactual worlds; specifically, terms may designate in counterfactual worlds via fitting a descriptive content or regardless of such content. But what about the actual world? Can't we apply the same distinction to the actual world, namely, isn't it the case that a term may designate in the actual world via fitting a descriptive content or regardless of such content?

This idea seems to fit some strong intuitions about two types of rigidified terms. With respect to counterfactual worlds, all rigidified terms designate the same referent that they designate in the actual world. However, how do such terms pick out their referent in the actual world? Kripke proposes two such ways:

> An initial baptism [i.e., designation in the actual world] takes place. Here the object may be fixed by ostension, or the reference of the name may be fixed by a description.[2]

Naming by ostension is pretty much self-explanatory. For the latter case, Kripke provides the instructive example of the name "Neptune"; it was fixed by the astronomer Leverrier to name "the planet which caused such and such discrepancies in the orbits of certain other planets".[3] Following the "baptising" event – whether by ostension or via a

descriptive content – all rigidified terms designate that same baptised object in all counterfactual worlds. For example, if "water" is indeed a rigidified term, then whether "water" baptised H_2O directly by ostension, or via the descriptive content "the watery stuff", when it comes to counterfactual worlds, "water" designates H_2O in all those worlds. In other words, although all rigidified terms designate the same referent in all counterfactual worlds, there are indeed two ways of picking out their referent in the actual world. The first is directly, regardless of any descriptive content; the second is via fitting a descriptive content.

Let us call rigidified terms that designate in the actual world directly, *directly-referring* terms (following Kaplan[4]), and terms that designate in the actual world via fitting a descriptive content *reference-fixing* (following Kripke[5]). Accordingly, we may say that the category of rigidified terms has two sub-categories: directly-referring terms and reference-fixing terms.

These observations lead to the following structure. We have two basic distinctions: (a) designation in the actual world vs. designation in counterfactual worlds; and (b) designation via fitting a descriptive content vs. designation regardless of descriptive content. Each referring term designates in both the actual world and in counterfactual worlds. In each of these realms, the term may designate by virtue of fitting its descriptive content or regardless of such descriptive content. This generates the following fourfold classification of semantic functions:

Designation in the actual world	Designation in counterfactual worlds
By fitting a descriptive content	By fitting a descriptive content
By fitting a descriptive content	Regardless of descriptive content
Regardless of descriptive content	By fitting a descriptive content
Regardless of descriptive content	Regardless of descriptive content

Table 2.1

Table 2.1 generates four categories. The three semantic functions discussed above, namely, descriptive, directly-referring and reference-fixing, each fit into one of these categories: a descriptive term is a term that designates by virtue of its descriptive content in both the actual world and in counterfactual worlds. It therefore belongs in the first category. Directly-referring terms designate regardless of descriptive content in both the actual and counterfactual worlds. They therefore belong in the fourth category. A

reference-fixing term designates in the actual world by means of its descriptive content, whereas in counterfactual worlds it designates regardless of descriptive content, i.e., it simply designates the same referent that it designates in the actual world. It therefore belongs in the second category. Note that there is one more possible category (the third one in our table): a term that designates regardless of descriptive content in the actual world, and by virtue of fitting its descriptive content in counterfactual worlds. Yet I could find no example for such a term.[6] In sum,

	Designation in the actual world	Designation in counterfactual worlds	
1	By fitting a descriptive content	By fitting a descriptive content	**Descriptive** *non-rigidified*
2	By fitting a descriptive content	Regardless of descriptive content	**Reference-fixing** *rigidified*
3	Regardless of descriptive content	By fitting a descriptive content	**---**
4	Regardless of descriptive content	Regardless of descriptive content	**Directly-referring** *rigidified*

Table 2.2

(We can see now why the title "nondescriptive" is not very suitable to cover the group of rigidified terms: some rigidified terms are reference-fixing; they have a descriptive element to them, namely, designating in the actual world via descriptive content. However, calling such terms "descriptive", as various philosophers do,[7] is also misleading, as it may confuse this category with the "pure" descriptive one. Calling these terms "rigidified" thus better serves its purpose.)

2.2 Natural-language terms – the Controversy over their Semantic Function

Our categories of semantic functions – descriptive, reference-fixing and directly-referring – are defined *formally*. They are not, as such, categories of terms in natural languages. As regards natural-language terms, it is a matter of controversy what their semantic functions are. However, it is proposed that the above classification of formal types of terms, based on the distinction between designation in the actual world vs. designation in counterfactual worlds, is essential to the analysis of natural-language terms.

We shall examine four types of referring terms: definite descriptions; proper names; complex demonstratives and natural-kind terms. The first three are types of singular terms; the latter is a type of a general term. Our discussion of the first type of terms (as it happens, proper names) will be relatively elaborate, whereas discussions of subsequent terms will, to a large extent, build on ideas stated in this first discussion, and will thus be considerably shorter.

2.2.1 Proper names

Let us briefly sketch some of the arguments that led to the main views about proper names.

Since it is natural to see the different views about the semantics of proper names as developing historically, in the form of responses to each view's predecessor, we shall deviate from our custom of not engaging in the arguments for or against views, and will briefly present them. Nevertheless, we shall, as usual, refrain from taking sides.

2.2.1.1 The main views about the semantics of proper names

2.2.1.1.1 *Millianism and its discontents*

Mill famously held that,

> ...proper names are not connotative: they denote the individuals who are called by
> them; but they do not indicate or imply any attributes as belonging to those
> individuals.[8]

Thus, according to Mill, names are something like "labels", which simply apply to their referents directly, and any descriptive content is left completely out of the process.[9] As a theory of meaning, Millianism is the view that the *meaning* of a name is simply its bearer. This invited the following main objections to Millianism, initially raised by Russell:[10] (a) Frege's puzzle: if Millianism is true, then the terms "Hesperus" and "Phosphorus" have the same meaning, and hence are substitutive *salva veritate*. It follows that "Hesperus is Phosphorus" amounts just to "Hesperus is Hesperus". But the latter, unlike the former, is trivial and uninformative; (b) Empty names: the sentence "Santa Claus lives in the North Pole" seems meaningful and false, yet if Millianism is true, then "Santa Claus" refers to nothing, i.e., means nothing, and hence the sentence is just meaningless; (c) Negative existentials: "Vulcan does not exist" seems meaningful and true. However, according to Millianism, the sentence comes out absurd: it denies the existence of

something the existence of which it presupposes; and (d) Propositional attitude reports: the sentence "Fred believes that Cicero, but not Tully, was Roman" seems to express a mistaken, yet nonetheless rational, belief. Millianism will simply render it a self-refuting, and hence irrational, belief.

Millians commonly respond to these objections by recourse to Gricean *speech-acts*.[11] Thus, a sentence in a speech act may *pragmatically communicate* an informative, or meaningful, or rational proposition, whereas at the same time, what it *semantically expresses* is trivial, or meaningless, or absurd, or irrational.[12]

Many were not impressed by these types of solutions. Thus these problems (as well as other considerations) led to the development of descriptive accounts of proper names.

2.2.1.1.2 *Frege-Russell descriptivism*

Frege (1892) held that names have sense, and that they designate their referents via their sense. Russell (1917) held that a proper name is a disguised definite description, and that it designates its referent by means of fitting the definite description.[13] Thus, both Frege and Russell held that proper names designate their referents not directly, but rather via a mediating descriptive content. By virtue of this common feature, and despite some important differences,[14] their views are commonly referred to as the "Frege-Russell" view.[15] Thus (to follow Kripke's example), on a Frege-Russell descriptivist view, "Aristotle" abbreviates something like "The last great philosopher of antiquity" (or a conjunction of some such descriptions). Such a descriptivist view easily avoids the above problems that Millianism faces: by substituting each name with the description that it abbreviates, all the above objections are barred.

However, this view, as Kripke showed, is not free from problems as well. Kripke launches two main arguments against this view, both based on the fact that although Aristotle actually fits the above description, he might not have fitted it; e.g., Aristotle might have chosen a different career and not become a philosopher at all. In other words, it is a contingent fact about Aristotle that he was a philosopher.

The first argument – sometimes called *unwanted necessity*,[16] or the *epistemic argument* – is the following. Consider the sentence:

(1) Aristotle might not have been a philosopher.

Given the contingency of the fact that Aristotle was a philosopher, (1) is true. However, if Aristotle abbreviates "the last great philosopher of antiquity", then (1) amounts just to

> (2) The last great philosopher of antiquity might not have been a philosopher.

Yet surely any person who satisfies the relevant description is thereby a philosopher, and hence (2) seems false. So descriptivism seems false.

The second argument – sometimes called *the modal argument* – states that if Aristotle had not become a philosopher, the description "the last great philosopher of antiquity" would apply to Plato, and thus descriptivism implies that "Aristotle" would just name Plato, which seems no less problematic for the theory.[17] These problems led to the development of the New Theory of reference.

2.2.1.1.3   *The New Theory of reference*

The central feature of the New Theory is that names are (*de jure*) rigid designators. A name is designed to designate the same referent in all counterfactual worlds. Kripke, an exponent of the theory, is less concerned about designation in the actual world. As was stated above (Section 2.1) This may be either in a direct way, i.e., not mediated by descriptive content, or by fitting a description.

As a mechanism to achieve the rigidity of names, Kripke (as well as Putnam) offers a causal theory. A name is causally "hooked" onto its referent (in the above baptising event), and then that link is preserved by a process of "reference borrowing", i.e., passing the name from one speaker to another in the community. Thus when using the name, speakers designate the original referent by virtue of the causal chains that track it, regardless of any description that that referent may fit:

> 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.[18]

The strength of the New Theory is that it conforms to some strong intuitions, by rendering names rigid designators. Accordingly, Kripke's main criticism of the Frege-Russell type of descriptivism is that it entails that names are nonrigid designators. Thus, in order to avoid Kripke's criticism, descriptivists need to find a way of rigidifying the descriptions that names are said to abbreviate, so that such descriptions designate, with

respect to all counterfactual worlds, the same object that they designate in the actual world.

2.2.1.1.4   *Actualised descriptions*

One way to rigidify descriptions, and thus to avoid Kripke's criticism of descriptivism, is to actualise them, i.e., to hold names to abbreviate not just descriptions, but *actualised* descriptions.[19] As Kaplan puts it,

> A Fregean who takes the name "Aristotle" to have as its sense the pupil of Plato and teacher of Alexander the Great need only add something like actuality to the content in order to account for the rigidity of proper names.[20]

Such an actualisation device will avoid Kripke's arguments against a Frege-Russell descriptivism. An actualised description referring to Aristotle designates Aristotle not only when discussing the actual world, but also when discussing counterfactual situations, in which Aristotle was not a philosopher. Thus, whereas (2) may sound false, it makes perfect sense to say that,

> (3) The *actual* last great philosopher of antiquity might not have been a philosopher.

In sum, since the "actualised" operator guarantees the rigidity of a description, a name that abbreviates such a description is thereby rigid, and avoids Kripke's objections.[21]

2.2.1.1.5   *Widescopism*

Another attempt to avoid Kripke's criticism, famously suggested by Dummett[22] (inspired by Russell[23]) is by using *scope-distinction*. Definite descriptions, when embedded in intensional context (e.g., modal, temporal, or propositional attitude contexts) are subject to a *de re*/*de dicto* ambiguity.[24] Let me explain this distinction. The above

> (2) The last great philosopher of antiquity might not have been a philosopher

is ambiguous between a *de dicto* reading, and a *de re* reading. On the *de dicto* reading, the sentence expresses the inconsistent claim according to which it is possible that there was someone who was the last great philosopher of antiquity and was not a philosopher; on the *de re* reading, the sentence expresses the consistent claim that it is possible that a

certain person, Aristotle (who happened to be the last great philosopher of antiquity), was not a philosopher.

It is customary to analyse this ambiguity in terms of scope.[25] On this analysis, the *de dicto* reading of (2) is

> (2') it is possible that: there is someone such that, that someone was the last great philosopher of antiquity and was not a philosopher

whereas the *de re* reading of (2) is,

> (2") there is someone who was the last great philosopher of antiquity, and it is possible that: *that* someone was not a philosopher.

Thus, on the interpretation (2') of (2), the existential quantifier "there is someone" is within the scope of the modal operator "it is possible that", and thus has a *narrow* scope; this is a *de dicto* reading of (2), which, in this case, is inconsistent. On interpretation (2") of (2), by contrast, the existential quantifier is outside the scope of the modal operator, and thus has a *wide* scope; this is a *de re* reading of (2), which, in our case is consistent. Thus the ambiguity is between a narrow and a wide scope of the existential quantifier relative to the intensional operator (which, in this case, is a modal operator).[26]

Kripke's criticism is based on reading (2) as expressing the *de dicto* statement (2'); thus read, the sentence is indeed false. However, when taken to express the *de re* statement (2"), (2) comes out true, and as a consequence Kripke's objection is blocked. Hence, all the descriptivist needs to do in order to avoid Kripke's criticism is restrict the existential quantifier involved in the relevant definite description to have a wide-scope.

The problem with this solution is, that names are frequently embedded in no context, let alone a modal context, and in these cases no scope analysis is available.[27] Adherents of such *widescopism* propose further modification to address this problem.[28]

2.2.1.2   Analysis of views on proper names in terms of the fourfold categorisation

In the outset of this chapter, we introduced a fourfold categorisation of semantic functions that gave rise to three formal semantic functions (the fourth logical possibility was deemed un-instantiated); descriptive, reference-fixing and directly-referring, as detailed below.

Designation in the actual world	Designation in counterfactual worlds	
By fitting a descriptive content	By fitting a descriptive content	**Descriptive**
By fitting a descriptive content	Regardless of descriptive content	**Reference-fixing**
Regardless of descriptive content	By fitting a descriptive content	---
Regardless of descriptive content	Regardless of descriptive content	**Directly-referring**

Table 2.3

In the previous section five views on the semantics of proper names were presented: the Millian view, the Frege-Russell view, the New Theory, actualised descriptions, and widescopism. The purpose of this section is to analyse these views by using our proposed categorisation. In order to thus analyse them, we will need to extract from each view its implications with regard to designation in possible worlds, i.e., actual and counterfactual. This will prove rather straightforward in the case of the first four views. Widescopism will require a slightly longer discussion.

2.2.1.2.1   Analysis of Millianism

According to Millianism, at least with respect to the actual world, a proper name designates regardless of descriptive content. How does it designate in counterfactual worlds? Although, understandably, Mill did not consider counterfactual worlds, direct-reference theorists, who share the spirit of Mill's view, did. With respect to the sentence

(4) John is suspicious

Kaplan formalises it as

(5) <John,P>

where P is a function "which assigns to each possible world w and each time t the set of all those individuals in w which, in w, are suspicious at t."[29] Sentence (5), says Kaplan,

> will determine that function F which assigns Truth to w and t if and only if John is a member of $P(w,t)$. If John is an individual of w at time t (i.e., John exists in w and is alive at t) but is not a member of $P(w,t)$, then $F(w,t)$ is falsehood.[30]

Thus, the guy who is checked for suspicion in each world, *actual or counterfactual*, is John, *regardless of any description* that may apply to him. "John" therefore designates just John in both the actual world and in all counterfactual worlds.

If this analysis is sound, then with respect to our fourfold categorisation, the above view of proper names belongs in our fourth formal category – that of *directly-referring* terms.

2.2.1.2.2   *Analysis of Frege-Russell descriptivism*

Frege and Russell took names to designate their referent via a mediating descriptive content. Thus, with respect to the actual world, a proper name designates that which fits its descriptive content. What about counterfactual worlds? Like Mill, Frege and Russell did not discuss designation in counterfactual worlds. Kripke, however, did (as noted in Chapter 1). He took the Frege-Russell view to imply that names designate in the same way in counterfactual situations. Namely, a name designates, with respect to each counterfactual world, that which fits its descriptive content in that world. Thus, on this interpretation of the Frege-Russell view, a proper name designates, with respect to *both* the actual and to each counterfactual world, that which fits the name's descriptive content.

If this analysis is sound, then the Frege-Russell descriptivist view of names falls under our first formal category, the *descriptive* category. In other words, on the Frege-Russell view, proper names are descriptive terms (according to our formal definition of "descriptive").

2.2.1.2.3   *Analysis of the New Theory's view*

As Kripke states his view in terms of possible worlds, there is no need in this case to extrapolate from the actual to the counterfactual, like we did in the above two cases. With respect to counterfactual worlds, Kripke explicitly takes names to designate the same referent that they designate in the actual world. As for the actual world, some names designate by directly "labelling" their referents, whereas others (like "Neptune") designate by fitting a description. Thus, with respect to our fourfold categorisation, some names are directly-referring, i.e., belong to our fourth category, whereas others are reference-fixing, i.e., belong to our second category. No proper name belongs to either the first or third category. In other words, on Kripke's view names are rigidified terms – a category that encompasses both the *directly-referring* and the *reference-fixing* sub-categories.

2.2.1.2.4   Analysis of the actualised descriptions view

According to the actualised descriptions view, names abbreviate not descriptions but actualised descriptions, and hence are also rigid designators. In terms of possible worlds semantics, this should be analysed as follows. Proper names have descriptive content. Each proper name designates in the actual world that which fits its content. It then designates, with respect to every counterfactual world, that same actual referent, i.e., whether or not it fits the descriptive content in that world.

If this analysis is sound, then the actualised description view of proper names belongs in our second category, namely, the *reference-fixing* category. In other words, proper names, according to the actualised description view, are reference-fixing terms (according to our formal definition).[31]

2.2.1.2.5   Analysis of widescopism

As discussed above, a standard analysis of the *de re/de dicto* ambiguity is in terms of scope-distinction. Based on this analysis, widescopism is the view that a name abbreviates a definite description whose existential quantifier has a wide scope relative to an intensional operator (rather than a narrow scope). The problem with this suggestion, as noted, is that names also occur in sentences that are not embedded in an intensional context, or, indeed, in any context at all. Thus this account, if successful, can only apply to the very limited number of cases, namely, cases of clauses embedded in intensional contexts.

However, it has been suggested that the *de re/de dicto* ambiguity applies directly to the definite description, and in fact, it is this ambiguity of the definite description which account for the ambiguity of the whole embedded clause. Such ambiguity of the definite descriptions is accounted for in terms of possible worlds semantics. Kripke (1977) puts it very clearly (although in the context of discussing definite descriptions and not proper names):

> If definite descriptions, $\iota x \varphi(x)$, are taken as primitive and assigned reference, then the conventional non-rigid [i.e., not *de jure* rigid[32]] assignment assigns to such a description, with respect to each possible world, the unique object, if any, which would have φ'd in that world. … Another type of definite description, $\iota x \varphi x$, a '[*de jure*] rigid' definite description, could be introduced semantically by the following stipulation: let $\iota x \varphi x$ denote, with respect to all possible worlds, the unique object that (actually) φ's. Both kinds of definite descriptions can obviously be introduced, theoretically, into a single formal language, perhaps by the notations just given. Some have suggested that

definite descriptions, in English, are *ambiguous* between the two readings. It has been
further suggested that the two types of definite descriptions, the nonrigid [i.e., non *de
jure* rigid] and [*de jure*] rigid, are the source of the *de dicto – de re* distinction…[33]

Kripke himself, by the way, says he has "an open mind on the subject"[34] and
hence neither endorses nor opposes such ambiguity.

So the idea is that the *de re/de dicto* ambiguity applies directly to the definite
description, and hence the ambiguity is not confined to embedded clauses. In fact, it is
not confined to any type of clause, as it equally applies to independent definite
descriptions, i.e., outside any clause or sentence. In other words, it is suggested that a
definite description is ambiguous between a *de re* and a *de dicto* readings of *it*.[35]

The implication of this to the analysis of proper names is that a proper name can
be simply said to abbreviate a *de re* definite description.

Recall that widescopism suggested a way of rigidifying the description that a
proper name allegedly abbreviates by invoking the *de re/de dicto* distinction. The above
possible-worlds analysis similarly provides a way of rigidifying the relevant definite
description by invoking the *de re/de dicto* distinction. Thus, the two suggestions share the
same spirit. In this sense, the above possible-worlds analysis can be seen as an extended
version of widescopism, with the advantage of having fewer limitations.

This suggestion fits very well within our fourfold categorisation of semantic
functions. On this suggestion, a proper name is said to abbreviate a *de re* definite
description. In terms of our categorisation, a *de re* definite description is just a reference-
fixing term, i.e., it designates in the actual world that which fits its descriptive content
(i.e., the definite description that it is said to abbreviate), and then designates, with
respect to every counterfactual world, that same referent. It is therefore a rigidified term.
Thus, widescopism, or better, the *de re* view of proper names, belongs in our second
category – the *reference-fixing* category.

Now if this is correct, then we have discovered that widescopism is, in essence,
just the actualised descriptions view. Both views simply take proper names to be
reference-fixing terms, i.e., terms that designate in the actual world that which fits their
descriptive content, and then designate, with respect to all counterfactual worlds, that
same referent.[36]

Interestingly, this, so-called by many *descriptivist*, view, now becomes very similar
to Kripke's view. For Kripke is happy to admit that at least some proper names fix their

reference in the actual world by fitting a description (see quote in Section 2.2.1.1.3). Thus, both views agree that names designate in counterfactual worlds that which they designate in the actual world; both views also agree that *some* names designate in the actual world by fitting a description. The disagreement narrows down to whether *some* names designate in the actual world directly (as is suggested by Kripke), i.e., regardless of descriptive content, or not (as is implied by the *de re*/widescopism view of proper names).

Finally, I would like to suggest a further support for the claim that the *de re/de dicto* ambiguity exists outside intensional contexts. This, I believe, can be achieved by acknowledging a very simple fact that is implicit in every sentence, namely, that it is *uttered* (or at least entertained). Once this is acknowledged, each un-embedded sentence of the form 'The F is G' can be transformed into an embedded one, in the form of 'it is uttered (/entertained) that, the F is G', which is thereby subject to a scope-distinction. Specifically, the *de dicto* reading of 'the F is G' is to be paraphrased as, 'it is uttered that, there is an x such that x is F and x is G'; whereas the *de-re* reading of 'the F is G' is to be paraphrased as: 'there is an x such that x is F, and it is uttered that *it* is G'. Thus, by adding the "utterance" operator, the existential quantifier can have a wide or narrow scope relative to this operator. Specifically, in the *de re* case the existential quantifier has a wide scope, whereas in the *de dicto* case the existential quantifier has a narrow scope. Just like in all intensional contexts.

2.2.1.3   Proper names: summary

In sum, there were five main views about the semantics of proper names: the Millian view, the Frege-Russell view, the New Theory view, actualised descriptions and widescopism, or the *de re* view. We have introduced a fourfold semantic categorisation, which is based on the distinction between designation in the actual world vs. designation in counterfactual worlds. Based on this categorisation, we came up with three basic semantic functions: *descriptive*, *reference-fixing* and *directly-referring*. Using our categorisation to analyse the above five views resulted in the following conclusion: the Millian view takes proper names to be directly referring terms. This view thus belongs in our fourth category. The Frege-Russell view takes proper names to be descriptive; it thus belongs in our first category. The New Theory takes proper names to be directly-referring *or* reference-fixing. In short, proper names belong to the higher *rigidified* category, which

includes both these sub-categories. The actualised descriptions view takes names to be reference-fixing; and so is widescopism, or the *de re* view of proper names. These views thus belong in our second category. The following Table 2.4 summarises this:

Designation in the actual world	Designation in counterfactual worlds	
By fitting a descriptive content	By fitting a descriptive content *Not rigidified*	**Descriptive** names according to Russell-Frege
By fitting a descriptive content	Regardless of descriptive content *Rigidified*	**Reference-fixing** names according to actualised descriptions; names according to widescopism; *some* names according to the New Theory
Regardless of descriptive content	By fitting a descriptive content *Not rigidified*	---
Regardless of descriptive content	Regardless of descriptive content *rigidified*	**Directly-referring** names according to Millianism *some* names according to the New Theory

Table 2.4

2.2.2   Definite descriptions

As we have already discussed definite descriptions at some length in the course of our examination of the semantics of proper names, our examination of the semantics of definite descriptions will, to a large extent, rely on the above discussion, and as a result will be significantly shorter. Consequently, we shall join the presentation of each view with its analysis in terms of our proposed fourfold categorisation.

2.2.2.1   The main views about the semantics of definite descriptions

2.2.2.1.1   *The Frege-Russell view of definite descriptions*

Both Frege and Russell intended their theory to apply to *all* singular terms, i.e., terms that designate an individual. In particular, both Frege and Russell believed that all singular terms designate their referent by fitting a descriptive content. Proper names are one

instance of singular terms; definite descriptions are another instance. According to the Frege-Russell view then, definite descriptions have a descriptive content, and they designate that which fits their descriptive content.

As previously mentioned, Frege and Russell did not discuss designation in counterfactual worlds. But it is customary to take the Frege-Russell view to imply that singular terms – definite descriptions included – designate in the same way in counterfactual worlds as they are said to designate in the actual world; namely, a definite description designates, with respect to each *counterfactual* world, that which fits the description in that world. If this analysis is sound, then, in terms of our fourfold categorisation, definite descriptions on the Frege-Russell view (just like proper names) belong in our first category, i.e., the *descriptive* category.

2.2.2.1.2   *The de re/de dicto ambiguity view of definite descriptions*

As we have seen in Section 2.2.1.1.5, it is commonly believed that sentences involving definite descriptions, when embedded in intensional context, are subject to a *de re/de dicto* ambiguity. This fact alone does not entail an ambiguity in the definite description itself. As Russell himself ingeniously demonstrated, the ambiguity can be elegantly expressed in terms of scope analysis, while preserving a descriptive account of definite descriptions. However it was demonstrated above (Section 2.2.1.2.5) that when analysing the same ambiguity in terms of possible worlds semantics, the ambiguity is shown to be independent of intensional contexts, and in fact, even independent of any sentence. Viz., the *de re/de dicto* ambiguity seems to apply directly to definite descriptions, regardless of any sentence. In detail, a definite description is ambiguous between a *de dicto* definite description, which designates, with respect to every counterfactual world, that which fits the description *in that world*, and between a *de re* definite description, which designates, with respect to every counterfactual world, that which fits the description *in the actual world*. While the Frege-Russell view only expresses the *de dicto* reading of definite descriptions, there is also the *de re* reading of it.

Thus, in terms of our fourfold categorisation, according to the *de re/de dicto* ambiguity view, definite descriptions belong in both the first *and* the second categories, i.e., they may be *descriptive* or *reference-fixing*.[37]

2.2.2.1.3   Referential definite descriptions

Some philosophers thought that definite descriptions have another use, which is neither descriptive (i.e., *de dicto*) nor reference-fixing (i.e., *de re*) in the above sense.

Consider the following example from Linsky. Observing a certain couple, a bystander says, "her husband is kind to her". It so happens that the two are not married. In this case, says Linsky, "the speaker might very well be referring to someone using these words, for he may think that that someone is the husband of the lady (who in fact is a spinster)."[38] Donnellan provides a well-known similar example:

> Suppose one is at a party and, seeing an interesting-looking person holding a martini glass, one asks, "Who is the man drinking a martini?" If it should turn out that there is only water in the glass, one has nevertheless asked a question about a particular person, a question that it is possible for someone to answer.[39]

Donnellan calls definite descriptions that function in this way *referential* (which he distinguishes from *attributive* definite descriptions that are used to state "something about whoever or whatever is the so-and-so"[40]). Thus, in both Linsky and Donnellan's cases, it is claimed that the definite description succeeds to refer to the individual intended by the speaker, despite the fact that that individual fails to fit the description. In other words, an alleged referential definite description picks out its referent directly, regardless of any descriptive content.[41]

The standard objection to referential definite descriptions is that in cases like the above "the man drinking a martini", the success of the reference is merely *pragmatic* and not *semantic*. i.e., although the speaker has managed to *communicate* a referent that fails to fit the description, the definite description failed to *express* such reference.[42] This, however, has been contested.[43]

This alleged referential use is clearly different from both the *de dicto* use and the *de re* use discussed above, as in both those uses the actual referent is picked out by means of fitting the description. So we have here a third use of definite descriptions.

This third use can be clearly distinguished from the above two uses in terms of our fourfold categorisation. A referential definite description may[44] designate in the actual world that which is intended, regardless of any descriptive content. Since it is meant to refer to a certain individual, it thus designates that same individual in all counterfactual worlds as well, i.e., again, regardless of descriptive content. Thus referential definite descriptions fit in our fourth category – that of *directly-referring* terms. By contrast (as was

pointed above), *De dicto* definite descriptions and *de re* definite descriptions, by contrast, fall within our first and second categories respectively.

2.2.2.2   Definite descriptions: summary

The ambiguity view of definite descriptions suggests that definite descriptions are ambiguous between a *de re* reading and a *de dicto* reading. Donnellan's view suggests that definite descriptions have a referential reading. The Frege-Russell view suggests that definite descriptions are not ambiguous, and that they only have their *de dicto* reading.

Applying our fourfold categorisation to these claims yields the following results: on the ambiguity view, a definite description is ambiguous between being descriptive and reference-fixing, i.e., it may belong in either our first or second category; on Donnellan's view, a definite description may also be directly-referring, i.e., it may belong in our fourth category. On the Frege Russell view, a definite description may *only* be descriptive, i.e., it may only fall in our first category:

Designation in the actual world	Designation in counterfactual worlds	
By fitting a descriptive content	By fitting a descriptive content *Not rigidified*	**Descriptive** *All* definite descriptions according to the Frege-Russell view; Only *de dicto* definite descriptions according to the ambiguity view
By fitting a descriptive content	Regardless of descriptive content *Rigidified*	**Reference-fixing** Only *De re* definite descriptions according to the ambiguity view
Regardless of descriptive content	By fitting a descriptive content *Not rigidified*	---
Regardless of descriptive content	Regardless of descriptive content *rigidified*	**Directly-referring** *Referential* definite descriptions according to Donnellan

Table 2.5

2.2.3 Complex demonstratives

The next type of natural-language term that we will be looking at is complex demonstratives - terms that have the general form, "that/this F" (as opposed to definite descriptions that have the form "the F", and to simple demonstratives such as "she").

Frege has a short discussion of demonstratives.[45] As with other singular terms, Frege takes demonstratives to have both sense and reference, where the sense determines the reference. Thus, a demonstrative has a descriptive content, and designates that which fits that content, e.g., "that guy" (pointing at Aristotle) abbreviates something like "the last great philosopher of antiquity", and designates Aristotle by virtue of fitting the description. As expected, objections to this Fregean view are similar to the ones directed at Frege-Russell views of proper names: the problem is that the sentence "that guy is a philosopher" is contingent, whereas "the last great philosopher of antiquity is a philosopher" seems to be necessary.[46] In other words, the problem is that on such a view, demonstratives become nonrigid. Whereas people may be in different minds about the rigidity of definite descriptions, demonstratives are widely agreed to be rigidified terms. E.g., whereas the simple definite description "the last great philosopher of antiquity" may be taken to designate, with respect to each counterfactual world, the individual who is the teacher of Alexander in *that world*, the complex demonstrative "*that* teacher of Alexander the Great" (pointing at Aristotle,) is widely agreed to designate the *actual* teacher, Aristotle, in every counterfactual world.

In terms of our fourfold categorisation, it turns out that a complex demonstrative designates, with respect to *counterfactual* worlds, that which it designates in the actual world, i.e., regardless of descriptive content. But what about the actual world? How does a complex demonstrative picks out its referent there – is it via descriptive content or independently of descriptive content?

Syntactically, what distinguishes a complex demonstrative from a simple definite description is the word "that". Thus, it must be the operator "that" that rigidifies the description. But how exactly? One natural view is that the operator "that" functions somewhat similarly to the operator "actual" in actualised descriptions. This seems to be Kaplan's view in his "Dthat".[47] Thus understood, the semantic function of "*that* teacher of Alexander", as opposed that of "*the* teacher of Alexander", is that of a reference-fixing term; it designates, in actual world, that which fits the description, and then designates

with respect to all counterfactual worlds, just that same actual referent (whether it fits the description or not in the counterfactual world).

This view, however, quickly runs into trouble. For consider a case in which that which actually fits the description is not the one gestured at, e.g., I use "*that* man drinking a martini" while pointing at Jones, who happens to have water in his glass. In that case, the above view entails that the actual referent of "*that* man drinking a martini" is not Jones, but Smith in the other room, who indeed has a martini in his glass, and this is a counterintuitive result (indeed, even more counterintuitive than in the case of a definite description "the man drinking a martini").

Others took the operator "that" to effectively annul the description.[48] David Braun calls theories of this sort *minimal theories*, which he characterises as follows:

> The common noun phrase in a complex demonstrative plays no semantic role in determining the referent of the complex demonstrative; so a person could be the referent of an utterance of "that crook" even if she is not a crook. Furthermore, this type of view says that the content of the common noun phrase is not a constituent of the content of the complex demonstrative; the content of the utterance of a complex demonstrative is just its referent.[49]

It follows that according to this view, complex demonstratives are in essence like referential descriptions. I.e., a complex demonstrative "that F" designates in the actual world that which is demonstrated (gestured at, or intended[50]), and then designates, with respect to all counterfactual worlds, that same referent. In our terminology, such a view takes complex demonstratives to belong to the *directly-referring* category.

But now Kripke's objection to referential definite descriptions resurfaces: in cases like that of "that crook", the success of the reference can be said to be merely *pragmatic* and not *semantic*. I.e., although the speaker has managed to *communicate* a referent that fails to fit the description, the complex demonstrative failed to *express* such reference. Kripke here expresses our intuitive discomfort in the face of reference to a man who has water in his glass by using, "that man drinking a *martini*".

The objection to the first view suggests that the *description* is not enough to determine the reference of a complex demonstrative in the actual world. The objection to the second view shows that the *gesture* is not enough to determine the reference of the complex demonstrative in the actual world. Consequently, an additional view is a conjunction of these two views. It suggests that in order for an object to be designated by a complex demonstrative in the actual world, it needs to satisfy *both* conditions,

namely, it has to be the one gestured at, *and* it has to fit the description. As Emma Borg puts it:

> An object cannot be the referent of "that cat" unless it is both the speaker's demonstrated referent and it is a cat. If no object meets both criteria, then the demonstrative utterance fails of reference: it is literally empty, lacking a truth condition.[51]

In other words, generally speaking, a complex demonstrative "that F" comprises two elements: the "that"-element and the "F"-element. Each generates a condition for designation in the actual world. The condition set by the "that"-elements is that of being gestured at; the condition set by the "F"-element is that of fitting the description "F". In order for an object to be designated by the complex demonstrative "that F" in the actual world, it has to satisfy *both* conditions.

We can use the example of "that man drinking a martini" to contrast the three views: someone is uttering "that man drinking a martini" while pointing at Jones, but the person gestured at, Jones, does not fit the description, and the person that fits the description, Smith, is not gestured at. In this case, the combined view holds that the complex demonstrative does not designate Smith, for it is Jones who is gestured at, (contrary to the reference-fixing view), nor does it designate Jones, for it is Smith who has a martini in his glass, and not Jones (contrary to the above "minimal view"). Rather, since in this case no candidate both fits the description and is gestured at, the complex demonstrative simply fails to refer. [52,53]

2.2.3.1   Complex demonstratives: summary

We can summarise the different views by referring to our fourfold categorisation. On a Frege-Russell view, a demonstrative has a descriptive content and designates that which fits that content. Thus on this view, a complex demonstrative designates, with respect to each world (actual and counterfactual) that which fits its descriptive content in that world. It follows that complex demonstratives, on this view, belong in our first category – descriptive terms. Contrary to what such a Frege-Russell view entails, complex demonstratives are widely agreed to be rigidified terms, i.e., they designate, with respect to each counterfactual world, that which they designate in the actual world. Yet views are divided about the designation of complex demonstratives in the actual world. Kaplan's (1978) view takes them to be reference-fixing, i.e., to designate in the actual world that which fits the descriptive content of the demonstrative. By contrast, the "minimal"

theory takes complex demonstratives to be directly-referring, i.e., to designate in the actual world that which is gestured at, i.e., regardless of the description. These two views seem problematic. The problem that the first view appears to have is that the *description* is not enough to determine the reference of a demonstrative in the actual world. The problem that the second view appears to have is that the *gesture* is not enough to determine the reference of the demonstrative in the actual world. Thus, an alternative view is that a complex demonstrative "that F" designates in the actual world that which both fits the description "F", and which is *also* gestured at by "that". This additional view does not fit into our original categorisation, and in fact generates a new, conjunct, category:

Designation in the actual world	Designation in counterfactual worlds	
By fitting a descriptive content	By fitting a descriptive content *Not rigidified*	**Descriptive** Complex demonstratives according to a Frege-Russell style view
By fitting a descriptive content	Regardless of descriptive content *Rigidified*	**Reference-fixing** Complex demonstratives according to Kaplan (1978)
Regardless of descriptive content	By fitting a descriptive content *Not rigidified*	---
Regardless of descriptive content	Regardless of descriptive content *rigidified*	**Directly-referring** Complex demonstratives according to the "minimal" view

Directly *and* by fitting a descriptive content	Regardless of descriptive content *rigidified*	**Combined** Complex demonstratives according to Borg

Table 2.6

2.2.4   Natural-kind terms

The last type of natural-language terms that we will examine is *natural-kind* terms. So far we have discussed natural-language terms which are singular terms. Natural-kind terms are general terms. However, due to various important similarities between proper names

and natural-kind terms, views on proper names are commonly extended to natural-kind terms as well.

Natural-kind terms are terms like "tiger", "gold", (arguably) "water", "lemon", or "heat". They are contrasted with terms of non-natural-kinds, typically human constructions, such as "city", "ten-pound note", and "bachelor".

A Millian-style view of natural-kind terms is that they designate their referents directly, regardless of any descriptive content. "Gold" for instance simply designates the stuff gold, regardless of any description that may be true of gold.

This view is subject to the same types of objections that Millianism faced with regard to proper names:[54] (a) Frege's puzzle: if Millianism is true, then the terms "Furze" and "Gorse" have the same meaning, and hence are substitutive *salva veritate*. If follows that "Furze is Gorse" amounts just to "Furze is Furze". But the latter, unlike the former, is trivial and uninformative; (b) Empty natural-kind terms: the sentence "gnomes are mythical creatures" seems meaningful. Yet if Millianism is true, then "gnomes" refers to nothing, i.e., it means nothing, and hence the sentence is just meaningless; (c) Negative existentials: "unicorns do not exist" seems meaningful and true. However, according to Millianism, the sentence comes out absurd: it denies the existence of something the existence of which it presupposes; and (d) Propositional attitude reports: the sentence "Fred believes that filberts, but not hazelnuts, are sweet" seems to express a mistaken, yet nonetheless rational, belief. Millianism will simply render it a self-refuting, and hence irrational, belief.

A version of the Frege-Russell view that applies to natural-kind terms will avoid these objections. According to such a view a natural-kind term has some descriptive content by virtue of which it designates its referent. Replacing each natural-kind term with the alleged abbreviated description will bar the above objections.

However, such a descriptive view faces the modal charges raised by the New Theory of reference: suppose that the descriptive content of "tiger" is, "a large carnivorous quadrupedal feline, tawny yellow in colour with blackish transverse stripes and white belly."[55] Yet the fact that, for example, tigers have stripes seems contingent. Tigers might not have been striped. But if "tiger" abbreviates the above description, then the claim that tigers are striped comes out necessary, which is false. Hence descriptivism is false.

Thus the problem with Frege-Russell descriptivism about natural-kind terms is that it renders such terms nonrigid (or better, non-rigidified), thus lending this view to Kripke's modal arguments. As in the case of proper names, a remedy to descriptivism seems to require a rigidification of the alleged descriptive content of natural-kind terms. Again, such rigidification can be achieved by actualising the description, E.g., "gold" stands for "the actual golden stuff", or simply by taking natural-kind terms to be *de re* terms, i.e., abbreviations of *de re* descriptions. [56]

The New Theory of reference extended to natural-kind terms simply claims that natural-kind terms are rigid (or better, rigidified) designators. "Tiger" designates just tigers in every counterfactual world, regardless of any description that may apply to them in that world. There is, however, an important difference between the New Theory of reference as applied to proper names and as it is applied to natural-kind terms. The difference is that whereas it is clear what a proper name rigidly designates, it is less clear what a natural-kind term rigidly designates. "Aristotle" rigidly designates Aristotle, yet what does "tiger" rigidly designate? "tiger" cannot rigidly designate the set of individual tigers that it actually designates, since in other counterfactual worlds "tiger" designates a different set of individuals. [57] As an alternative, it has been suggested that "tiger" designates an abstract entity, the kind, or property, of being a tiger. And it designates that kind, or property, in every counterfactual world. However, this suggestion seems to allow too much. Artificial kinds like "city", "bachelor", and even general terms like "watery", "hot", and "red" also each designate a kind, or property, in every possible world, and are thus rendered rigid by that view. In fact, some claim that according to this view, any kind-term, or property-term, turns out rigid. In other words, the charge is that this view trivialises rigidity. [58] As a response, Devitt (2005) suggests the following alternative sense for the rigidity of natural-kind terms:

> A general term 'F' is a rigid applier iff it is such that if it applies to an object in any possible world, then it applies to that object in every possible world in which the object exists. Similarly for a mass term
>
> …
>
> Clearly, if 'F' is a rigid applier then any individual *F* must be essentially *F*. So the view that there are any such 'F's entails a fairly robust metaphysical thesis. Still, that thesis has been popular from ancient times to the present and I think that it is plausible.[59]

It seems to me that this suggestion can be paraphrased as follows: a natural-kind term designates, with respect to every counterfactual world, objects of the same (natural) *kind* that it designates in the actual world. (Although I believe Devitt's definition is entailed by the paraphrase, I am not entirely sure that they are equivalent). Thus the

referent of a natural-kind term is a set, but it is not the same set in every counterfactual world. Now surely, being of the same natural-kind involves criteria of identity, but the same is true of proper names as well: a rigid name designates, with respect to every counterfactual world, *the same* individual that it designates in the actual world. So involving criteria of identity is no special feature of natural-kind terms, but rather a mark of rigidified designators in general. And it is this link to criteria of identity that restores the distinction between natural-kind terms, and other predicates. If this last proposal of Devitt's is successful, (as I believe it is,) then the New Theory of reference can be extended to natural-kind terms.

As mentioned in our discussion of proper names, the New Theory is primarily concerned with designation in counterfactual worlds, and not so much with how the referents are picked out in the actual world. Designation in the actual world can be either by ostension or by fitting a description. Thus, to take a mass natural-kind as an example, "gold" may initially be "baptised" either by pointing at a sample (or some samples) of gold (a directly-referring term), or it can be fixed to name "that which is the (actual) golden stuff" (a reference-fixing term). Both terms – the directly-referring "gold" and the reference-fixing "gold" – designate the same stuff in all counterfactual worlds. The difference between the two however, is the following: whereas a directly-referring "gold" designates the same stuff in all *actual* worlds as well, a reference-fixing "gold" will designate different stuffs in different *actual* worlds.

2.2.4.1   Natural-kind terms: summary

The analysis of views about natural-kind terms in terms of our fourfold categorisation is as follows. On a Millian-style view, natural-kind terms designate in the actual world directly some individuals, i.e., regardless of any descriptive content, and then designate, with respect to every counterfactual world, objects of the same (natural) kind. Thus, according to this view, natural-kind terms belong in our fourth category, *directly-referring* terms. According to a Frege-Russell style descriptivism, a natural-kind term designates in all worlds (actual and counterfactual) that which fits its descriptive content. Thus according to that view, natural-kind terms belong in our first category – *descriptive* terms. According to rigidified descriptions views, a natural-kind term designates in the actual world those referents which fit the term's descriptive content, and it then designates, with respect to all counterfactual worlds, referents of *the same* kind as the actual ones. So

on this view, natural-kind terms belong in our second category – *reference-fixing* terms. Lastly, according to the New Theory, a natural-kind term designates, with respect to every counterfactual world, referents of the same (natural) kind as the referents that it designates in the actual world. In the actual world, natural-kind terms may either designate directly, or via fitting the term's descriptive content. Thus, according to this view, natural-kind terms belong both in the second category – *reference-fixing* terms, and in the fourth category – *directly-referring* terms. In other words, natural-kind terms are simply *rigidified* terms. Table 2.7 summarises this:

Designation in the actual world	Designation in counterfactual worlds	
By fitting a descriptive content	By fitting a descriptive content	**Descriptive** All natural-kind terms according to a Russell-Frege descriptivism
By fitting a descriptive content	Regardless of descriptive content	**Reference-fixing** All natural-kind terms according to rigidified descriptivism *Some* natural-kind terms according to the New Theory
Regardless of descriptive content	By fitting a descriptive content	---
Regardless of descriptive content	Regardless of descriptive content	**Directly-referring** All natural-kind terms according to Millianism *Some* natural-kind terms according to the New Theory

Table 2.7

2.2.5 Summary of the proposed analysis of natural-language terms

We have been discussing the semantics of natural-language terms; specifically, we have discussed proper names, definite descriptions, complex demonstratives and natural-kind terms. We have presented different views about each of these terms. We aimed at comparing and contrasting those view. For this purpose, we have suggested a fourfold categorisation of semantic terms, based on two distinctions: (a) designation in the actual world vs. designation in counterfactual worlds; and (b) designation by fitting a descriptive content vs. designation regardless of such content. Based in this categorisation, we have defined three basic semantic functions: descriptive, reference-fixing, and directly-

referring. The comparison of the different semantic views about the various natural-language terms culminates in an interesting overall picture. The obvious thing is that different views ascribe to a given term different semantic functions. But more interestingly, these semantic functions were shown to always be one of our three proposed basic functions (Borg's view on complex demonstratives combines two of these functions). More interestingly still is that with respect to each type of term, for each of the three basic functions, it turned out that there is always a view that takes the term to have that basic function (or at least to be able to, in the case of ambiguity views). Table 2.8 specifies this:

	Proper names	Natural-kind terms	Definite descriptions	Complex demonstratives
Descriptive	Frege-Russell	Frege-Russell	Frege-Russell; ambiguity views *(de dicto* definite descriptions)	Frege-Russell
Reference-fixing	Rigidified descriptivism; New Theory *(some* names)	Rigidified descriptivism; New Theory *(some* natural-kind terms)	Ambiguity views *(De re* definite descriptions)	Kaplan (1978)
Directly-referring	Millianism; New Theory *(some* names)	Millianism; New Theory *(some* natural-kind terms)	Donnellan *(Referential* definite descriptions)	"Minimal" view
Combined view				Borg

Table 2.8

An important lesson to draw from this analysis is that there is no use in talking about *the* semantics of a certain kind of term, as this is a matter of much controversy. Rather, it is more constructive to talk about formal terms, namely, about descriptive terms, reference-fixing terms or directly-referring terms. If this is sound, then it follows that once we've discussed the formal language terms, we've thereby discussed the semantics of all these natural-language terms. (Whatever these semantics turn out to be).

2.3   Intension as a Means to Distinguish among Semantic Functions

We could hope to compare the different referring terms in our natural-language by means of differences in their intensions. But, as we have seen, there are ongoing controversies about the semantics of these natural-language terms, and want of a consensus about the semantics of a term, how can we determine its intension?

The above analysis of these views in terms of our fourfold categorisation can solve this problem.

Some views regard proper names as descriptive, others as reference-fixing, and yet others as directly-referring. Something similar holds for the other natural-language terms: definite descriptions, demonstratives and natural-kind terms. Thus, since all views fit into our categorisation, we can simply compare our formal categories – descriptive, reference-fixing, and directly-referring. The semantics of these formal types of terms are well defined, and thus can be easily compared and contrasted by means of their different intensions. In other words, rather than comparing the natural-language terms directly, we shall compare the underlying semantic structures that they *may* have.

Note: as this chapter's main concern is semantics, we shall assume, for the sake of simplicity, that there is no problem with criteria of identity, and hence will ignore this factor in the present discussion.

Suppose Jones is Taylor's neighbour. There are various types of terms that we can use to refer to him. We can refer to Jones by using a descriptive term (e.g., a *de dicto* definite description "Taylor's neighbour"), or by using a reference-fixing term (e.g., a *de re* definite description "Taylor's neighbour", or an actualised description "Taylor's *actual* neighbour"), or by using a directly-referring term (e.g., a referential definite description "Taylor's neighbour", or a Millian-name "Jones").

The descriptive term - e.g., a *de dicto* "Taylor's neighbour" – designates, with respect to every possible world, that which fits the descriptive content, i.e., the person who is Taylor's neighbour, in that world. Thus, the designation of a descriptive term in counterfactual worlds is entirely independent of its designation in the actual world. In other words, the designation of a descriptive term is constant across actual worlds, but varies across counterfactual worlds. E.g., in the case of a descriptive term whose descriptive content is "Taylor's neighbour", the intension of that term is,

Designation in the actual world	Designation in counterfactual worlds	State of the actual world	**W_1** **Taylor's neighbour is Jones**	**W_2** **Taylor's neighbour is Smith**	**W_3** **Taylor's neighbour is Brown**
Via fitting a descriptive content	Via fitting a descriptive content	**W_1** **Taylor's neighbour is Jones**	Jones	Smith	Brown
		W_2 **Taylor's neighbour is Smith**	Jones	Smith	Brown
		W_3 **Taylor's neighbour is Brown**	Jones	Smith	Brown

Table 2.9

By contrast, the intension of a reference-fixing term for Jones – e.g., the actualised definite description "Taylor's *actual* neighbour" – designates, with respect to every counterfactual world, that which fits the description – e.g., the person who is Taylor's neighbour – in the *actual* world. Viz., a reference-fixing term designates in the actual world that which fits the descriptive content, and then designates, with respect to all counterfactual worlds, that same individual. Hence, the designation of the reference-fixing "Taylor's actual neighbour" in counterfactual worlds is entirely dependent on its designation in the actual world. The intension of such a reference-fixing term, e.g., "Taylor's actual neighbour", is thus constant across counterfactual worlds, yet it varies across actual worlds:[60]

Designation in the actual world	Designation in counterfactual worlds	State of the actual world	**W_1** **Taylor's neighbour is Jones**	**W_2** **Taylor's neighbour is Smith**	**W_3** **Taylor's neighbour is Brown**
Via fitting a descriptive content	Regardless of a descriptive content	**W_1** **Taylor's neighbour is Jones**	Jones	Jones	Jones
		W_2 **Taylor's neighbour is Smith**	Smith	Smith	Smith
		W_3 **Taylor's neighbour is Brown**	Brown	Brown	Brown

Table 2.10

Finally, a directly-referring term for Jones – e.g., a Millian-name "Jones" - designates, with respect to every possible world, simply Jones, i.e., regardless of fitting any descriptive content. The intension of "Jones" is thus constant across the board:[61]

Designation in the actual world	Designation in counterfactual worlds	State of the actual world	W_1 Taylor's neighbour is Jones	W_2 Taylor's neighbour is Smith	W_3 Taylor's neighbour is Brown
Regardless of a descriptive content	Regardless of a descriptive content	W_1 Taylor's neighbour is Jones	Jones	Jones	Jones
		W_2 Taylor's neighbour is Smith	Jones	Jones	Jones
		W_3 Taylor's neighbour is Brown	Jones	Jones	Jones

Table 2.11

In conclusion, there is a clear structural pattern that emerges here: a descriptive term is one that is constant across actual worlds, yet varies across counterfactual worlds; a reference-fixing term is constant across counterfactual worlds, yet varies across actual worlds, and a directly-referring term is constant across the board, i.e., both across actual worlds and across counterfactual worlds.

2.4 Fine-Graining "The Actual Referent" Factor in the Formula

How does our fourfold classification of semantic terms relate to our "calculator of intensions"?

Recall that our general formula of intensions accepts three arguments that together determine the intension of a term: (a) the semantic function; (b) the actual referent; and (c) criteria of identity. Thus, one evident way in which the fourfold classification contributes to such a formula is that it fine-grains the semantic function factor. The initial distinction, between descriptive and rigidified terms was shown to be focused on designation in *counterfactual* worlds: where the former designates via fitting a descriptive content, and the latter designates regardless of such content. Yet we found that the same distinction applies to the actual world as well: with respect to the actual

world, terms may designate via fitting a descriptive content or regardless of such content. In particular, this latter distinction divided the category of rigidified terms into two sub-categories: reference-fixing terms and directly-referring terms.

However, this fine-graining may at first seem to have little impact on *intensions*. The intension of a term is the reference of the term with respect to counterfactual worlds. Thus, given that, as far as the semantic function is concerned, the intension of a term can be determined on the basis of the way a term designates in counterfactual worlds, why bother finding the way the term designates in the actual world as well?

Interestingly, it will be argued that the way a term designates in the actual world does play a crucial role in determining intensions, but it does so through another factor: *the actual referent*. Specifically, the actual referent is determined by two other factors: (i) the way the term designates in the actual world; together with (ii) the state of the actual world. Therefore, since the intension of a term is dependent upon its actual referent, it is thereby dependent upon the two factors that determine this actual referent, one of which is the way the term designates in the actual world. Let me demonstrate this.

As in the previous section, we shall initially assume, for the sake of simplicity, that there is no problem about re-identification, and hence no need for criteria of identity. E.g., we shall take the intension of a rigidified "Newton" to simply assign Newton to each counterfactual world (i.e., assuming no problem in re-identifying Newton in other possible worlds). We shall thus initially develop our discussion while ignoring the factor of criteria of identity, and will incorporate it only at a later stage.

When the semantic function of a term is descriptive, the effect of the actual referent on intension is cancelled out, i.e., the intension of a descriptive term is in fact independent of the actual referent. However, the intension of any rigidified terms *is* determined by its actual referent. In other words, changing the actual referent of a rigidified term will result in a change of its intension. But what could make an actual referent be something other than it is?

I shall defend the following: the actual referent of a term is determined by both the way in which the term picks out its referent in the actual world, and by what the actual world is like. Changing each of these may result in a change of the actual referent. This breaks down into the following two claims. On the one hand, (holding the actual world constant) different ways the term designates in the actual world will yield different actual referents. On the other hand, (holding the way of designation in the actual world

constant) different states of the actual world will result in different actual referents as well. Let me illustrate these two claims in order.

The first claim is that (while holding the actual world constant) different ways the term designates in the actual world will yield different actual referents. There are two ways in which a rigidified term can pick out its referent in the actual world: by directly "labelling" (a directly-referring term), or by fitting a descriptive content associated with the term (a reference-fixing term). A directly-referring term "T" and a reference-fixing "T" will pick out the same referent in an actual world, as long as the thing "labelled" is the same as the thing that fits the descriptive content. For example, considering our world as the actual world, both a directly-referring "Aristotle" (i.e., a term that directly "labels" Aristotle), and a reference-fixing "Aristotle" (say, an abbreviation of the rigidified description "the *actual* last great philosopher of antiquity") will designate Aristotle in this actual world. Both being rigidified terms, they will also designate Aristotle in all counterfactual worlds. But now consider the possibility that the *actual* world is such that Aristotle was not so bright, and that it was his wife, Pythias, who secretly wrote all the books and prepared his lectures for him. Holding this actual world constant, each of the two semantic functions will yield a different actual referent. A directly-referring "Aristotle" will still yield Aristotle as the actual referent, and hence will designate Aristotle in all counterfactual worlds, whether he is a philosopher there or not; whereas a reference-fixing "Aristotle" will yield Pythias as the actual referent, and hence will designate Pythias in all counterfactual worlds, whether she is a philosopher there or not. Thus, changing the way that the term designates in the actual world may result in a change of the actual referent, and consequently, a change in the intension. (Recall that for simplicity we initially assume no problem with criteria of identity, of people or anything else, and hence ignore this factor in the formula for now.) The intension of a rigidified "Aristotle" is thus:

						W_1	W_2	...
						The last great philosopher of antiquity is Pythias	The last great philosopher of antiquity is Aristotle	
Designation in counterfactual words	Designation in the actual world	+	The state of the actual world	=	**Actual referent**			
Regardless of descriptive content ("rigidified")	Via fitting a descriptive content ("reference-fixing")	+	The last great philosopher of antiquity is Pythias	=	**Pythias**	Pythias	Pythias	Pythias
	Regardless of descriptive content ("directly-referring")	+		=	**Aristotle**	Aristotle	Aristotle	Aristotle

Table 2.12

The second claim was that (while holding the designation in the actual world constant) different states of the actual world will result in different actual referents as well (and, consequently, different intensions). However, as reference-fixing terms and directly-referring terms pick out their referent in the actual world in different ways, the contribution of the state of the actual world to determining reference is different in the two cases. We shall thus examine those separately.

Let us begin by looking at the reference-fixing term. We have seen that if the actual world is such that Pythias came up with all the ideas ascribed to Aristotle, a reference-fixing "Aristotle" whose descriptive content is "the last great philosopher of antiquity" picks out Pythias as the actual referent (and, consequently, as the referent in all counterfactual worlds). However, if the actual world is as we believe it to be, the same reference-fixing "Aristotle" yields Aristotle as the actual referent (and, consequently, also as the referent in all counterfactual worlds). Thus, changing the actual world may result in a change of the actual referent of a reference-fixing term. (As above, we assume that there is no problem with criteria of identity, and hence we ignore this third factor.) The intension of a reference-fixing "Aristotle" is thus:

	Designation in the actual world	+	The state of the actual world	=	Actual referent	W_1	W_2	…
Designation in counterfactual words	Designation in the actual world	+	The state of the actual world	=	**Actual referent**	The last great philosopher of antiquity is Pythias	The last great philosopher of antiquity is Aristotle	
Regardless of descriptive content ("rigidified")	Via fitting a descriptive content ("reference-fixing")	+	The last great philosopher of antiquity is Pythias	=	**Pythias**	Pythias	Pythias	Pythias
		+	The last great philosopher of antiquity is Aristotle	=	**Aristotle**	Aristotle	Aristotle	Aristotle

Table 2.13

A similar dependency holds for directly-referring terms, i.e., terms that designate their referent in the actual world directly, regardless of any descriptive content. Such a directly-referring "Aristotle" simply "labels" Aristotle. Now consider the possibility that Aristotle was not a human being, but rather a super-sophisticated robot from outer space (how else could one person contemplate and write this entire profound corpus in one lifetime?) For convenience, let *us* refer to this robot as "Rob". In that case, "Aristotle" picks out in the actual world a robot, Rob. Thus "Aristotle" will designate Rob in all counterfactual worlds. (Consequently, in a *counterfactual* world where the guy referred to as "Aristotle" is exactly like who we think of as Aristotle (i.e., human, philosopher, etc.) our directly-referring term "Aristotle" will *not* designate him.) However, the actual world may be as we believe it to be, i.e., Aristotle is a human being. For convenience, let *us* refer to this guy as "Hum". In case Aristotle is a human being, a directly-referring "Aristotle" will pick out that person, Hum, in the actual world, and, consequently, it will pick out that same human in all counterfactual worlds. It thus follows that changing the actual world results in a change of the actual referent of a directly-referring term as well. (Again, for simplicity, we assume that there is no problem with criteria of identity of people or, for this matter, of robots; hence we ignore that factor.) The intensions of a directly-referring "Aristotle" is thus:

Designation in counterfactual words	Designation in the actual world	+	The state of the actual world	=	Actual referent	*W₁* The last great philosopher of antiquity is the person, Hum, referred to as "Aristotle"	*W₂* The last great philosopher of antiquity is a robot, Rob, referred to as "Aristotle"	...
Regardless of descriptive content ("rigidified")	Regardless of descriptive content ("directly-referring")	+	Aristotle is human	=	**Hum**	Hum	Hum	Hum
		+	Aristotle is a robot	=	**Aristotle**	Rob	Rob	Rob

Table 2.14

To conclude, the intension of a rigidified term is dependent upon the actual referent of a term. So different possibilities of the actual referent of a term will generate different intensions. Having fine-grained semantic functions, we came to realise that the actual referent of a term is itself determined by two factors: (i) the way in which the term designates in the *actual* world (directly or via fitting a descriptive content), and (ii) the way the actual world is like. Changing the values of any of these variables results in a change of the actual referent of the term, and, consequently, in the counterfactual referents of the term; in other words, changing the values of these variables results in a change of the intension of the term.

In this sense, therefore, considering the way in which terms designate in the actual world (apart from the way they designate in counterfactual worlds), is in fact a fine-graining of the factor of the actual referent: it provides insight into the elements which determine this factor. We could refine our formula of intensions by replacing the variable of the actual referent, with the following two variables: the way of designation in the actual world, and the state of the actual world. The factor we called "the semantic function" should more accurately be called "designation in counterfactual worlds":

Designation in counterfactual words	Designation in the actual world	The state of the actual world	Criteria of identity	W_1	W_2	…
Via fitting a descriptive content (*descriptive*)	(Whatever)	(Whatever)	(Whatever)			
Regardless of descriptive content (*rigidified*)	Via fitting a descriptive content (*reference-fixing*)	W_1				
		W_2				
		…				
	Regardless of descriptive content (*directly-referring*)	W_1				
		W_2				
		…				

Table 2.15

Put it differently, we may conclude by saying that the intension of a term depends on the following *three* factors: (a) the semantic function of the term, i.e., the way the term picks out its referent in possible worlds (distinguishing picking out in actual vs. picking out in counterfactual worlds); (b) the state of the actual world; and (c) criteria of identity. Thus, overall, the intension of the term is dependent upon semantics, metaphysics and actuality. (This is in fact the formulation of the "calculator" as presented in the introduction to this dissertation).

Lastly, for simplicity, we so far ignored the metaphysical issue of re-identifying a referent in counterfactual worlds in this discussion. What if we did not ignore it?

If we want to take the further factor of criteria of identity on board, the intension of a rigidified term will be determined as follows. As we have just established, the actual referent of the terms is determined by the way the term designates in the actual world, together with the state of the actual world. E.g., if "Aristotle" is a reference-fixing name that is designed to designate in the actual world the last great philosopher of antiquity, and the state of the actual world is that the last great philosopher is Aristotle, then the actual referent of "Aristotle" is Aristotle. Once determined, the actual referent needs to be checked for its properties. Suppose Aristotle is human, and has certain mental features M and a body B. Once these have been fixed, then together with the criteria of personal

identity, we can re-identify the actual referent on counterfactual worlds. E.g., suppose the criteria of identity for people are psychological. In that case, Aristotle is identical to whoever has his mental features, i.e., he is identical to a guy in some counterfactual world who has mental features M yet another body C, but not to a guy on some counterfactual world who has body B but mental features N, or no mental features at all.

Endnotes

---

[1] With the exception of one "hybrid" view of complex demonstratives; but even this view takes complex demonstratives to be a combination of two of the above three.

[2] Kripke (1972/1980, p. 96)

[3] Kripke 1972/1980, p. 79, n. 33). Evans's (1977) descriptive name "Julius", that is stipulated to designate the (actual) inventor of the zip, is a similar case.

[4] Kaplan (1977/1989, p. 483)

[5] Kripke (1980, pp. 55, 135). There is a sense in which this title is not completely suitable; all rigidified terms have their reference fixed across possible worlds, and all have a way of fixing the reference in the actual world; e.g., directly-referring terms may be said to fix their actual referent by an act of ostension, or "labelling". Thus, *our* use of "reference-fixing" here is restricted to fixing the reference by means of fitting a descriptive content.

[6] Nathan Salmon (1981 p. 33 n.35) says that there is nothing to logically prevent directly-referring terms from being nonrigid, e.g., due to some device that makes them designate different referents in different possible worlds. However, in the case of *semantically simple* terms like proper names, such device is not likely to be a description, and hence such terms will not fit into our third category. And at any rate, Salmon admits that "to do so [i.e., to vary designation in counterfactual worlds] seems arbitrary and artificial", thus partly explaining why there are no such terms in English (and presumably in any other natural languages).

[7] Notably Dummett (1973; 1981). See Section 2.2.1.2.5.

[8] Mill (1867, reprinted in 1949, p. 20)

[9] Notable contemporary proponents of the Millian view of names include Kaplan (1989); Salmon (1986), and Soames (2002, p. 240, 243).

[10] Russell's (1905) original version involved definite descriptions rather than proper names, yet it is widely accepted that the objections equally apply to a Millian view of any singular term. It is a matter of controversy whether ordinary names in our natural language are indeed directly-referring or not. As regards names, Russell (1910) called such Millian names *logically proper names*, and thought that ordinary proper names, in natural language, are not such logically proper names.[10] The adaptations to proper names brought here are due to Reimer (2003).

[11] Grice (1975)

[12] See, for example, Salmon (1986) and Soames (1989).

13 More specifically, Russell's theory of descriptions (1905) argues that sentences containing definite descriptions of the form "the F is G" are equivalent to "there is an x such that, x is uniquely an F, and it is G". Thus, definite descriptions should be understood in terms of *quantifiers* (Russell took this to be an advantage over Frege's model. See Endnote 14). In that initial introduction of his theory, however, there was not yet mention of proper names.

14 Two main differences between the views, which Russell (1905) stressed, are, (1) Frege's view entails that the referent of a name is a constituent of the proposition that involved the name, whereas on Russell's view the referent is not a constituent of the proposition; (2) Frege's view is a two-layer view: it involves references and sense. Russell's view, by contrast, is a one-layer view. It involves no sense, i.e., it has less metaphysical commitment (which Russell of course considered an advantage).

15 Notably Kripke (1972/1980).

16 Devitt and Sterelny (1999, p. 49)

17 Another objection to descriptivism which is frequently raised (mainly stressed by Putnam (1975)) is based on the fact that different people associate different descriptions, or senses, with the same proper name. This leads to two related problems. Firstly, names become ambiguous, hence, the problem of *unwanted ambiguity* (Devitt and Sterelny, 1999, p. 49). Secondly, some people may associate an entirely wrong description with a name, for instance, some people may associate with the name "Einstein" the definite description "the scientist who invented the atomic bomb" (Kripke, 1980); for those people, the name "Einstein" refers not to Einstein but to Oppenheimer. This is known as the problem of *ignorance and error*. However, on some versions of descriptivism, the description that the name abbreviates is not "in the head" of the speaker, but rather an objective abbreviation, independent of the speaker. Such versions are therefore immune to these objections, and thus I believe these types of objections not to be principal.

18 Kripke (1972/1980, p. 96)

19 Searle (1983); Nelson (2002). For a detailed discussion of this proposal see Stanley (1997).

20 Kaplan (1989, p. 577). Kaplan, however, does not endorse this view. Rather, he sees proper names as directly-referring. (For exact references, see Endnote 61 in this chapter).

21 Ludlow (2005) mentions attempts to rigidify the descriptions that names are said to abbreviate by means of restricting the descriptions to a very particular kind: one such kind is that a name "Aristotle" abbreviates something like the description "the individual designated by the name 'Aristotle'" (Ludlow lists Loar (1976), McDowell (1977), and Schiffer (1978) as proponents). The weakness of the suggestion is evident in the title it received, *circular descriptivism* – "Aristotle" is

abbreviated by a description that involves the name "Aristotle". Another such kind is that "Aristotle" abbreviates something like "the individual linked by a causal chain to Aristotle", which Lewis (1984) called *causal descriptivism* (Ludlow lists Kroon (1987) and Jackson (1998b) as supporters). As these suggestions become much closer to Kripke's view than to the Frege-Russell descriptivism, I do not see them as interesting candidates for rescuing *descriptivism*.

[22] Dummett (1973, pp.111-135, 1981)

[23] Russell (1905/2001, p. 217)

[24] The distinction dates back to medieval philosophy. Literally, "*de dicto*" means "of, or concerning, a dictum", i.e., concerning a bearer of representative content (something like a sentence, a statement, or a proposition), whereas "*de re*" means "of, or concerning, a thing" (Galois, 1998). The example to follow should clarify this distinction.

[25] However, Quine (1956) has famously objected to the scope analysis of the *de re/de dicto* distinction, on grounds that the wide scope reading involves *quantifying into* an *opaque* (i.e., intensional) context (as is demonstrated below), which he considered illegitimate.

[26] This, however, does not mean that the two distinctions – *de re/de dicto* and narrow/wide scope – are interchangeable. As Kripke (1977) demonstrated, whereas the *de re/de dicto* distinction can be accounted for in terms of scope, the converse does not hold. The reason is that there are more than two scopes, and hence "no twofold distinction can do the job" (1977, p. 10).

[27] Kripke (1980, pp. 11-15)

[28] LaPorte (2006) mentions such suggestions made by Sosa (2001); Hunter (2005); Caplan (2005) (who actually argues against this view); and Everett (2005).

[29] Kaplan (1978/2001, p. 328)

[30] Kaplan (1978/2001, p. 329). Kaplan expresses this view elsewhere as well. E.g., (1977/1989; 1979). See also Endnote 61 below.

[31] I thus believe that Chalmers's two-dimensionalist view of names (e.g., 2002; 2006) also fits here: "Hesperus" stands for "the actual evening star", and thus designates Hesperus in every *counterfactual* world; yet it designates different stars (more correctly, planets) in different *actual* worlds. See also Endnote 60 below.

[32] As noted above, a term with the semantic function specified here by Kripke, may be *de facto* rigid, i.e., it nay *turn out* to designate the same referent in all counterfactual worlds, e.g., "the sum of 2+3". Hence, I believe that instead of rigid/nonrigid, Kripke should have used "rigidified" (or *de jure* rigid)/non-rigidified (or, not *de jure* rigid).

[33] Kripke (1977, pp. 10-11). For a more recent similar analysis, see Sidelle (1992, p.412).

[34] Kripke (1977, p. 11)

[35] With relation to the above quote, Kripke simply states "some" philosophers that hold this view without naming any. Kaplan (1978) proposes a *de re/de dicto* ambiguity of un-embedded *sentences*, which he accounts for in terms of possible worlds. (Kaplan devises a *dthat* operator, whose job is to rigidify descriptions):

> …the verbal form of ["the spy is suspicious"] might have been adopted by one who lacks "dthat" to express what is expressed by ["dthat (the spy) is suspicious"]. We seem to have here a kind of *de re – de dicto ambiguity* in the verbal form of ["the spy is suspicious"] and without benefit of any intensional operator. No question of an utterer's intentions has been brought into play. *There is no question of an analysis in terms of scope, since there is no operator.*
>
> (my *italics*)

But Kaplan stops short of ascribing the ambiguity to the definite description itself, rather than to the whole sentence. A similar line of reasoning is suggested by Partee (1970), where she suggests a *de re/de dicto* ambiguity in the non-embedded sentence, "The man who murdered Smith is insane". However, she takes this ambiguity to be the same as Donnellan's referential/attributive distinction – an assimilation that Kripke (1977) argued against. For details see Sections 2.2.2.1.3 and (in particular) 5.2.3.

[36] There is, however, one important difference to note between *de re* terms and actualised descriptions despite both having the same semantic function. The difference is that an actualised description, like "the actual spy", is distinguished *syntactically* from a descriptive term, like "the spy", but a *de re* definite description is not so distinguished from a *de dicto* definite description; the latter two *look* exactly the same – just "the spy". This, in fact, is the source of the *ambiguity* involved in the *de re/de dicto* distinction.

[37] Versions of this ambiguity view of definite descriptions are held, for example, by Kaplan (1978), Wettstein (1981) and Reimer (1998a; 1998b).

[38] Linsky L. (1963, p.80)

[39] Donnellan (1966/2001, p. 250)

[40] Donnellan (1966/2001, p. 249). We shall discuss this problematic distinction at considerable length in Section 5.2.

[41] It should be noted though that although both Linsky and Donnellan agree that the definite descriptions succeed in *referring* to the intended object, they are less committed to the *truth* of such statements. When it comes to the proposition as a whole, Linsky says that since the subject does not fit the description, the proposition itself lacks a truth-value – it is neither true nor false. By

holding such a view, he pretty much follows P. F. Strawson's (1950) intuition with regard to vacuous terms, who believed that "the present king of France" is neither true nor false. (As opposed to Russell (1919), who, in the light of his theory of description, took propositions that included such terms to be plainly false.) Donnellan is less clear on this matter. Discussing Linsky's example he intuits,

> it seems to me that we shall, on the one hand, want to hold that the speaker said something true, but be reluctant to express this by "It is true that her husband is kind to her."
>
> (1966/2001, p. 256)

At any rate, whatever the truth value of the proposition as a whole is, it is quite clear that both Linsky and Donnellan thought that definite descriptions may in some cases designate referents that fail to fit the description. Thus, it follows that a definite description on its own, is subject not only to a *de re* and a *de dicto* reading of it, but also to a referential reading of it.

[42] E.g., Kripke (1977); Salmon (1981; 1991); Bach (1987) and Neale (1990).

[43] E.g., Larson and Segal (1995, p. 348-9) argue that Donnellan's distinction is, *pace* Kripke, semantic, i.e., that a referential use expresses, semantically, reference to the man with the water in his glass. Quite a few others support a semantic interpretation, but most of them add a requirement, *pace* Donnellan, that in addition to being intended by the speaker, the referent *also* fits the description. This seems easier to defend for it does not involve designation of a referent that does not actually fit the description, as the above referential use allows. See Section 5.2.2.

[44] I.e., on some interpretations of the distinction it *need* not. That is to say, some take the referential to simply be rigidified (i.e., *either* directly-referring *or* reference-fixing). On such interpretation, a referential definite description designates, with respect to every counterfactual world, that which it designates in the actual world (however the actual referent is being picked-out). See discussion in Section 5.2.5.

[45] Frege (1918)

[46] E.g., Kaplan (1977/1989). For a non-modal version of the objection see Perry (1977).

[47] Kaplan (1978; 1989, p. 580). Similar views are held by Wettstein (1981); Devitt (1981a, who later, 2004, retracts from this view, and adopts a version of the "combined" view, to be presented shortly. See Endnote 53 below for a detailed description of his later view); and Neale (1993, p. 108).

[48] Notable defenders of this view are Larson and Segal (1995, p. 213).

[49] Braun (2001)

[50] There is a controversy whether it is the demonstration, e.g., gesturing, or the intention that does the picking-out job, and, as one can expect in philosophy, there are other controversies within each camp as well. Adherents of the demonstration thesis include Kaplan in his "Demonstratives" (1977/1989, pp. 589-91), McGinn (1981) and Stanley and Szabo (2000, pp. 220-1), whereas those favouring the intention thesis include Kaplan in his "Afterthoughts" (1989, pp. 582-4), Bach (1992a, 1992b), (arguably) Wettstein (1984), and Devitt (2004) who also presents arguments against the demonstration thesis. In want of a covering term, we shall use, for brevity, "gesture" to include both.

[51] Borg (2000, p. 242)

[52] As Borg (2000) indicates, Kaplan seemed to be attracted by such a view (see Kaplan, 1977/1989, p. 515). Others who also seem to broadly endorse a view of this type are Braun (1994), Richard (1993), King (2000) and Lepore and Ludwig (2000). The latter explicitly formalise "that F is G" as [the x: x = that *and* x is an F](x is G).

However, as Braun (2001) indicates, members of this camp are divided with regard to the question of whether the content of the noun phrase F is or is not a constituent of the content of the demonstrative: whereas Borg and Braun think it is not, Richard and King think it is.

[53] Devitt (2004) presents a somewhat different version of this view. On his view, a complex demonstrative designates, with respect to every possible world (actual *and* counterfactual), that which is intended in the actual world, *and* which fits the description in that world (actual or counterfactual). Thus, Devitt agrees with Borg on the designation in the actual world; yet he disagrees on the designation in counterfactual worlds: whereas on Borg's view a complex demonstrative, like any rigid designator, designates, with respect to every counterfactual world, that which it designates in the actual world, Devitt insists that it should also fit the description in that counterfactual world. Consequently, according to Borg, "that teacher of Alexander" will designate Aristotle in counterfactual worlds in which he is not Alexander's teacher, whereas on Devitt's view the term will fail to refer in such worlds. Accordingly, Devitt defines *weak rigidity*:

> *e* is weakly rigid iff it contains, implicitly or explicitly, an element that designates the same object in every possible world in which that object exists *and any descriptive element of e applies to that object.*
>
> (p. 296)

In other words, whereas a rigid term designates all and only things identical to the actual referent, a weakly rigid term designates only, but not all, objects identical to the actual referent. Devitt thus concludes that "'that F' is not rigid but only weakly rigid. (Devitt holds a similar view about referential definite descriptions.) However, Borg (2000, pp. 232-3) objects that since on such a

view "that philosopher" designates only philosophers, the sentence, "that philosopher might not have been a philosopher", which seems to be true, comes out false.

[54] The following examples are collected in Reimer (2003).

[55] Kripke's derivation from the *Shorter Oxford English Dictionary*.

[56] This seems to be Chalmers's view (see quotation in Section 1.4.1). "Water" stands for something like "the actual watery stuff". Since the actual watery stuff is (watery) H_2O, "water" designates H_2O (assuming materialist criteria of identity) in all counterfactual worlds. However, had the actual watery stuff been different, say, XYZ, "water" would designate XYZ in that actual world, and, subsequently, in all counterfactual worlds.

[57] Bernard Linsky (1984); Soames (2002)

[58] Salmon (1982); Linsky B. (1984); Schwartz (2002, pp. 268–9); Soames (2002, pp. 250–1). For recent defences of the rigidity of property terms see Marti (2004) and LaPorte (2006) but see Lopez de Sa (*forthcoming* 2007) and Devitt (2006) for respective counterattacks.

[59] Devitt (2005, p. 146). This definition, says Devitt, is a slight modification of the idea presented in Devitt and Sterelny (1999, pp. 85–6). A similar idea was advanced by Cook (1980).

[60] The following table expresses exactly Chalmers's account of names, such as "Hesperus", and of natural-kind terms, such as "water" (which he expresses in terms of such a two-dimensional apparatus). Thus, in our terminology, Chalmers takes names to be reference-fixing rather than directly-referring (e.g., Chalmers, 2006).

[61] The following exactly represents Kaplan's (1977/1989, p. 562; 1989, pp. 597-9; 1979, p. 404) view of names (which he expresses in a roughly similar two-dimensional apparatus. In his terms, names have *constant character*, where character is a function from contexts of utterance (that can be represented by possible actual worlds) to content (that can be represented as intensions – for Kaplan, a function from *circumstances of evaluation* to extensions.). In short, in our terminology, Kaplan takes names to be directly-referring rather than reference-fixing. Some of what Evans says about proper names makes it plausible to take him to hold a similar view (Cf. Chalmers, 2006, p. 585).

3   INTENSIONS: SOME METAPHYSICAL ASPECTS

Overview

This chapter is dedicated to an elaboration of the metaphysical aspect of intensions. At the centre of the discussion stands the notion of *sortal* property.

We begin by realising that the joint role of the factor of the criteria of identity and of the factor of the actual referent in our intensions-formula is in fact to determine together the sortal property of the actual referent.[1] In particular, the criteria of identity single-out one of the properties of the actual referent, and render it sortal. This is an instantiation of a more general relation between second-order properties and first-order properties: a second-order property, such as "the material constitution of the object" "homes-in" on some first-order property, such as being H_2O. The notion of sortal is dominant in the metaphysical tradition (ever since Locke, and mostly during the 1960's and 1970's). Thus, using the notion of sortal serves to highlight ways in which our discussion links to that tradition.

We then discuss two problems concerning sortal properties. The first is that sortal properties seem to be un-analysable *ad hoc* properties (e.g., the sortal property of H_2O appears to be just being H_2O.) This seems of little explanatory value. The second is the problem of criteria of identity for individual objects (as opposed to kinds, like tiger, or masses, like water), where sortal properties, although necessary, are not sufficient. E.g., whereas the criterion for being identical to water is having water's sortal property (on a material view, being constituted by H_2O), it seems not true that the criterion for being identical to Fido is having Fido's sortal property – which is, presumably, being a dog. Something can be a dog without being Fido. The notion of *individual essence* is discussed in this context.

Next, a classification of (first-order) properties is proposed, which is based on an analysis of properties in terms of sufficiency and necessity. This analysis leads to a possible-worlds account of the differences between various types of properties, such as

essential properties, sortal properties, secondary substances, contingent properties, and others.

Lastly, we shift to discuss supervenience, and in particular, whether physicalism is consistent with non-physical criteria of identity. Physicalism is the view that the manifest, as well as all other aspects of the world, supervenes upon the physical. The view that criteria of identity are physical is the view that the physical accounts for what objects ultimately *are*, whereas the manifest, as well as other aspects of objects, merely accounts for what objects *are like*. It is argued that despite appearance the two theses are distinct and mutually independent.

3.1 Coarse-Graining the Intensions Formula

A rigidified term designates, with respect to every counterfactual world, the same referent that it designates in the actual world. E.g., if "water" is a rigidified term, then it designates, with respect to every counterfactual world, that which is identical to its actual referent, the watery H_2O stuff. However, we have seen that being the same is dependent upon the following two factors: the criteria of identity for that referent, and (the properties of) the actual referent.[2] E.g., on the "material view", the criterion for being identical to the watery H_2O stuff is having the same material constitution as that stuff, namely, being constituted by H_2O (whether watery or not); on the "manifest view", by contrast, some stuff is identical to the watery H_2O stuff iff it has the same manifest properties as that stuff, namely, iff it is watery (whether it is constituted by H_2O or not). It follows that a criterion of identity somehow singles out one (or a certain collection) of the properties of the actual referent, and gives it a special status: whatever has *that* property is identical to the actual referent. For instance, in the case of the watery H_2O: on the material view this property is 'being constituted by H_2O', whereas on the manifest view this property (or, more accurately, this collection of properties) is 'being watery'.

Thus out of all of an object's properties, one (or a certain collection of properties) is privileged; it is strongly linked with what it is to be the same as that object. Different views differ on what this property is. (Realists on this issue hold the special status of such properties to be real, i.e., independent of description or perspective, whereas others relativise it to description or to perspective.[3] However, relativised or not, the special status of such properties remains.) Due to its privileged status, this type of property has gained enormous attention throughout the history of philosophy. The

traditional concepts which are used with relation to this privileged property are notably *substance*, *essence*, and *sortal.*

The concept of "sortal" was introduced by Locke,[4] but is strongly related to Aristotle's key distinction between the category of substance on the one hand, and the rest of the categories on the other.[5] It is thus generally associated with: criteria of identity[6] (across times – i.e., *persistence conditions*; and across worlds – i.e., *transworld identity*); the *what-is-it* question (as opposed to *what-is-it-like* questions); and, *essence* (as opposed to *accident*).

However, as is not uncommon with important philosophical notions, philosophers vary considerably in their application of the term. One such important difference is to do with what "sortal" is taken to stand for. Grandy observes that,[7] P. F. Strawson takes "sortal" to stand for universals;[8] Quine takes it to stand for predicates;[9] Wiggins takes it to stand for concepts (indeed, he also admits sortal-predicates);[10] and Geach uses his alternative term *substantival term*, which, like Quine, he takes to stand for linguistic expressions.[11] Another important difference is that many philosophers restrict the term to countable kinds of objects like people and trees (or, accordingly, to count-nouns), whereas others extend its use to apply to stuffs as well, such as water and gold (or, accordingly, to mass-terms).[12]

Where do we stand in relation to these differences? Our interest in intensions (and in particular, our interest in the intension of rigidified terms) incline us as follows: firstly, we are concerned with the privileged *properties* that are linked with criteria of identity, no less that we are concerned with concepts or predicates; secondly, we are interested in referents in general, i.e., in masses like water and gold just as much as in countable objects like dogs and people; and thirdly, we are interested in identity in counterfactual worlds in general, i.e., in being the same *kind* as Fido, or as water, as well as in being the same *individual* as Fido.

That said, should *we* use the term "sortal" for this purpose?

I believe we should. Fashioning new terms increases the terminology beyond necessity on the one hand, and conceals the striking resemblance of the corresponding terms' subject matter on the other hand. In addition, using "sortal" embeds our present discussion appropriately within the metaphysical tradition.

Thus, we shall use the term "sortal" in the following way: (a) to include not only count-nouns like "dog" and "tree", but also mass-terms like "gold" and "water"; (b) to apply to terms as well as to properties; thus, 'being a dog' and 'being gold' are considered

here sortal *properties*; and (c) We shall introduce the notion of *individual sortal property* (by which we extends the traditional use of "sortal"): the property of "being a dog" may be said to make Fido the *kind* of thing that it is; having this property guarantees being of the same kind, or of the same essence, as Fido is. It does not guarantee *numerical* identity. An *individual sortal* property may be said to make Fido the particular *individual* that it is; having this individual sortal property guarantees being *numerically* identical to-, i.e., having the same individual essence as-, Fido.

We shall now state the difference between different metaphysical views about criteria of identity in terms of sortals.

Take water again. On the material view, being the same stuff as the watery H_2O simply amounts to being constituted by H_2O; we may conclude then that on that view, the sortal property of the watery H_2O is 'being constituted by H_2O'. On the manifest view, by contrast, being the same stuff as the watery H_2O amounts to just being watery; hence, the sortal property of the watery H_2O stuff is 'being watery'. Overall then, the intension of a natural-kind mass-term like "water", which is a rigidified term, is simply determined by the sortal property of the actual referent. (The case of singular rigidified terms, like proper names and (*de re*) definite descriptions, is slightly different, as we shall shortly see.)

As we have noted at the beginning of this section, the sortal property is somehow determined by the criteria of identity together with (the properties of) the actual referents: the criteria of identity single out one of the properties of the actual referent, and render it sortal. Thus, as far as intensions are concerned, the job of the former two factors can be seen as simply that of determining the sortal property. In other words, we could replace these two variables in our formula with just the one variable of the sortal property. This would be a "coarse-grained" version of our formula. Yet using the two variables makes explicit *how* this sortal property is exactly determined. At any rate, the two versions are, of course, versions of the same formula.

In the next section, we expose the structure of the relations between these three elements: the sortal property, and the two factors that determine it: criteria of identity and (the properties of) the actual referent.

3.2 Sortal Properties: An Analysis in Terms of First-Order and Second-Order Properties

According to an appealing view inspired by Aristotle, properties of objects "fall under" some fixed categories, e.g., the object's height; its weight; its colour; its function etc.[13] So if a given object's colour is white, then the property under the category of 'being the object's colour' with respect to this object, is '(being) white'.

'Being a colour' is a property of (being) white; hence it is a property of a property. In other words, 'being a colour' is a *second-order* property. It follows that, 'being the object's height'; 'being the object's weight'; 'being the object's function' etc. are all second-order properties. Accordingly, '(being) white' is a *first-order* property. Thus we may say that an object's first-order properties *fall under* second-order properties. Correspondingly, a second-order property, with respect to a certain object, *homes-in* on some first-order property.

Now the same second-order property may be applied to different objects and thus home-in on different first-order properties. For instance, the second-order property 'being the object's colour', when applied to Mars, homes-in on the first-order property '(being) red', but when applied to the moon, homes-in on the first-order property '(being) white'. We may thus ask what first-order property falls under a given second-order property when that second-order property is applied to a certain object.

We can use this structure to analyse the relations between sortal properties and criteria of identity.

A material criterion of identity for x states that y is identical to x iff it shares with x the same material constitution; now 'being the object's material constitution' is a property of a property, e.g., with respect to the watery H_2O, it is a property of the property '(being constituted by) H_2O'. Thus 'being the object's material constitution' is a second-order property. It follows that some second-order properties are linked with criteria of identity (for example, on the material view, 'being the material constitution of an object'), whereas others are not (on that same view, 'being an object's colour' for example).

Second-order properties, when applied to objects, home-in on first order properties. The special feature of the second-order properties that are linked with criteria of identity is that the first-order properties on which they home-in are the *sortal* first-

order properties of their objects. Or, to follow the tradition, simply, they are the objects' sortal properties. For example, on the material view, the second-order property 'being the object's material constitution' when applied to watery H_2O, homes-in on the first-order property '(being constituted by) H_2O', thereby rendering it sortal. We can thus call this type of second-order property, a "*sortal* second-order property", to distinguish it from other second-order properties.

From a rigorous philosophical point of view, in principle, *any* second-order property may be the sortal one. E.g., considering the second-order property 'the object's function' to be sortal suggests that the object is primarily a *functional* object, which merely happens to have both its manifest properties, and its material constitution (we seem to have such strong intuitions when attending to artefacts or to organisms). In other words, on this view, the object's material constitution and manifest properties are merely what the object *is like* and not what it *is*. Similarly, (and highly counter-intuitively,) considering the second-order property 'being the object's colour' to be sortal suggests that the object is primarily a coloured object, which merely happens to have manifest, material and functional properties.

In sum, it is proposed that the relation between (i) criteria of identity; (ii) the actual referent; and (iii) the referent's sortal property, are the following. A metaphysical view is characterised by the second-order property that it takes to be sortal. Like any other second-order property, this sortal second-order property homes-in on one of the object's first-order properties. This first-order property is thereby rendered a sortal (first-order) property. (Another way to put it is to say that a sortal-second order property is simply a *function* from objects to (first-order) sortal properties. It assigns to each object its first-order sortal property.) Being the same as that referent (at least as far as kinds and mass-objects are concerned) is simply having that sortal property. Different metaphysical views take different second-order properties to be sortal, and, consequently, different first-order properties to be sortal.

This structure has the following implications with respect to intensions. The intension of a rigidified term is determined by the sortal property of the actual referent. The sortal property is one (or a collection) of the first-order properties of the referent. Which of the object's first order properties is the sortal one is fixed by two factors: the list of the object's first-order properties, and the criteria of identity, as detailed above. Different metaphysical views differ in the second-order property which they take to be

sortal. And that entails difference in the (first-order) sortal properties of objects. Eventually, such differences result in different intensions.

(It follows that there are in fact two ways to specify an object's criteria of identity. One way is in terms of second-order sortal properties, e.g., on the material view, the criterion for being identical to water is having the same material constitution as water. Another way is in terms of first-order properties, e.g., on the same view, the criterion for being identical to water is being constituted by H_2O. The first is indirect, implicit and more general; the second is direct, explicit and particular. (This difference is significant from an epistemological point of view; to be discussed in Section 4.1.3). However, both ways are but different descriptions of the same criterion of identity.)

3.2.1 Two problems concerning sortal properties

3.2.1.1 The ad hoc-ness of sortal properties

One may object to sortal properties on the following grounds. What is the sortal property of this dog? Well, it is the property 'is a dog'… but this seems hardly informative. In fact it is trivial. Putnam expresses this worry thus:

> If we are allowed to invent unanalysable properties *ad hoc*, then we can find a single property – not even a conjunction – the possession of which is a necessary and sufficient condition for being a lemon, or being gold, or whatever. Namely, we just postulate the property of being a lemon, or the property of being gold, or whatever may be needed.[14]

(Although Putnam is talking about these properties in terms of necessary and sufficient conditions rather than in terms of sortals, we shall soon demonstrate the close link between the two in Section 3.3.5. For the time being, we can rely on an intuitive sense of this link.)

In other words, the worry is that those who use sortal properties to account for what an object is, are committing the so-called *virtus dormitiva* fallacy, termed after Moliére's doctor in *The Hypochondriac*, who explains that opium makes one sleep "because there is a dormitive virtue in it whose nature is to cause the senses to become drowsy."[15]

Thus, it seems that instead of postulating that the sortal property of gold is 'being gold' we should come up with something more informative, like, say, 'having an atomic number 79' (or, alternatively, contrary to the consensus, 'being golden'). Similarly, instead of postulating that the sortal property of water is 'being water' we are better off

suggesting something like 'being H_2O' (or, alternatively, 'being watery', i.e., drinkable, potable, etc.'). Thus, when it comes to dogs, we may simply expect a sortal property to tell us what being a dog *consists in*, i.e., to reveal the underlying nature, or "essential nature",[16] of our subject. Yet however appealing this may sound, the suggestion has its limits. Consider the view according to which the sortal property of water is 'being H_2O'. But what is the sortal property of H_2O? It seems that there is very little to say more than, again, 'being H_2O' (i.e., 'being stuff compounded of molecules each of which is composed of one oxygen atom and two hydrogen atoms'). Thus the problem is that after some recursive applications of the question "what is the sortal property of …?" it seems we are bound to arrive at a stage in which the sortal property of an F is simply 'being F'.

However, as I understand Putnam's worry here, it is not about the mere fact that the sortal property of F is 'being F'. It seems that Putnam will have little trouble with the fact that the sortal property of H_2O is 'being H_2O'. I.e., Putnam seems not too concerned with a rock-bottom sortal property, as long as this property is *informative* (however 'being informative' is to be cashed out). What seems to be worrying Putnam is rather "abusing the system". I.e., abusing the fact that we accept rock-bottom sortal properties by introducing *uninformative* properties as sortal properties. E.g., 'being a lemon' is uninformative (for Putnam, but maybe not for the Martians), and hence should not be introduced as the sortal property of lemons; whereas 'being H_2O' is informative (again, for Putnam), and hence may be introduced as a sortal property of water, or, indeed, of H_2O.[17]

Overall then, the mere fact that the sortal property of some F is 'being F' does not in itself indicate that 'being F' is an *ad hoc* property. It is only when this property is uninformative, and is introduced as a pseudo criterion of identity, that it can be blamed of being introduced *ad hoc*. The problem remains, of course, how to clearly define what being informative amounts to. We shall not pursue this here however.

3.2.1.2   The problem of criteria of numerical identity

The criterion for being identical to water – the watery H_2O stuff – is having the sortal property of that watery H_2O stuff. Some believe this sortal property to be '(being constituted by) H_2O'; others may take it to be '(being) watery'. In general, the criterion for being identical to certain stuff is having the sortal property of that stuff. The same seems to hold for (natural) kinds. The criterion for being of the same (natural) *kind* of

Fido, is having the sortal property of Fido, namely, 'being a dog'. But what about individual objects? What about numerical identity? Could we say that the criterion for being numerically identical to a certain object is having that object's sortal property? It seems that although this is a necessary condition for identity, it is by no means a sufficient one: merely having the property '(being a) dog' does not guarantee being Fido; for something that instantiates this property may be *any* old dog. We therefore need something more here. In other words, having the sortal property of an object is a *necessary* condition for being identical to it, yet it is not *sufficient*. Something additional is required to make it sufficient.

In the case of identity over time, there seems to be a relatively simple solution at hand. Consider the clay statue David again. Or, more neutrally, the object that has a David shape and that is made out of clay. On a temporal version of the puzzle, David is smashed, so that at t_2, we have a clay ball. Is the clay ball at t_2 identical to the clay statue David at t_1? If we hold the manifest view (and hence think that David is distinct from the clay ball), we cannot simply say that an object is identical to David as long as it *has* David's shape, for that will entail that any David-shaped statue is identical to David.[18] However, we can say instead that an object remains the same so long as it *continuously keeps* its shape, which will get us where we want. Similarly, on the 'material view', we may say that an object remains the same as long as it continuously keeps its material constitution (hence on this view David is indeed identical to the clay ball). Thus, whereas the criterion for being the same *stuff* as S is merely *having* the sortal property of S (and similarly, the criterion for being of the same (natural) kind as object O is *having* the same sortal property of O), the criterion for being *numerically* identical at t_2 to an object O at t_1 is *continuously keeping* the sortal property of O from t_1 to t_2 (and not merely *having* it).

This solution, however, cannot be applied to transworld identity. Since it seems to make little sense to talk about *continuously keeping* the sortal property from one counterfactual world to another. Worlds do not succeed each other like moments in time, and are not causally related to one another like the temporal stages of an object.

The difficulty to provide criteria of transworld identity and the need to find such criteria are nicely manifested in Chisholm's role-switching puzzle:[19] The puzzle is about Adam and Noah. There is a counterfactual world in which Adam is a bit more like actual Noah than in the actual world, and Noah is a bit more like actual Adam than in the actual world. In another counterfactual world, Adam is even more like actual Noah, and vice

versa. Continuing like this, we shall arrive at a counterfactual world in which Adam is entirely like Noah and vice versa. It follows that there is a counterfactual world which is exactly like ours in every respect, apart from the fact that Adam and Noah have switched roles, i.e., that the guy that is called "Adam" in the counterfactual world, and has all the attributes of our Adam (including the relational ones, like being coupled with Eve etc.), is in fact distinct from our actual Adam; in other words, that that guy and our Adam are indiscernible in *every* respect yet are numerically distinct; and, furthermore, that that guy and our Noah, who are different in almost every respects, are nevertheless numerically identical.[20] Similarly, the guy called "Noah" in the counterfactual world and our Noah, although qualitatively-identical, are numerically distinct, whereas that guy and our Adam who are different in almost every respects are nonetheless numerically identical. Moreover, it seems that the same can be applied to *any* two objects in the world. Which entails that there is a counterfactual world that is qualitatively entirely identical to ours, and yet that all objects have switched roles; e.g., you are identical to the Eiffel tower, I am identical to your bicycle, etc. This seems like a very unfortunate consequence that needs to be somehow avoided.[21]

One may hope to avoid this consequence by alluding to *essential properties*. E.g., Adam cannot switch roles with an apple since Adam is *essentially* a person, and the apple is *essentially not*. So in the move from a world in which Adam is a man to the world in which he is an apple, and thus not a person, we left an essential property of Adam "(and therefore Adam himself) behind."[22] But this is of little help, since Adam can still switch roles with Noah, who is also essentially a person (as well as with any other person). Thus, in order to completely block *any* role-switching possibility, Adam has to have some essential property (or collection of properties) which *no other individual* has. In other words, Adam has to have a property which is not only *necessary* but also *sufficient* for him. In Chisholm's words, in order to block the possibility of role-switching with an object x, x has to have a property E, such that: "x has E in every possible world in which x exists; and, moreover, for every y, if y has E in any possible world, then y is identical with x."[23] Thus, this property (or collection of properties) is fundamentally a *unique* essential property of Adam; something like 'Adamness'[24]. Some called this property *individual essence*. A version of this view, which takes individual essences to be non-qualitative, i.e., a *primitive thisness*, dates back to Duns Scotus (1266-1308), who famously called such an individual essence *haecceity*.[25] Whatever you call it, such individual essences in fact provide criteria of transworld identity. In other words, in order to block role-switching scenarios

as described above, we need criteria of numerical transworld identity, and these seem to take the form of individual essences.

(This is the reason why the same problem does not arise for stuffs: stuffs, like water, have a unique essential property, which is just their sortal property. We cannot construct a series of counterfactual worlds in which water and gold switch roles, since the moment water ceases to posses its sortal property (e.g., being H_2O according to the material view), it stops being water; the same goes for gold. So a similar role-switching scenario is barred.)

Kripke proposed something like an individual essence, in the form of his necessity of origin thesis (discussed in Section 1.5.2).[26] Thus, for instance, it is essential for Noah to be the product of the union of a particular sperm and a particular egg. Anyone who originated from a different gamete cannot be Noah. Indeed, it seems not only essential, i.e., necessary, but also sufficient. Anyone originating from this gamete is by that mere fact Noah; no-one else can originate from this gamete. However, as we have seen, this suggestion came under heavy attack from Salmon (Section 1.5.2. Note that Salmon's objection is in fact a version of Chisholm's argument).

Kripke's thesis of the necessity of origin is a specific thesis about the individual essence of objects. It may be wrong. Refuting it, however, does not refute the claim that objects have *some* individual essences. A rigidified term that refers to an object, designates, with respect to every counterfactual world, *the same object* that it designates in the actual world. The notion of "the same object" presupposes some criteria of numerical identity. It is hard to think of any such numerical criteria of identity that do not involve *some* individual essence. Admittedly, however, it is extremely difficult to come up with suggestions for such non-primitive individual essences; e.g., with what "Adamness" consists in.

Given the equivalence between essences and sortal properties, we may call such individual essences, "*sortal* individual properties".

3.3 First-Order Properties: An Analysis in Terms of Necessary and Sufficient Conditions

Sortal properties are but one category of first-order properties. Other types of first-order properties are, for example, essential properties and contingent properties. In this

section, I propose a classification of (first-order) properties, based on their analysis in terms of necessary/not necessary and sufficient/insufficient conditions. This analysis allows for a more precise and less intuitive account of the distinction between different types of first-order properties. By fitting all those properties into a unified, common framework, I hope that an illuminating comparison is provided. This classification, however, stands quite independently of the main argument advanced in this dissertation, which is the general formula of intensions.

3.3.1   Properties as conditions

The issue of properties – their existence, nature, and criteria of identity – is a matter of considerable controversy in philosophy. Thus, although it is commonly accepted that Mars is a planet and that Earth is a planet too – viz., that both Mars and Earth have the property 'being a planet' – there are very different ways to spell this fact out. One traditional way, for example, is to say that Mars and Earth share a universal (*realism*)[27]. Another way is to say that properties are particulars, named *tropes*, rather than universals; i.e., Mars's property of being a planet is numerically distinct from Earth's property of being a planet, so the fact that both objects are planets merely amounts to the fact that both have *similar* such particular properties (*trope-nominalism*).[28] Another way to avoid alluding to universals is to say that both Mars and Earth simply belong to the same class (*class nominalism*).[29] A related view holds that terms like "planet", that seem to refer to properties, in fact refer to concepts (*conceptualism*).[30] Yet another way would be to say that Mars and Earth do share a universal, but not of being a planet, for such a universal does not exist; rather, they share many "fundamental" universals, possibly physical, which being a planet is but an abbreviation of, or, alternatively, *supervenes upon* (we shall discuss the supervenience relation later in this chapter). It is also commonly accepted that Mars weighs more than three kilograms. This fact may be accounted for in terms of *truth-makers*, namely, that Mars is of a certain *given* weight, which is the truth-maker of the above fact.[31] Other analyses would provide different accounts for this fact.

Our discussion involves properties. Yet it is not in the interest of our study to engage in an analysis of the *nature* of such properties. In view of such controversies, we should thus seek a *neutral* use of properties, which would be committed to none of the above specific accounts. The more neutral our use of properties will be in this respect, the wider its application can be.

What would such a neutral conception be?

As was stated above, all views agree on some basic neutral facts; e.g., that Mars and Earth share the property of 'being a planet', and that Mars has properties such as, 'weighing more than three kilograms', as well as, 'being red'; 'being larger than the Earth'; 'being the fourth planet from the Sun'; etc. Now the feature that is common to all these attributes is that each states some *condition* that Mars meets. And this is regardless of how meeting these conditions is spelled out in terms of the nature of properties. So having a property, most neutrally, can be simply conceived as meeting a condition. We may thus use "property" in our discussion just in this way; i.e., whenever we say that an object has a certain property this simply amounts to asserting that the object meets a certain condition.

3.3.2   Necessary and sufficient conditions

The relation between an object and its properties is that of instantiation: objects instantiate properties.[32] There are logical relations between properties and the objects that instantiate them. Two such logical relations are necessity and sufficiency. A property is necessary to an object iff the mere presence of the object entails the instantiation of that property, i.e., iff the absence of the property entails the absence of the object. A property is sufficient to an object iff the mere instantiation of the property entails the presence of the object, i.e., iff the absence of the object entails the absence of the property. A property may be necessary to its object or not; a property may also be sufficient to its object or not. Consequently, there are four different options for logical relations between a property and the object that instantiates it (the following examples for each such relation will rely on everyday common views[33]): (a) the property is neither necessary nor sufficient for its objects, e.g., 'being pale' seems neither necessary nor sufficient to Socrates (to follow Aristotle's classic discussion); (b) the property is necessary but not sufficient to that object, e.g., 'being mammal' is necessary but not sufficient to Socrates (since the presence of Socrates entails that he instantiates this property, but the mere instantiation of the property does not guarantee that it is instantiated by Socrates – it may be Parmenides, or, for that matter, Fido); (c) the property is sufficient but not necessary for that object; e.g., 'knowing that one is Socrates' (for the mere instantiation of this property entails that it is instantiated by Socrates – no other individual can possibly have this property, yet Socrates' mere presence does not guarantee this property's instantiation

– Socrates might have doubts, or be mistaken, about who he is); and (d) arguably – according to the necessity of origin thesis – 'originating from *this* gamete (i.e., from the sperm and egg from which Socrates actually originated)', for Socrates. According to this thesis, the mere presence of Socrates entails the instantiation of this property, and the mere instantiation of this property entails that the individual who instantiates it is Socrates. Less controversial examples are: (i) 'having atomic number 79' for gold; the presence of gold entails the instantiation of this property, and the instantiation of this property entails being instantiated by gold. Neither can exist without the other; and (ii) 'being the sum of 2 and 3' is necessary and sufficient for the number 5. An object cannot be the number 5 without having this property, and something cannot have this property without being the number 5.

Interestingly, whereas properties of the first, second and fourth categories are widely discussed in the literature, I could not find any discussion, or even mention, of properties of the third category as such.

The above definitions are stated in terms of logical dependence. The domain of application may vary. It may be confined to the actual world. It may, however, extend to all possible worlds. This will be the modal interpretation of necessity and sufficiency: A property is necessary to an object iff that object instantiates that property in every counterfactual world; and a property is sufficient for its object iff in every counterfactual world in which that property is instantiated, it is instantiated by that very object. Thus, the above four categories will be: (a) a property is neither necessary nor sufficient for its object iff there are some worlds where the object does not instantiate the property and some worlds in which the property is instantiated by another object (e.g. 'being pale' for Socrates); (b) a property is necessary but not sufficient for its object iff there are no worlds in which the object fails to instantiate this property, but there are some worlds in which it is instantiated by other objects (e.g. 'being mammal for Socrates); (c) a property is sufficient but not necessary for its object iff there are no worlds in which it is instantiated by other objects, but there are worlds in which the object fails to instantiate it (e.g., 'knowing that one is Socrates' for Socrates); and (d) a property is necessary and sufficient for its object iff there are no worlds in which the object fails to instantiate the property and there are no worlds in which that property is instantiated by another object (arguably, 'originating from this gamete' for Socrates).

This framework proves very useful for the purpose of accounting for various central distinctions concerning properties. The next five short sections aim to illustrate this claim.

3.3.3   Essential vs. contingent properties

Let us begin with one such common distinction, namely, between *essential* and *contingent*, or *accidental*, properties. A property may be necessary to the object that instantiates it, or not necessary to its object, regardless of whether it is sufficient to its object or not. An essential property is just a property which is necessary – whether sufficient or not – to its object; an object simply cannot survive its loss.[34] E.g., being a mammal is an essential property of Socrates (and also of Fido) – he could not survive losing this property; similarly, having the atomic number 79 (on the material view) is essential to gold – gold could not survive losing this property. A contingent property, by contrast, is one that is not necessary – whether sufficient or not – to the object that instantiate it, viz., an object can survive its loss. E.g., being pale is contingent to Socrates. Similarly, knowing that he is Socrates is contingent to Socrates.

A similar distinction can be drawn regarding sufficient properties. A property may be sufficient to the object that instantiates it, whether it is necessary to it or not; or, it may not be sufficient to its object, whether it is necessary to it or not. Yet I know of no special title for either of these two categories. This fact may indicate that, for some reason or other, we seem to have no special interest in such properties.

3.3.4   Criteria of identity vs. secondary substances

The category of essential properties includes *all* necessary properties; it thus has two sub-categories: essential properties that are also sufficient to their object on the one hand, and essential properties that are not sufficient to their object on the other. An example of the latter category would be being a mammal for Socrates. There are good reasons to believe that the properties in this sub-category are what Aristotle called *secondary substances*. The other sub-category within essential properties, namely, properties that are both necessary and sufficient to their objects, is in fact the group of properties that serve as the criteria of identity of their objects. Let me demonstrate this.

Criteria for being identical to X are defined as some necessary and sufficient *conditions* for being identical to X. Now, recall that according to our neutral definition of properties, having a property is just meeting some condition. Thus, meeting some necessary and sufficient conditions is just having some necessary and sufficient *property*. It thus follows that the criterion for being identical to X is simply having a *property* which is necessary and sufficient for X. E.g., the criterion for being identical to gold is having a property which is necessary and sufficient for gold – widely believed to be, 'being stuff with atomic number 79'. Anything that fails to meet the condition/to have that property, is not gold, and anything that meets the condition/has that property, is gold. Similarly, the criterion for being identical to Kripke is having a property which is necessary and sufficient for Kripke – believed by some to be, 'originating from the gamete from which he actually originated'.

(Recall that criteria of identity can be given in terms of second-order properties, as well as in terms of first-order properties. The latter is simply a more explicit and particular way of specifying them. See end of Section 3.2.)

To conclude, within the group of necessary properties, which is just the group of essential properties, the sub-group of necessary *and sufficient* properties is the group of properties the having of which form the (explicit) criteria of identity of the objects that instantiate them. The other sub-group within the group of essential properties if that of properties that are necessary but not sufficient (also known as secondary substances).

3.3.5   Sortal properties

In Section 3.2 we established that *a* criterion for being identical to X is having the (individual) sortal property of X. Thus, a criterion for being identical to water is having the sortal property of water (whether it is 'being watery' or 'being constituted by H_2O'); a criterion for being numerically identical to Kripke is having the *individual* sortal property of Kripke, (whether it is originating from a certain gamete or something else). In the previous section we established that *the* criterion for being identical to X is having a property which is *necessary and sufficient* for X. It thus follows that all sortal properties are necessary and sufficient for the stuffs that instantiate them; similarly, all sortal *individual* properties are necessary and sufficient for the individual objects that instantiate them. However, does the converse hold as well? Are all necessary and sufficient properties sortal properties?

It is tempting to think that they are. Indeed, it is tempting to think that necessary and sufficient properties are just sortal properties. This, however, appears to be wrong.

Intuitively, a sortal property states what the object *is*, as opposed to what it *is like*. Thus, according to the material view, the sortal property of water, for instance, is (being constituted) by H_2O, i.e., H_2O is what water *is*, whereas being watery is merely what water *is like*. (The manifest view takes it to be the other way around). Now consider the number 5.[35] The number 5 has the property 'being the sum of 2+3'. This property is necessary and sufficient for the number 5; every object that instantiates that property is the number 5, and every object that is the number 5 instantiates that property. But, so is 'being the quotient of 10 divided by 2'. And so on; an infinite list of properties, all of which are necessary and sufficient for being the number 5. Yet we would not wish to claim that each one of these properties makes the number 5 what it *is*, as opposed to what it *is like*. Firstly, because we feel that being the quotient of 10:2 is only a *characteristic* of 5, and not what being 5 consists in. And secondly, because it seems wrong to suppose that the same thing will have many properties, each of which *on its own* makes the number 5 what it is. Similarly, consider the property 'being identical to Socrates', for Socrates. Merely having this property guarantees being Socrates, and the mere presence of Socrates guarantees the instantiation of this property. Being identical to Socrates is thus a necessary and sufficient property of Socrates. Yet, again, we would not wish to say that being identical to Socrates is what Socrates *is*; i.e., that being identical to Socrates is the (individual) *sortal* property of Socrates.

We may conclude then, that there are necessary and sufficient properties that are not sortal properties. And generally, although every sortal property is necessary and sufficient for the stuff that instantiates it (and every *individual* sortal property is necessary and sufficient for the individual object that instantiates it), interestingly the converse does not hold; there are necessary and sufficient properties that are nevertheless not sortal-properties. It turns out then that sortal (individual) properties are a sub-category of the category of necessary and sufficient properties.

It should be noted though that for the most part, criteria of identity do seem to overlap with sortal properties (and individual sortal properties). It seems that the only cases in which the two diverge are the case of properties of mathematical objects, and the case of the property 'being identical to X'. Outside these, the two types of properties –

(individual) sortal properties and necessary and sufficient properties – do appear to coincide.

Why is it that the two *almost* completely overlap, yet not *entirely* overlap?

Here is a suggestion. The reason for the divergence lies in a difference in orientation between sortality on the one hand and necessity and sufficiency on the other. Specifically, necessity and sufficiency are about *spotting* something, whereas sortality is about *accounting for* something. Fundamentally, these are different orientations. E.g., the property of being Smith's favourite drink *spots* water, but it is very different from an *account of* what makes water what it is. This difference can be clearly shown in terms of counterfactual worlds: although being Smith's favourite drink spots water in the actual world, it spots other stuffs in other worlds. However, extending the requirement of spotting to be modal, i.e., extending it to spot water, and only water, in *every* counterfactual world – which is what necessity and sufficiency (as used in our discussion) is about – will bring the two types of orientations much closer. For it would be difficult to find a property that merely spots all and only water, and yet is not what makes water the stuff that it is (the only such property I am familiar with is the above 'being identical to water'). Think of a property that is necessary and sufficient for some non-mathematical stuff (or object), excluding the property of being identical to that stuff (or object), and I believe you will come up with the sortal property of that stuff (or the individual sortal property of that individual object). However, lacking a proof for that principle, it remains a mere conjecture.

3.3.6   Essential properties, sortal properties and essence

We have established that sortal properties are a sub-category of necessary and sufficient properties. Essential properties, by contrast, are necessary properties – whether sufficient or not. Thus necessary *and* sufficient properties are a sub-category of essential properties. It follows that sortal properties are *ipso facto* a sub-category of essential properties.

Thus every sortal property is an essential property but not vice versa. This last point is a rather important one. On the material view, the *sortal* property of gold is 'having atomic number 79'. Clearly, 'having atomic number 79' is also an *essential* property of gold on that view (gold has atomic number 79 in every counterfactual world). However, gold has other essential properties as well. For instance, being a chemical substance. Gold is a chemical substance in every counterfactual world. Yet merely being

a chemical substance is not *sufficient* for being gold, i.e., it does not guarantee that something is gold; it may be, say, lead.

Now sometimes people talk about *essence*. Essence has to be distinguished from essential properties. E.g., although 'being a chemical substance' is *essential* to gold – gold cannot survive its loss – 'being a chemical substance' is nonetheless not *the essence* of gold – otherwise it will be the case that gold and hydrogen share the same essence. However, 'having atomic number 79' (on the material view) is both the essence of gold as well as an essential property of gold. Thus, essence is a sub-category of essential properties. In our terminology, essence is just a sortal property. Or, as Locke put it:

> … the *Essence* of each *Genus,* or Sort, comes to be nothing but that abstract *Idea*, which the General, or *Sortal* (if I may have leave so to call it, from *Sort*, as I do *General* from *Genus*) Name stands for. And this we shall find to be that, which the word Essence imports, in its most familiar use.[36]

(Although Locke's view involves an additional anti-realist component, to which we need not commit. See our discussion of nominal substances in Section 4.2.1.1)

However, some may feel that being a chemical substance, although not *the* essence of gold, may be held *part of* the essence of gold. This view implies that the essence of something is just the collection of its essential properties; e.g., that the essence of gold *includes* the essential property of being a chemical substance, as well as the other properties that are essential to gold. Thus, according to such a view, the condition that is embodied in the sortal property amounts exactly to the sum of conditions that is entailed by the set of all of an object's necessary (i.e., essential) properties. E.g., the condition of having atomic number 79 includes the condition of being a chemical substance etc.

In sum, necessary and sufficient properties (e.g., being the quotient of 10:2 for 5) are a subclass of essential properties. Sortal properties (arguably, being the successor of 4) are a subclass of necessary and sufficient properties. The complementary subclass is that of necessary yet not sufficient properties (e.g., being a number for 5). Thus sortal properties are a subclass of essential properties. Essences, as opposed to essential properties, are just sortal properties. The two should not be confused. Some believe essences, i.e., sortal properties, to simply be the sum of an object's essential, i.e., necessary, properties.

3.3.7   Summary: the analysis of properties in terms of necessary and sufficient conditions

To sum up: the relation between an object and each of its properties is that of instantiation. Two logical relations between properties and the objects that instantiate them are necessity and sufficiency. A property may be necessary to its object or not; a property may also be sufficient to its object or not. Consequently, a property's relation to the object that instantiates it may be any of the following four (a) both necessary and sufficient; (b) necessary but not sufficient; (c) not necessary but sufficient (a rather neglected category for some reason); and (d) neither necessary nor sufficient.

These relations provide means to clearly compare and contrast the following different types of properties. *Contingent property*: a property that is not necessary – whether sufficient or not – to its object (e.g., both being pale, and knowing that one is Socrates are contingent to Socrates. He would have survived losing either. The first is neither necessary nor sufficient for Socrates; whereas the second, while also not necessary, is nonetheless sufficient for Socrates). *Essential property*: a necessary property, whether sufficient or not – to its object. The two sub-categories of essential properties are thus, on the one hand, the category of necessary and sufficient properties (e.g., arguably, originating from the gametes from which Socrates actually originated, for Socrates; as well as being identical to Socrates for Socrates; or, less trivially, the property of being the quotient of 10:2, for 5), and, on the other hand, the category of properties that although necessary are not sufficient to their objects (e.g., being mammal for Socrates). The first is in fact the category of *properties that serve as the criteria of identity* of their objects; the second is the category of *secondary substances*. Both, therefore, are types of essential properties. The sub-category of properties that are the criteria of identity of their objects, i.e., of necessary and sufficient properties, itself includes a sub-category, which is that of *sortal properties*, more traditionally referred to as *essences* (e.g., arguably, originating from the gametes from which Socrates actually originated, for Socrates). However, it seems that for the overwhelming majority of cases these two types of properties, namely, criteria of identity and sortal properties, overlap. Specifically, it seems that the only cases of necessary and sufficient properties that are not sortal properties are mathematical properties, and the trivial property 'being identical to x'. The following table summarises this classification:

Logical relations		Examples (for Socrates)	Types of properties		
Necessary	Sufficient	Originating from the gamete from which Socrates actually originated	*Sortal properties/ Essences*	*Criteria of identity properties*	*Essential properties*
		Being identical to Socrates			
	Not sufficient	Being mammal		*Secondary substances*	
Not necessary	Sufficient	Knowing that one is Socrates			*Contingent properties/*
	Not sufficient	Being pale			*Accidental properties*

Table 3.1

3.4 Criteria of Identity vs. Supervenience: A Note About Hierarchical Relations

Finally, and to a large extent independently of the above discussions, I would like to analyse the link between material criteria of identity and physicalism, and more generally, the link between criteria of identity and supervenience. In the course of this analysis, two more important hierarchical relations will emerge. This will culminate in a proposed structure of the links between these four hierarchical relations.

Physicalism is the view according to which the manifest, as well as any other aspect of reality, (at least[37]) *supervenes* upon the physical.[38] In its broad sense, supervenience is intuitively defined as following: for a set of properties *A* to supervene upon a set of properties *B* is for a *B*-duplicate of something to also be an *A*-duplicate of it (i.e., if two things are *B*-indiscernible they are thereby also *A*-indiscernible), but not necessarily vice versa.[39] Or, in its contra-posed version, *A* supervenes upon *B* iff there can be no *A*-difference without there being a *B*-difference. (We shall fine-grain this broad definition shortly.)

It is tempting to think that material criteria of identity – i.e., a universe that primarily contains material objects (such as lumps of clay and H_2O) that merely happen

to have manifest properties (such as being a statue or being watery) goes hand in hand with physicalism. This linkage is ill-fated; or so at least I shall argue. More generally, it will be argued that supervenience and criteria of identity are mutually independent.

In view of the varieties of supervenience relations in the literature, we should make some preliminary remarks before getting to the argument.

(A terminological remark: as not everything that is physical is plainly material (e.g., forces, electric fields etc.), it is more customary now to talk about the *physical* rather than about the mere *material*. We shall thus follow this trend, and talk about physical criteria of identity instead of mere material criteria of identity.)

The above broad definition of supervenience has undergone considerable fine-graining through the introduction of various distinctions, most of which are concerned with the *domain of application* of the supervenience relation. Two such major distinctions (due to Kim[40]) are between *individual* (aka *local*) supervenience and *global* supervenience, and between 'weak' supervenience and 'strong' supervenience. In its individual version, the relation applies to individual objects or events. As Davidson puts it (with relation to the mental and the physical):

> . . . supervenience might be taken to mean that there cannot be two events alike in all physical respects but differing in some mental respects, or that an object cannot alter in some mental respects without altering in some physical respects.[41]

In its global counterpart, the relation applies to the whole world. As Lewis puts it:

> If two possible worlds were exactly isomorphic in their pattern of coinstantiation of fundamental properties and relations, they would thereby be exactly alike *simpliciter*.[42]

Both versions – local and global – have each a weak and a strong version (as well as most other types of supervenience[43]). With respect to individual supervenience, weak supervenience applies merely to the actual world, whereas global supervenience applies to all possible worlds. Specifically, A individually *weakly* supervenes on B iff "for any possible world w and any individuals x and y in w, if x and y are B-indiscernible in w, then they are A-indiscernible in w"; By contrast, A individually *strongly* supervenes on B iff "for any possible worlds w_1 and w_2 and any individuals x in w_1 and y in w_2, if x in w_1 is B-indiscernible from y in w_2, then x in w_1 is A-indiscernible from y in w_2".[44] This strong supervenience entails weak supervenience but not vice versa.

Our discussion below is concerned with the relation between supervenience and criteria of identity. Similar distinctions – weak vs. strong and local vs. global – may apply

to criteria of identity as well. I.e., criteria of identity may apply merely to the actual world or to all counterfactual worlds; in addition, we can talk about the criteria for being the same object (or stuff), as well as about the criteria for being the same world, e.g., whether these are manifest or material. The main claim to be advanced here is that supervenience and criteria of identity are independent from one another. This claim is not restricted to a certain domain; thus, the claim is that supervenience and criteria of identity are mutually independent, whether we consider individuals or the entire worlds, and whether we consider merely the actual world or all possible worlds.

(However, note that it is by no means claimed that supervenience is *true*, and in particular, that physicalism is true; nor is it claimed that physical criteria of identity are *true*. The claim is that physicalism and physical criteria of identity, whether true or not, are independent of one another).

We can now proceed to the argument.

To be sure, both views – physical criteria of identity and physicalism – express some sort of *hierarchy*, in which the physical has some priority over other aspects, such as the manifest, the mental, the functional etc. Yet it is suggested that each of these views expresses a *different* type of hierarchy, and the two hierarchies are quite independent of one another.

Before we get to analyse and compare the two types of hierarchies, we can consider a simple intuitive example that demonstrates the independence of the two hierarchies. Think of organisms. As far as material constitution is concerned, they – in fact, we – are all in a constant flux; molecules constantly join and leave our bodies, and given a sufficient period of time, we (i.e., organisms) even undergo a complete material change. Hence, organisms seem to merely happen to have their physical constitution. In other words, if organisms exist (indeed, some believe they don't exist, as such)[45], then, since the physical is something that such objects merely happen to have, their criteria of identity are not physical. Yet at the same time, many believe that a physical duplicate of an organism is also a duplicate of the entire organism, i.e., of all aspects of the organisms, but not necessarily vice versa. In other words, many believe that organisms supervene upon their physical constitution. Intuitively, these two views are in no conflict. It seems perfectly consistent to hold that while organisms supervene upon their physical constitution, they each merely happen to have their *particular* constitution, and might just as well have had a different constitution. It follows that the supervenience of everything

(including the living world) upon the physical, is compatible with non-physical criteria of identity.

And now for some more rigorous analysis.

The supervenience of B on A amounts to the idea that an A-duplicate is also a B-duplicate but not vice versa. This a-symmetrical, i.e., hierarchical, relation is in fact a combination of two, simpler, hierarchical relations. The first is a formal type of relation, which has to do with the mathematical correlation between things of type A and things of type B. Namely, for B to supervene upon A is for B to stand in a one-many correlation to A; i.e., every A is correlated to only one B, whereas some B's may be correlated to more than one A. For example, in a community where all members of each family drive one car and every car is driven by members of but one family, there is a one-many correlation between cars and drivers. This correlation entails that given a certain A – driver – we can determine a corresponding B – car – but not vice versa; we cannot determine a certain A – driver – on the basis of a particular B – car. (Other formal relations are one-one; many-one; many-many; and no correlation. Supervenience requires *at least* one-many correlation, i.e., either one-many, or one-one correlation.) Supervenience relations involve such a formal hierarchy. E.g., in the case of physicalism, the manifest (as well as other aspects) allegedly stands in a one-many correlation to the physical. Thus, we can determine the manifest on the basis of the physical, but not vice versa (as the same manifest may be correlated with different physical states).

However, supervenience relations involve more than just a formal one-many correlation. Clearly, despite the one-many correlation between cars and people in the above community, cars do not *supervene* upon people; a human-duplicate of the community is by no means thereby a car-duplicate of that same community. What's missing is a further *spatial* hierarchy. When B supervenes upon A, B is not only correlated to A but B and A also *coincide*, in a way that B is *instantiated by*, or *realised in*, A; i.e., it is required that A somehow *underlies* B. (Admittedly this may not be very satisfactory a definition of this relation, but hopefully it will do for our present purposes).

It thus follows that supervenience is a case in which two conditions are met. Firstly, B carries a one-many correlation to A; and secondly, A underlies B. In other words, supervenience involves *both* a formal hierarchy and a spatial hierarchy between A and B. Only the combination of the two hierarchies brings about the situation that an A-duplicate is also a B-duplicate. In particular, to say that the manifest (and other aspects)

supervenes upon the physical is to say that, (a) the manifest (as well as other aspects) is *determined by* the physical but not vice versa; and (b) physical objects *underlie/instantiate/realise* the manifest (and other aspects) but not vice versa. It is the *combination* of the two hierarchies that brings it about that a physical duplicate is also a manifest duplicate, (as well as any other duplicate), i.e., that the manifest *supervenes* upon the physical.

The two hierarchies that make the supervenience relation are independent of each other: it was already demonstrated that we could have a formal hierarchy without a spatial one, as in the above cars-drivers case; it is also possible, however, to have a spatial hierarchy without a formal one, as the following example illustrates. Consider a possible world in which *all* the following stuffs exist: watery H_2O, watery XYZ, non-watery H_2O and non-watery XYZ (note that such a world is clearly logically as well as metaphysically possible, even if its *physical* possibility may be questioned[46]). In such a world, the correlation between the physical and the manifest is many-many, and so it is not possible to determine the manifest on the basis of the physical or vice versa. Nevertheless, it is possible that the physical in this world still *underlies* the manifest, i.e., that the manifest is realised, or instantiated, in the physical. This would be a case of spatial hierarchy without formal hierarchy. Overall then, we may conclude the mutual independence of the formal and the spatial hierarchies.

Furthermore, these formal and spatial hierarchies are distinct from yet another type of hierarchy, which is the one involved in criteria of identity. This is a *metaphysical* hierarchy. In the context of criteria of identity, for A to be hierarchically superior to B means that whereas A is what objects *are*, B is what (some) objects *are like*. (If our above analysis of criteria of identity in terms of sortal properties is sound, we may simply say that A is metaphysically superior to B if A is a sortal property, whereas B is not). Clearly, this metaphysical form of hierarchy is different from both the formal hierarchy – which is about correlation – as well as from the spatial hierarchy – which has to do with instantiation.

Consequently, metaphysical hierarchy is also distinct from the *combination* of the two other hierarchies, namely, metaphysical hierarchy is distinct from supervenience hierarchy.

Metaphysical hierarchy and supervenience are not only distinct from each other, but are also independent of each other. As a support for this claim consider the above

organism example: there is a view according to which it is possible for organisms to supervene upon their physical constitution and yet at the same time to have this physical constitution contingently. Whether this view is correct or not, the important point is that it is coherent. In other words, it is coherent to hold that the two hierarchies are in opposite order. However, to prove that there is no dependence, it should be also possible for the two hierarchies to be in the same order. This seems to be no less coherent. The manifest (and other aspects of the world) may supervene upon the physical while at the same time the universe may be such that it primarily contains physical objects that merely happen to have their manifest properties (as the material view has it). (On this latter view, what we call organisms are in fact collection of molecules that happen to have organismic properties, which they might have equally lacked).

To sum up, hierarchical relations between things of type A and things of type B are a-symmetrical relations between the two types. We have discussed four such sorts of hierarchical relations: (a) if things of type B carry a one-many correlation to things of type A, then A determines B but not vice versa. This is a *formal* hierarchy, in which A is superior to B; (b) if things of type A and things of type B coincide, while the former underlies/instantiates/realises the latter, then there is a *spatial* hierarchy between A and B, in which A is superior. These two hierarchies are distinct and independent; (c) a combination of the formal hierarchy and the spatial hierarchy results in *supervenience* hierarchy: B supervenes upon A iff an A-duplicate is also a B-duplicate, but not necessarily *vice versa* (in this case A is superior to B); and (d) if the universe is such that it primarily contains things of type A, some of which merely happen to be B's, then there is a metaphysical hierarchy between A and B (in which A is superior to B). This is a *metaphysical* hierarchy.[47] It has been argued that the metaphysical hierarchy and the supervenience hierarchy are distinct and independent. We have argued as follows. Metaphysical hierarchy is clearly distinct from formal hierarchy. It is also distinct from the so-called spatial hierarchy. It thus also distinct from the combination of these two hierarchies, namely, from supervenience hierarchy. Furthermore, metaphysical hierarchy is not only distinct from supervenience but it is also independent of it. As a support for this claim, we have considered two views about organisms: (a) a common view that takes organisms to supervene upon their physical constitution, and at the same time to have the specific physical constitution that they have contingently; and (b) the converse view – namely, that equally takes organisms to supervene upon their physical constitution, but,

contrary to the former view, holds them to have their manifest properties contingently. The fact that the two views are coherent provides a strong support for the independence of the two types of hierarchies. Thus, generally, we conclude that the fact that A supervenes upon B has no entailment as to criteria of identity, i.e., whether the universe primarily contains A's that merely happen to be B's, or whether the universe primarily contains B's that merely happen to be A's. Both options are compatible with each supervenience hierarchy.

Endnotes

---

[1] To remind, it is assumed in this dissertation that identity requires criteria of identity. See Endnote 23 in Chapter 1.

[2] See Section 2.3.

[3] E.g., Quine (1960, p. 199) argues as follows:

> Mathematicians may conceivably be said to be necessarily rational and not necessarily two-legged; and cyclists necessarily two-legged and not necessarily rational. But what of an individual who counts among his eccentricities both mathematics and cycling? Is this concrete individual necessarily rational and contingently two-legged or vice versa? Just insofar as we are talking *referentially* of the object, with no special bias towards a background grouping of mathematicians as against cyclists or vice versa, there is no semblance of sense in rating some of his attributes as necessary and others as contingent. Some of his attributes count as important and others as unimportant, yes; some as enduring and others as fleeting; but none as necessary or contingent.

In other words, if Paul is the mathematician who is also a cyclist, then Paul *qua* mathematician is necessarily rational, whereas *qua* cyclist he is only contingently rational. (While regardless of description, i.e., simply *qua* Paul, he is neither necessarily nor contingently rational.) Indeed this argument for the description-relativity applies to necessary properties and to contingent properties, and not explicitly to sortal properties. However, as we shall see shortly, sortal properties are a sub-class of necessary properties, and hence the argument applies *ipso facto* to sortal properties as well.

[4] Locke (1689, III, iii, 15)

[5] Aristotle, *Categories*

[6] The relation between sortal properties and criteria of identity will be later refined. In particular, it will be shown that although all sortal properties serve as criteria of identity, some criteria of identity are not sortal properties (Section 3.3.5).

[7] Grandy (2006)

[8] P. F. Strawson (1959, 169 ff.; 1974, Chapters 4, 5)

[9] Quine (1960)

[10] Wiggins (1967, Part 2 and Appendix; 1980, Chapters 2, 3)

[11] Geach (1962)

[12] Geach (1962, p. 64) distinguishes his *substantival* terms from sortals on just this point:

we can speak of the same gold as being first a statue and then a great number of coins, but 'How many golds?' does not make sense. Thus 'gold' is a substantival term, though we cannot use it for counting.

[13] Indeed, the Aristotelian picture is rather specific, and posits, for example, a limited number of categories that apply to all objects.

[14] Putnam (1977, p. 103)

[15] Molière (1673, act 3)

[16] Putnam (1977, p.104)

[17] I am indebted to José Zalabardo for helping me to formulate my ideas on this matter.

[18] Unless we have some *trope* conception of properties. But that will make the shape an individual essence. See below.

[19] Chisholm (1967)

[20] It is important to note that indiscernibility here is not merely in the epistemic sense; i.e., it is not confined to what properties are available to *us*, but it applies to *all* properties. Thus, it is different from the qualitative indiscernibility discussed above in relation to possibilities such as that "Nixon might have been an automaton" (see Section 2.3). In the latter case the subjects resemble each other for all *we* can tell, yet they do differ in some properties that are hidden from us, namely, one is a person while the other is an automaton. No such difference applies in the present context.

[21] Incidentally, in one place it seems that Kripke (1980, pp. 15-20) in fact prefers such an option of qualitatively-identical yet numerically distinct worlds (while explicitly not committing to their existence) to its alternative on which qualitatively-identical worlds are *ipso facto* held identical. This is because the latter view entails that a situation in which, for example, dice A shows 6 and dice B shows 5 is identical to a situation in which dice A shows 5 and dice B shows 6; and, consequently, that the probabilities related to two dice amount to merely 21 rather than to 36, which seems a highly undesirable result (see Endnote 97 in Chapter 1).

[22] Chisholm (1967, p. 6)

[23] Chisholm (1967, p. 5)

[24] Echoing Quine's (1953) famous replacement of "Pegasus" with "the thing that Pegasizes".

[25] Other adherents of such *primitive thisness* are Kant and Pierce. Famous anti-haecceitists include Leibniz, Russell, and Ayer. For contemporary applications of haecceitism to modal contexts see Kaplan (1975); Adams (1979); and, Lewis (1986, pp. 220-48).

[26] Kripke (1972/1980, pp. 110 ff.)

[27] The view dates back to Plato (see *Phaedo*, Books V-VII of the *Republic*, and *Parmenides*). Notable contemporary defenders include Russell (1912), Donagan (1963) and Armstrong (1989)

[28] A notable advocate is Williams D. C. (1953). But according to D.W. Mertz (1996, ch. IV), versions of this view can be found in Plato, Aristotle, Boëthius, Avicenna, Averroës, Thomas, Scotus, Buridan, Suárez, Leibniz, Husserl, Russell (early, 1911), Stout, Cook Wilson, and P. F. Strawson.

[29] Loux (1998) notes that versions of this view can be found already in the medieval thinkers Abelard and Ockham: McKeon (1929, pp. 208-58). The recent name that is most strongly linked with class nominalism is of course Quine (1953).

[30] E.g., Cocchiarella (1986, Ch. 3)

[31] Armstrong (1997; 2004)

[32] The instantiation relation famously opens the door to Bradley's regress problem, namely, if an object x is related to a property F by the relation of instantiation R, this relation requires another two relations: one that links the instantiation relation R to x; the other that links the instantiation relation R to F. This in turn requires four more relations to link these two relations to their relata. And so forth, *ad infinitum*. As Swoyer (2000) notes, the common solution to the problem is to view instantiation as a relation that links objects and properties directly, with no need for further intermediate relations. Thus, Strawson (1959) talks about a "non-relational tie"; Frege suggested a model of a function-argument link, i.e., a property is like a function that takes objects as arguments; Wittgenstein suggested something like links in a chain; and Broad (1933, p.85) talked about "metaphysical glue". All these modes demonstrate relations that bind the relata directly, with no need for further relations to attach the binding agent to the relata.

[33] Having reached this stage in our study, we are aware of the fact that everyday metaphysical assumptions are by no means indisputable; in principle, there is nothing to rule out other, less common, views. E.g., although (as our first example below assumes) it is commonly believed that being pale is neither necessary nor sufficient for Socrates, still, there is nothing to rule out a metaphysical view that takes paleness to be necessary, or sufficient, or, even both, to Socrates.

[34] Although this is very widely accepted, there is some opposition nonetheless. Kit Fine (1994) draws a distinction between necessary properties and essential properties. Specifically, he argues that although all essential properties are necessary to their object, the converse does not hold; some necessary properties are not essential. For example, being distinct from the Eiffel tower is necessary to Socrates but, according to Fine, it is not essential to him. This claim is based on

Fine's notion of essential properties as properties that are part of the object's *essence*, or nature; being distinct from the Eiffel tower is not part of Socrates' nature, and hence not an essential property. I believe this to be a mere terminological disagreement, about the meaning of the term "essential" and thus not a substantial disagreement. That is to say, the debate merely reveals that different speakers may use the term "essential" in different ways. We may say then that the above characterisation does not apply to Fine's particular notion of "essential".

[35] Note that numbers are more like stuffs than they are like individual objects: just as there is only one stuff which is watery H_2O in each world (albeit many *samples* of that stuff), similarly there is also only one number 5 in a world (albeit many groups of five objects). By contrast, there are many individual people, dogs, trees, etc.

[36] Locke (1689, III iii 15)

[37] A stronger relation would be, of course, that the non-physical not only supervenes, but is also identical to the physical.

[38] Although the idea appears to have been around for quite a while (as McLaughlin and Bennett (2005) note), Davidson (1970) is normally credited for the contemporary understanding of supervenience, and for stirring the discussion; particularly within the philosophy of mind.

[39] Thus the relation has two sub-relations: one in which there can also be no *A*-difference without *B*-difference; the other in which there *can* be an *A*-difference without there being a *B*-difference. The first case is one of mutual supervenience (e.g., the surface area of a sphere and its volume mutually supervene upon each other (McLaughlin and Bennett (2005) credit the example to Lombard (1986)). However, the latter case is interesting as it allows for the so-called *multiple realisability*: the same *B* may be realised in more than one *A*, but one *A* can only realise one *B*. Putnam (1967) famously used multiple realisability to argue against the identity theory of mind: pains can be felt by humans as well as by other animals (and thus, possibly, by Martians as well); so if pains are realised in physical states, then it follows that they might be (if not actually are) realised in physical states quite different from one another. And generally, the same mental state is realisable in more than one physical state. But not vice versa: once the physical state has been determined, so is the mental state that it realises. So, although pains (and other mental-state types) may supervene upon physical-states, they may not be (type-) identical to them. (This, however, does not rule out token-identity; i.e., that every pain *token* is identical to a certain physical-state *token*.)

Note also that whereas the first sub-relation, namely mutual supervenience, is a symmetric one, the second is not. Consequently, whereas two properties may supervene upon one another (as in the case of the surface and the volume of a sphere), no two properties may be

multiply realised in one another; if B is *multiply* realised in A it follows that A is *not* multiply realised in B (in fact, if B is realised in A, A cannot be realised in B at all, let alone multiply realised).

[40] Kim (1984)

[41] Davidson (1970, p. 214)

[42] Lewis (1994)

[43] In addition to individual and global supervenience, McLaughlin and Bennett (2005) mention also *regional* (due to Horgan 1982), *similarity-based* (Kim 1987) and *multiple domain* (Kim 1988) types of supervenience – all of which have weak and strong versions. (For a detailed classification of types of supervenience and formal definitions, see McLaughlin and Bennett (2005), especially the supplement: "Appendix: List of definitions").

[44] Kim (1987). With respect to global supervenience, the distinction between weak and strong supervenience (drawn by Stalnaker (1996), McLaughlin (1996; 1997), and Sider (1999)) is defined by McLaughlin and Bennett (2005) as follows: A weakly globally supervenes on B iff "for any worlds w_1 and w_2, if there is a B-preserving isomorphism between w_1 and w_2, then there is an A-preserving isomorphism between them"; whereas A strongly globally supervenes on B iff "for any worlds w_1 and w_2, every B-preserving isomorphism between w_1 and w_2 is an A-preserving isomorphism between them."

[45] Notably, Unger (1980)

[46] See also Endnote 4 in Chapter 1.

[47] In other words, A is a sortal property whereas B is not.

4 INTENSIONS: SOME EPISTEMIC ASPECTS

Overview

This chapter looks into some epistemic issues, mainly our state of knowledge or ignorance, and the way these are related to intensions.

Our epistemic access to the world is limited. There is always a possible gap between how things are and how we take them to be. When the two coincide, we are right; when they don't, we are wrong. This possible gap applies to each of the three factors in our intensions formula (assuming, of course, that there is a fact of the matter in each case). Specifically, on the metaphysical level, there are the real criteria of identity, and there are what people take to be the criteria of identity; on the linguistic level, there is the real semantic function of a term, and there is what speakers take the semantic function of the term to be; and with respect to actuality, there is the real actual state of affairs, and there is what people take the actual state of affairs to be.

In the first part of this chapter we discuss issues to do with the relation between the way things are and the ways we take them to be. One common reason for changing the way we take things to be is following (scientific) discoveries. Thus discoveries loom large in this discussion. We begin by considering rigidified terms (i.e., terms with constant intensions), and their crucial role in securing reference in states of ignorance, and in light of possible future discoveries. We then discuss different ways in which we may lack knowledge of sortal properties. This analysis will teach us that it is not in the ken of science to determine the sortal properties of things. Lastly, we analyse a complication related to the intension of natural-kind terms that does not apply to singular terms. The problem is that whereas the name-givers of Aristotle cannot be mistaken in initially naming him "Aristotle", the name-givers of "fish" – even under the assumption that, like "Aristotle", "fish" is a rigidified designator that is introduced by stipulation – were apparently partly mistaken in initially naming all the swimming organisms "fish". Specifically, the discovery that whales were not fish, but mammals, seems to indicate such a mistake.

In the second part of this chapter, we discuss the relation between different ways we may take things to be. The discussion focuses on reduction, which is a particular sort of hierarchical relation between one way we take the world to be, and another. We analyse different ways in which reduction may be described, and how these bear upon the intensions of the terms involved. We shall uncover two new semantic functions – to be termed "role" and "occupant" – which will prove very useful in explaining some reduction-related puzzles.

4.1 The Relation between the Way Things Are and the Ways They are Taken to Be

4.1.1 Rigidified vs. descriptive as "absorbent" content vs. stipulated content

Let us recall first the relations between intensions, criteria of identity and sortal properties.[1]

Our prime semantic distinction is between descriptive and rigidified terms. The intension of a descriptive term assigns to each possible world that which fits the term's descriptive content in that world. Thus, we may describe the "manual" attached to a descriptive term "T" which is associated with a descriptive content D as follows: "pick out, with respect to every possible world, that which D's (i.e., is D) in that world". This is independent of any criteria of identity. By contrast, determining the intension of a rigidified term does require criteria of identity. A rigidified term is designed to designate, with respect to every counterfactual world, the *same* referent that it designates in the actual world. Being "the same referent" is dependent on criteria of identity.

Criteria of identity are spelled-out in terms of sortal properties: X is identical to Y iff X and Y share the same sortal property (or individual sortal property, in case X and Y are individuals. See Section 3.2). Thus, the "manual" attached to a rigidified term can be said to be the following: "go to the actual world, find the referent of the term, find its (individual) sortal property, then go to other counterfactual worlds, and pick out that which has the same (individual) sortal property." Thus, if this (individual) sortal property is S, we may say that the manual attached to the rigidified terms is simply, "pick out, with respect to every counterfactual world, that which S's (i.e., is S) in that world".

But now, once the means by which a rigidified term designates across counterfactual worlds is taken into account, namely, once the (individual) sortal property

is located, it turns out that there is a striking resemblance between the semantic functions of descriptive terms and that of rigidified terms: a descriptive term designates, with respect to each counterfactual world that which D's in that world; whereas a rigidified term designates, with respect to each counterfactual world, that which S's in that world. In other words, both designate, with respect each counterfactual world, that which meets a certain *condition*, in that world; namely, D in the case of descriptive terms, and S in the case of rigidified terms. This is another way of saying that both designate, with respect to every possible world, that which fits some descriptive content. It thus follows that rigidified terms eventually have *some* descriptive content as well, which is "has the (individual) sortal property of the actual referent". (This last claim may, of course, be very unpopular among some New Theorists. However, at present I can see no way around it, and am therefore compelled to hold it a necessary implication of the New Theory.)

There are, nevertheless, crucial differences between the descriptive content of a descriptive term and that of a rigidified term. These differences make them categorically distinct. Let me specify those. (Indeed all three differences below are aspects of one central difference.)

Firstly, the descriptive content of a rigidified term is mediated by the actual referent of the term. No such mediation is involved in the case of descriptive terms. The content of a descriptive term is simply given directly. This has the important implication that whereas the content of a descriptive term is determined by *us*, i.e., by the semantic conventions of our language, the descriptive content of a rigidified term is determined by *nature*, independently of us. In other words, whereas the descriptive content D of a descriptive term is *stipulated* by the language community, the descriptive content S of a rigidified term, by contrast, "absorbs" its content from (the sortal property of) the actual referent. E.g., the descriptive content of "water" is "absorbed" from the sortal property of the actual referent of "water"; thus, a sortal property 'being H_2O' will generate a different descriptive content than a sortal property 'being watery'. As we shall see shortly, this fact has some crucial epistemic implications.

Secondly, there are different possibilities with respect to the actual referent of a term. E.g., it is a possibility of the actual world that Aristotle was indeed a person, but it is also a possibility of the actual world that Aristotle was not a person but a robot, i.e., there is a possible *actual* world in which Aristotle was a robot (although, if the "real"

actual world is one in which Aristotle *is* a person, then there is no possible *counterfactual* world in which he is a robot. see Sections 1.1.3 and 2.3). Thus, there are two possible actual referents of the rigidified term "Aristotle". These two possible actual referents, the person and the robot, have different sortal properties. So there are different possibilities as to the sortal property of the actual referent of a given rigidified term. And this further entails that there are different possibilities as to the descriptive content attached to such a rigidified term. E.g., if Aristotle was actually a person, then – assuming the necessity of origin thesis – the descriptive content of a rigidified "Aristotle" in that case is "he who originated from the same gamete as the actual referent". This descriptive content will obviously be different in case Aristotle was actually a robot; it will involve the sortal property of robots (whatever that might be). By contrast, no such variations apply to descriptive terms; e.g., a descriptive "the last great philosopher of antiquity" designates, with respect to each counterfactual world, that which fits the description in that world, regardless of whether the actual great philosopher of antiquity was a person, a robot, or anything else. Thus, relative to the possible *actual* worlds, the descriptive content of a descriptive term is *constant* across the intensions table, whereas the descriptive content of a rigidified term *varies* as a function of what the actual is like.

Lastly, when considering the actual world, the two types of terms, descriptive and rigidified, have different implications regarding the relations between the sense and the reference of the term. In the case of a descriptive term, the sense determines the reference, e.g., the referent of "the last great philosopher of antiquity" is determined by the sense of that term. By contrast, in the case of a rigidified term, the sense to be attached to "Aristotle" (if indeed rigidified) is determined by the sortal property of the actual referent, i.e., by the sortal property of Aristotle. In other words, in the case of rigidified terms, it is the (actual) reference that determines sense and not vice versa.

4.1.2 Referring terms in a state of ignorance

One may wonder, why have rigidified terms at all? Why not simply use only descriptive terms? After all, all we need to do is to use some descriptive predicate that will uniquely refer to our object; and these seem to be in abundance.

An intuitive reply may be that rigidified terms are there to guarantee designation of the *same* individual in all counterfactual worlds.

Yet, as we have witnessed, a descriptive term may be rigid as well, in which case it designates the same individual in all possible worlds (recall "stuff with atomic number 79"). This happens whenever the property involved in the description is an (individual) sortal property. Thus, apparently, we can replace a rigidified term with a descriptive one, simply by using a description which is linked to the referent's (individual) sortal property. For example, instead of using the (rigidified) term "gold" we may simply use "stuff with atomic number 79", and get the same results. And in general, if the actual referent of a certain term "T" has the (individual) sortal property S, then instead of using (a rigidified) "T", we may just use "that which S's". Thus, apparently, we could in principle discard all rigidified term, by replacing them with their corresponding descriptive term.

This replacement would be successful in a state of omniscience. The fact of the matter, however, is that we are far from such a state. Specifically, in many cases, we simply do not *know* what the sortal property of a referent is, and thus we do not *know* what the appropriate description to replace it would be. Given this human epistemological limitation, rigidified terms come incredibly handy. This is because they are latched onto their referent regardless of our state of knowledge about that referent. We can simply name objects and stuffs, without knowing their sortal properties. Thanks to this facility, a directly-referring "Aristotle", for instance, designates Aristotle in every counterfactual world, whatever Aristotle actually turns out to be in the actual world, e.g., if it turns out that he was not the last great philosopher of antiquity (say, if his wife Pythias was), and even if he was discovered not to be actually human at all, but, say, a robot or even a non-material hologram. No discovery about Aristotle would deprive him of his name; not in the actual world, and not in any counterfactual world. Whatever Aristotle actually is, the (rigidified) term "Aristotle" would designate that individual in every counterfactual world.

The situation is entirely different when it comes to descriptive terms. Such terms refer to whatever fits the description associated with the term. Hence, if we were to discover that a certain object that we believed to fit the description in effect doesn't, we would deprive it of that descriptive term. Specifically, in the case of sortal terms. For instance, if we were to discover that gold, contrary to our belief, does not actually have atomic number 79, we will cease to apply the descriptive term "stuff with atomic number 79" to gold. In fact, part of the discovery would be that the description *never* actually designated gold (despite the fact that we, wrongly, used it in this way). By contrast, such a discovery would have no effect on the designation of the (rigidified) term "gold".

Overall then, no matter how ignorant we are about a certain object or stuff, by using a rigidified term to designate it, we are guaranteed that that term would refer to it. No future discovery about our referent could ever change this designation.

(A qualification needs to be made here. To remind, there are two types of rigidified terms: directly-referring terms and reference fixing terms. Whereas both pick out their *counterfactual* referents in a similar way, namely, both designate, with respect to every counterfactual world, that which they designate in the actual world, nonetheless the two pick out their *actual* referent differently: a directly-referring term designates its actual referent independently of descriptive content, while the actual referent of a reference-fixing term is picked-out by fitting the term's descriptive content. Hence, what was said above strictly applies to directly-referring terms, but should be slightly restricted in the case of reference-fixing terms: e.g., a reference-fixing "gold" whose descriptive content is "the (actual) golden stuff" indeed designates gold in all counterfactual worlds no matter what we discover about it, *save for the (unlikely) discovery that it was not golden.*)

The same equally applies to natural-kinds that are not stuffs like gold; e.g., to tigers. According to the New Theory of reference, natural-kind terms are rigidified terms. We have proposed above (Section 2.2.4) to spell this out as follows: a natural-kind term designates, with respect to every counterfactual world, objects *of the same kind as* the objects that it designates in the actual world. (This definition allows a natural-kind term to designate different *sets* of individuals in different counterfactual worlds.) E.g., "tiger" designates, with respect to every counterfactual world, the individuals that are of the same (natural) kind as the ones referred to by "tiger" in the actual world. However, we may be *uncertain* about what being a tiger consists in, i.e., what is its sortal property. (We may have a *theory* about what it is, yet, clearly, any theory is fallible, i.e., we may be proved wrong). Thus, by using "tiger" as a rigidified term (as Kripke believes it to be the case[2]), we guarantee that whatever being a tiger consists in, i.e., whatever the sortal property of tigers is, the term "tiger" will designate members of *this* kind, in every counterfactual world.

In sum, the content of a descriptive term is stipulated by the community of speakers. By contrast, the content of a rigidified term is "absorbed" from the (individual) sortal property of the actual referent to which it is applied (regardless of what we believe this (individual) sortal property to be). In other words, the content of a descriptive term is *constant*, whereas that of a rigidified term is *variable*. This provides rigidified terms with

the capacity to secure reference to an intended object of reference regardless of our state of ignorance (whether the object of reference is an individual, some stuff, or a natural-kind). The same does not apply to descriptive terms.

A final note: talking about a state of ignorance makes sense only under realist assumptions, i.e., that there is a fact of the matter regarding the sortal property of objects. If there is no fact of the matter, and these sortal properties are merely stipulated, then, indeed, it seems that rigidified terms do amount simply to descriptive terms whose descriptive contents state the stipulated sortal properties. E.g., it makes sense to hold that there is no fact of the matter regarding the sortal property of cars, i.e., that cars have no mind-independent sortal property; by designing cars to serve a certain function, we thereby stipulate what it is to be a car, i.e., we stipulate its sortal property. Having stipulated it, there is little sense to engage in exploring the sortal property of cars. Correspondingly "car" *is* an abbreviated description – simply a description of its function – whose descriptive content *is* constant and not variable. This descriptive content by no means depends on external facts, but rather solely on our stipulation. If, as anti-realism has it, there are no facts of the matter about *any* sortal properties, then all referring terms in our language are just like "car" – stipulated and constant.

4.1.3   Sortal properties in a state of ignorance

In Section 3.2 we discerned the following structure. One of the factors that determine the intension of a term is the criteria of identity of the actual referent of the term. Criteria of identity are (mostly[3]) spelled out in terms of sortal properties: x is identical to y iff x shares the (individual) sortal property of y. Thus, the intension of a term is dependent upon the sortal property of the term's actual referent. The sortal property itself is determined by two factors: the sortal second-order property (i.e., the second-order property that defines the criteria of identity) and the actual first-order properties of the object in question: with respect to a certain object, the first-order property on which the sortal second-order property homes-in is thereby rendered sortal.

From an epistemological point of view, this structure entails three ways in which we may be incapable of determining the sortal property of an object, and consequently also the intension of a term that actually refers to this object: (a) we may *know* one factor but fail to know the other; or (b) the other way around; or (c) we may fail to know both. In detail, (a) we may know the sortal second-order property, but lack the knowledge of

the first-order property on which that property homes-in. This seems to be the state of affairs before 1750 with regard to water. People (arguably) assumed that the sortal property of water was whatever the material constitution of water was, i.e., they assumed material sortal second-order property. However, at that time they did not know what that material constitution of water in fact was, i.e., they did not know what first-order property fell under that sortal second-order property. Thus they did not know what the sortal (first-order) property of water was. Consequently, they were not in a position to determine the intension of "water"; (b) by contrast, we may be well aware of the first-order properties of the actual referent, yet at the same time fail to know the sortal second-order property, and as a result, again, fail to know which of the first-order properties should count as sortal. This state seems to apply to the case of "David". We know its first-order properties – being constituted by a lump of clay, and having a certain statue-shape (and, in principle, we may know all of its other first-order properties as well), but unless we know the sortal second-order property, i.e., the one that defines its criteria of identity, we cannot determine which of these first-order properties is the sortal one. This, again, will also prevent us from determining the intension of a rigidified "David"; (c) lastly, we can simply be ignorant of both factors.

4.1.3.1 Can scientific practice refute metaphysical antirealism?

Science *discovered* that water is H_2O; that lightning is an electrical discharge; that heat is a mean kinetic molecular energy; and the like. It may seem then, that metaphysical antirealism – i.e., the view that there is no fact of the matter with regard to criteria of identity – is at odds with scientific progress: for science tells us what things are, and thus allegedly proves that there is a fact of the matter in this respect. However, the distinction between first-order and second-order properties may suggest otherwise.

It is true that physical science, for instance, uncovers the physical structure that underlies certain stuffs and phenomena. But this practice merely amounts to uncovering the first-order properties on which a certain type of second-order properties homes-in. For instance, in discovering that water was H_2O, we came to realise that the first-order property that falls under the second-order property 'being the stuff's material constitution' was '(being) H_2O'. Very few would doubt that there is a fact of the matter here. Even if this discovery is to be proved wrong in the future, most will expect it to be merely replaced by another discovery about this first-order property, i.e., a discovery

about what is really the case. However, what may remain in doubt is whether this first-order property is a *sortal* property. This would depend, of course, on whether the second-order property 'being the stuff's material constitution' is sortal or not, i.e., whether it is the one linked with criteria of identity. And the answer to *this* question is simply outside the scope of physical science. And indeed of any *special science.*

To conclude, applying a second-order property, like 'being an object's material constitution' or 'being an object's manifest property (or properties)', to an object will home-in on some first-order property, whether we know this first-order property or not. It takes a special strong type of antirealism to deny that there are facts of the matter in this respect. However, endorsing such facts of the matter does not suffice to enable us to determine which of the second-order properties is the sortal one, and, as a consequence, to being able to determine which of the first-order properties is the sortal property. In other words, antirealism about criteria of identity does not conflict, as such, with our scientific practices.

4.1.4   Artefacts in a state of ignorance

Alongside natural-kinds, i.e., natural objects and stuffs, there are artefacts – objects that are not nature's work but are human constructs. Such human constructs include, for example, cars, bridges, money, states, and football teams. As human constructs, it seems there is little question about their sortal properties, since these sortal properties appear to be just part of the very construction of these objects. When we stipulate a note to have a ten pound value, we thereby stipulate what it is to be a ten pound note. Similarly, (as was noted above,) when we design cars to serve a certain function, we thereby stipulate what it is to be a car, i.e., we stipulate its sortal property. It seems then that unlike in the case of natural-kinds, it makes little sense to engage in exploring the sortal properties of artefacts. Surprisingly, though, this does not seem to generally reflect all our practices relating to such human constructs.

Take for example the political scientists' endeavours to define "state", i.e., to find its essence, or its sortal property. A simplified portrayal of their activity may strike one as very similar to the scientific practice of searching for the sortal property of water: political scientists "collect" a set of entities referred to as "state", and then engage in attempts to reveal what these entities consist in, i.e., to reveal their sortal property. Yet

given that political entities like states, unlike water, are human constructs, how can such a practice of exploring their sortal property take place?

I understand the matter as follows. Many of our own constructs are formed as a result of some *intuition*, rather than through some *conscious* definition. Once such an intuitive construct is formed, we often look for the underlying essence, i.e., sortal property, of this intuition of ours. Thus, although the essences that underlie artefacts as such are not determined by nature but by our own stipulations and conventions, the fact that these essences were not *explicitly* stipulated leaves room for their exploration. And in *this* respect, namely, being non-explicit, the sortal properties of natural-kinds and those of artefacts are alike.

4.1.5   Scientific discoveries: whales, fish, cats and robots

A widely discussed example of scientific discovery is the discovery that whales were not fish but rather mammals. Following the discovery, whales were no longer called "fish". How should such episodes be accounted for?

Descriptivism offers the following account.[4] "Fish" designates that which fits its descriptive content. On that view, what happened in the discovery process, is simply that the term "fish" changed its descriptive content. Scientists helped to refine the concept of the layman, roughly, from "a swimming animal" to something like "animal with gills",[5] thus excluding whales from the group of fish. In other words, before the scientific refinement, the term "fish" applied to whales, but not afterwards. There was a *conceptual change.*

Kripke denies this. He insists that

> scientific discoveries of species essence do not constitute a "change of meaning"; the possibility of such discoveries was part of the original enterprise. We need not even assume that the biologist's denial that whales are fish shows his 'concept of fishhood' to be different from that of the layman; he simply corrects the layman.[6]

Kripke's objection is of course based on his proposed semantics of natural-kind terms, i.e., that they are rigidified terms. On this view, (some[7]) natural-kind terms, like proper names, are introduced via an act of baptising a certain referent (by ostension or intending), and then designate that same referent in all counterfactual worlds (i.e., on our understanding as discussed in Section 2.2.4, objects *of the same kind* of the actual referents). In particular, "fish"[8] designates, with respect to every counterfactual world,

just that which "fish" baptised in the actual world, i.e., regardless of any descriptive content. And if there is no descriptive content involved, how can there be a change of content? So descriptivism is wrong. (We shall see below, however, that even if "fish" is indeed rigidified, still descriptivists are onto *something* in their claim for a conceptual change.)

Yet Kripke's proposal faces a serious problem which its descriptive competing proposal does not. The New Theory, that Kripke endorses, takes some natural-kind terms to be directly referring. One aspect that seems to follow from such semantics is that since the initial actual referents of such terms are introduced by stipulation, there can be no mistake in the initial application of the term. E.g., if "Aristotle" was introduced in the actual world by his name-givers – presumably his parents – in a *baptising ceremony*, then it seems his parents could not have possibly been wrong in *naming* him "Aristotle". Indeed other people, not the initial name-givers of Aristotle, may be wrong in applying the name; e.g., (adapting Kripke's example here[9]) noticing from afar, at twilight, a man raking the leaves in Aristotle's garden, this man may be taken for Aristotle while being in fact Aristotle's new gardener. No such mistake can be made in the process of the initial naming of an intended individual. After all, this is a mere stipulation. Now if natural-kind terms are also directly-referring, as the New Theory contends, it seems the same should apply to them as well; i.e., there can be no mistake in the initial application of natural-kind terms in the actual world. However, the above fish-case seems to testify to the contrary. According to the New Theory, people have initially coined the natural-kind term "fish" by referring to a certain collection of referents. This initial collection included whales. Yet (as Kripke himself points-out[10]) others later *discovered* that whales were not fish but rather mammals. So apparently, we have a case in which the initial application of the term "fish", the naming ceremony, was mistaken; the term "fish" was wrongly used to initially include whales in its extension. (We shall see below however, that there is nevertheless *some* sense in which the idea of infallibility of naming applies in such episodes as well).

So it seems that the fish case serves as a counterexample to the view that natural-kind terms are directly-referring terms. And so are numerous similar episodes in the advent of our knowledge. Unless, however, the New Theory can account for this apparent counterexample. Kripke hints at a suggestion but does not elaborate:

> the substance is defined as the kind instantiated by (almost all of) a given sample. The 'almost all' qualification allows that some fools' gold may be present in the sample. If

the original sample has a small number of deviant items, they will be rejected as not really gold.[11]

Here is a proposed elaboration of this.

First, we locate the source of the problem. The complication with natural-kinds springs from the fact that natural-kinds, such as gold and tigers, unlike individual objects, do not inhabit the world in one unit, but are rather scattered in many samples across it. Mass-terms like "gold" refer to a collection of samples; similarly, terms like "tiger" apply to many individuals, spatio-temporally separated from one another. This poses the following problem to the intension of (rigidified) natural-kind terms. If natural-kind terms like "gold" and "tiger" are rigidified (as New Theorists contend), then they are stipulated to designate, with respect to every counterfactual world, whatever has the same sortal property as that of the actual referents, i.e., the actual collection of samples. But, from an epistemological point of view, we simply *assume* that all the samples of stuff that we call "gold" (or "tiger") are indeed of the same kind. The problem is that we cannot *know* it for certain. I.e., we cannot know for certain that all the samples that we take to be of the same kind actually share the same sortal property.

The solution is not too complicated. Given our epistemic imperfection, it seems safer to assume that it is not *all* samples of the group that share a sortal property, but rather *most* of them. Consequently, the semantics of natural-kind terms should be: a natural-kind term is designed to designate, with respect to every counterfactual world, that which has the sortal property of the *majority* of the actual (group of) referents of the term. Hence, the reference of these terms is not fixed by the actual referent – as in the case of proper names – but rather by the majority of the actual (group of) referents. Such a semantic function may result in excluding some, *yet by no means most*, of the initial referents of the term, from the eventual group of referents. Just like whales were excluded from the designation of "fish".

Thus, a New Theorist's description of this scientific episode should run as follows. The term "fish" was used to refer to a group of organisms. In particular, it was used to refer to the group of swimming organisms. Being a natural-kind term, it implied that most of its referents were taken to share some essential property. Given the assumed biological criteria of identity, it was discovered that the essential property shared by *most* of the referents of "fish" was 'has gills'. Despite manifest appearance – swimming in water – not all members of the group shared it; in particular, whales didn't. Thus, following the discovery, whales were excluded from the group of fish; they were no

longer designates by the rigidified term "fish". In other words, it was found that they were wrongly referred to as "fish" all along, as they were not part of that group. They belonged to the minority that share the (salient) manifest property of the group – swimming in water – but not the essential property – having gills.

Based on this analysis, I believe that there is *some* truth in both the idea of conceptual change implied by descriptivism, as well as in the idea of the infallibility of naming implied by the New Theory. Let me explain this.

In Section 4.1.1 we concluded that whereas the content of descriptive terms is stipulated by the language-community, the content of a rigidified term, by contrast, is determined by the actual referent of the term, specifically, by the essence (i.e., the sortal property) of the term. The whale-fish episode involves a change in what *we* took the essential property[12] of fish to be – from swimming in water to having gills. Hence, even if "fish" is a rigidified term, this episode also involved a change in what *we* took the content of the rigidified term "fish" to be, namely, from "that which swims in water" to "that which has gills". In other words, although the content of a rigidified term "fish" is determined by the actual referent, and is thus independent of our discoveries, still, what *we* take this content to be does change in accordance with our discoveries. If we are allowed to talk about what *we* take some *concept* to be as opposed to what the concept really is, then we may say that although the real concept "fish" underwent no change, what *we* took the concept "fish" to be did undergo a change. So in this sense, there was some *subjective* conceptual change.

As regards the infallibility of naming, there is *some* version of the thesis which is applicable to natural-kinds as well. Rather than saying that we cannot be mistaken in naming (i.e., applying a rigidified term to) an entire collection of samples (stuffs or individuals), we should say that we cannot be mistaken in naming the *majority* of a collection of samples. E.g., although we can be wrong in naming all members of a certain collection "fish" (e.g., by calling whales "fish") we cannot be wrong in calling *most* of them "fish". Such a scenario is simply impossible. For instance, were we to find out that most initial referents of "fish" had no gills, but rather some other essential biological property E, the (rigidified) term "fish" would in that case apply to whatever instantiates E, and not to whatever has gills.

Lastly, based on this analysis, I would like to make a small comment on Putnam's well-known robot-cats thought-experiment:

> Suppose … that the cat as we know it is and always was an artefact. Every movement of the cat … is thought out by a man in a control centre on Mars and is then executed by the cat's body as the result of signals.

For which Putnam, being a New Theorist, draws the following conclusion,

> It seems to me that in this last case, once we discover the fake, we should continue to call these robots that we have mistaken for animals and that we have employed as house pets "cats," but not "animals".[13]

> However, before that Putnam says,

> If some cats are animals in every sense of the word, while others are automata, then there is no problem. I think we would all agree that these other were neither animals nor cats but only fake cats.[14]

According to our proposal, this is not quite accurate. As the whale example taught us, Putnam should have qualified here that only if *most* initial referents were animals, then the robots are to be excluded from the extension of "cat". If, however, most initial referents were robots, and only a minority were animals, then a rigidified "cat" would designate the robots, and it is the animals that would have to be excluded from the term's extension.

4.2   Reduction: A Relation between Different Ways things are Taken to Be

So far we have discussed the relation between the way things are and the way we take them to be. Let us move on now to look at relations between different ways we take things to be. One such particular relation is between a reductionist view of (a certain portion of) the world and a non-reductionist view of it.

4.2.1   The metaphysical implications of reduction

A reduction of one type of entities, M, to another type of entities, P, has the following general form: M is nothing but P; what we call "M" is just P that occupies the M-role.[15] David Lewis put this pattern very clearly in his famous model for reductive method, exemplified through the reduction of the mental to the physical:

> Mental state M = the occupant of the M-role (by analysis),
> Physical state P = the occupant of the M-role (by science),
> Therefore M=P[16]

> (Note: Throughout this section, I will use "occupying the M-role"; "having the

property M", "being M" and simply "M" interchangeably).

Some other celebrated examples of reduction are: what we call "lightning" is just an electrical discharge that occupies the lightning-role; what we call "water" is just H_2O which occupies the watery-role, i.e., being potable, drinkable, odourless, tasteless liquid, that flows in rivers etc.; what we call "heat" is just motion of molecules that occupies the heat-role, i.e., feeling hot.

If the reduction of M to P means that what we call "M" is in fact P that merely occupies the M-role, then different reductionist theories may diverge in (at least) two respects; (a) in what they take the roles to be; and (b) in what they take the occupants of these roles to be.

For example, functionalism takes the role to be a functional causal-role, and the occupants to be physical entities. So what we call "pain", for the functionalist, is just a physical state, which occupies the functional pain causal role, i.e., causes winces and groans etc.[17] E.g. Lewis asserts:

> My argument is this: The definitive characteristic of any (sort of) experience as such is its causal role, its syndrome of most typical causes and effects. But we materialists believe that these causal roles which belong by analytic necessity to experiences belong in fact to certain physical states. Since these physical states possess the definitive character of experiences, they must be experiences.[18]

Another form of reductionism is, for instance, mereological essentialism. The mereologist takes occupants to be collections of elementary particles, and a role to be, being arranged in a certain way.[19] Thus, what we call "cat", for the mereologist, is a collection of particles arranged cat-wise. So, again, "cat" stands for "that which occupies the cat-role".

Generally then, reduction involves a change in the way we take things to be: we think that there are objects of kind M, and then, following the reductive process, we come to think that they are really objects of kind P, albeit M-*like* P's.

What are the implications of reduction; of shifting from thinking that there are M's to thinking that what we took to be M's are in fact P's, that happen to occupy an M-role?

From a metaphysical point of view, the most significant aspect of reduction is the following. An object P that occupies an M-role, in principle, might not have occupied the M-role. Similarly, the M-role, in principle, might have been occupied by something other than P. It thus follows that occupying an M-role is *contingent* to an occupant P. Reduction therefore entails that the property of occupying the M-role, or, in short, being an M, is

contingent to its bearer. In other words, following the reduction, we come to *believe* that being an M is not what the object *is*, but rather what it *is like*. And what the object *is*, i.e., sortally, is P, and this P is M-*like*.[20] For instance, the functionalist holds that the occupant of the pain role – the thing that causes winces and groans etc. – is a physical state; this is what it *is*; occupying the pain-role is merely what this physical state is *like*. Similarly, the mereologist holds that occupying the cat-role, i.e., being arranged cat-wise, is not what the occupant of the cat-role *is*, but rather what it *is like*; what it in fact *is*, is a collection of particles.

So we are quite clear now about the way things are taken to be following the reduction. But what about the way things were taken to be prior to the reduction? Before the reduction took place we thought that there were things of kind M, whereas following the reduction we came to think that there are only things of kind P that happen to occupy the M-role. What are those things of kind M that we no longer believe in? E.g., what did we think cats and pains were before we came to believe that there are merely collections of particles that are arranged cat-wise, or that there are merely physical states that cause winces and groans? Well, it seems we simply thought that there *were* cats, not merely things arranged cat-wise that might have been arranged dog-wise; and that we thought that there *were* pains, not merely painful things that might not have been painful. Moreover, we thought that although cats are constituted by some collections of particles, they might have been otherwise constituted; and that although pains may be realised in some physical states, they might have been otherwise realised just as much. And generally, we thought that M's are sortally M, and merely contingently P's, and not the other way around.

We can thus offer the following schematic structure for the difference between a reductionist and a non-reductionist way we take the world to be. Following a reduction we take the things we use to call "M" to be P's that occupy the M-role, i.e., to have the property P sortally (i.e., necessarily and sufficiently[21]), and the property M contingently; before the reduction, by contrast, we took those things to have the property M sortally and the property P (if we were aware of it) contingently.

4.2.1.1 Nominal essences

The non-reductionist thinks that there are things that are sortally M and contingently P. The reductionist takes them to be merely contingently M and sortally P. Both cannot be

right at the same time. At least one of them must be wrong. At least one of them believes in things that do not really exist. Locke thought that we are in such a state; that most of what we believe to exist does not really exist as such. In Locke's terminology, what we take to be *real essences* are in fact merely *nominal essences*.[22] A nominal essence, as opposed to a real essence, is something that *we* take to be an essence, whereas in *reality* it is not. But whence these nominal essences? How do we arrive at such illusions?

Locke believed that they were the product of our own ideas and thoughts projected onto real essences in the world. Locke's view thus has a realistic aspect – his belief in real essences – as well as a conventionalist aspect – his belief that what we take to be essences are in fact our own constructions and are thus merely nominal. In more detail, nominal essences are constructed

> by collecting such combinations of simple ideas, as are by experience and observation of men's senses, taken notice of to exist together, and are therefore supposed to flow from the particular internal constitution, or unknown essence of that substance. Thus we come to have the ideas of a man, horse, gold, water, *etc*…[23]

In the same spirit, talking about species Locke later says,

> I do not deny, but nature, in the constant production of particular beings, makes them… very much alike and of kin one to another: but I think it nevertheless true, that *the boundaries of the species, whereby men sort them, are made by men.*[24]

Accordingly, we may say that a nominal sortal *term* is a term that *we* take to be a sortal term, whereas in reality it is not, i.e., it does not correspond to a sortal property. In short, the adjective "nominal" indicates the way we (wrongly) take things to be, as opposed to the way things really are.[25]

4.2.2   The role/occupant distinction

In the remainder of this chapter, I would like to propose two additional semantic functions – to be termed "role" and "occupant" – that we haven't so far encountered in our discussion. The great importance of these functions will be argued through their ability to solve three related problems. Most notably, it will be argued that none of the semantic functions we encountered so far – namely, descriptive or rigidified (reference-fixing or directly-referring) – can explain the behaviour of "cat" in the following ordinary expressions of a reductionist view:

> (1) Strictly speaking, there are no cats; there are only collections of particles arranged cat-wise.

It will be claimed that it is only an occupant term (to be defined shortly) that can do the work here.

Below are three associated problems.

First problem: Reductionists believe that M is nothing but P that occupies the M-role. E.g., the mereologist believes that cats are nothing but collections of elementary particles arranged cat-wise. This view can be expressed by saying (1) above. Alternatively, some will choose to express the *same* view by saying,

(2) Of course there are cats; they are just collections of particles arranged cat-wise.

But how can the same metaphysical view be expressed by two seemingly contradicting statements – one admitting the existence of cats, while the other denying it?

Second problem: In his robot-cats thought experience, Putnam said that "Even if cats turn out to be robots remotely controlled from Mars we will still call them 'cats' … Not only will we call them cats, they are cats".[26] This claim is grounded in Putnam's view that "cat", like other natural-kind terms, is a rigid (indeed, rigidified) designator. However, some may feel rather uncomfortable with this assertion. They may feel that what we should say following such a discovery is that, "we thought that there were cats in the world; however we were wrong. What we called "cats" were not really cats but in fact robots. Thus, we should stop calling them "cats", for they are not cats".

How shall we account for this intuitively reasonable approach?

Third problem: Reduction is often taken to be a vindicative process. Thus, if P's are credible entities – e.g., physical states – whereas the existence of M is doubted – e.g., of mental states such as pains – then by showing that M's are in fact just P's, the existence of M's is thereby vindicated. E.g., by showing that mental states are just physical states, mental states are allegedly thereby vindicated. However, some may feel uncomfortable with this. In particular, some non-reductionists may feel that as a matter of fact the reduction has eliminated M's rather than vindicated them. I believe that intuitively we sense there is something to this contention. How shall we account for it though?

I propose that the three required accounts are all based on a certain use of terms like "cat" and "pain" – a use that will be termed here *occupant*. This use will be specified while

addressing the first problem, and then employed in addressing the other two problems as well.

4.2.2.1 First problem: having reduced M to P, does M still exist?

A reductionist view can be stated by expressing statements such as (1) but also by stating statements such as (2). The problem is that one states that cats (or pains etc.), exist, whereas the other denies their existence. The solution lies in a difference in the way the two sentences use the term "cat". Let me demonstrate this.

A reductionist who states her view by expressing (2) may be (pragmatically) using "cat" in more than one way. She may be using "cat" descriptively, i.e., to simply designate that which has the property of being a cat, i.e., of occupying the cat-role. On such a use, "cat" designates, with respect to each counterfactual world, that which occupies the cat-role in that world (whether it is identical to the actual referent of "cat" or not). On this use, "cat" may turn out a nonrigid term (in particular, if occupying the cat-role is a contingent property, as reductionists in fact believe). The sentence then goes on to say something about those actual cat-like things; namely, that they are sortally collections of particles (and hence only contingently occupy the cat-role).

Alternatively, the reductionist who asserts the same sentence (2), may be (pragmatically) using "cat" as a rigidified term, i.e., to designate that which is of the same kind as *this*. On such a use, "cat" designates, with respect to every counterfactual world, that which is of the same kind of these actual referents of "cat" (whether it occupies the cat-role in that counterfactual world or not). Again, the sentence then goes on to say that these actual referents are sortally collections of particles (and hence, that when they happen to occupy the cat-role, they do so only contingently).

Thus, although a descriptive "cat" and a rigidified "cat" are, of course, crucially different – e.g., the former may be nonrigid whereas the latter may not – nevertheless both can refer to things that are sortally collections of particles and contingently occupy the cat-role. Consequently, both terms can be used to express the same reductionist view (i.e., that cats are just (sortally) collections of particles that are (contingently) arranged cat-wise,) by stating the same sentence (2).

Sentence (1), no less than (2), seems a very natural way of expressing the same reductionist view. However, it appears more difficult to account for the use of "cat" in

(1). On the one hand, "cat" here cannot be used as a descriptive term, since if it were, then "cat" would apply to these objects merely by virtue of them occupying the cat-role, i.e., of being cat-like (albeit only contingently). Just like "red" applies to red things by virtue of being (albeit only contingently) red. So in that use it makes no sense to claim that cats do not exist. On the other hand, "cat" in this case can neither be rigidified; for if it were, it would simply designate these things, whatever their sortal property is, in which case, again, there would be no reason to deny the existence of cats. It seems then that "cat" in the case of (1) is taken to have a function which we haven't thus far encountered.

I suggest that this function is the following. "Cat" in (1) is (pragmatically) taken to stand for "that which is not merely cat-*like*, but rather that which *is* a cat", in other words, "cat" in (1) means, "that which has the property of being a cat *sortally* and not merely contingently". And since according to reductionism the things we call "cat" are sortally collections of particles, and merely contingently arranged cat-wise, the term "cat", thus understood, does not apply to such entities. So the reductionist who uses "cat" in this way thinks in fact that cats do not exist, and hence states her view by uttering (1).

We can generalise the use of "cat" in (1). A term "M" may be taken, pragmatically, to mean "that which M's *sortally* (i.e., has the property M sortally)". (Those who use it in this way are likely to think, of course, that this is its *semantic* meaning). Correspondingly, there may also be a use of the term "M" that takes it to mean "that which M's *contingently* (i.e., contingently occupies the M-role)". E.g., some may use "cat" to mean "that which is (contingently) arranged cat-wise". (Again, those who use "cat" in this way take this meaning to be the *semantic* meaning of "cat"). In fact, "cat" as appears in (2) may be used in this way as well (i.e., in addition to the descriptive and rigidified options mentioned above); on this use, (2) will amount to "that which contingently occupies the cat-role, exists" – a statement that fits the reductionist's view. Let us call such "M" that stands for "that which (contingently) occupies the M-role" a *role-term*. Correspondingly, let us call "M" that stands for "that which *is* (i.e., sortally) an M" an *occupant term*.[27]

Overall then, we have considered four types of terms here: descriptive terms, rigidified terms, role terms and occupant terms. All are *possible* semantic functions of referring terms. i.e., there is a possible language that includes all four functions *semantically* (whether they are syntactically distinguished in that language or not). However, we are

not engaged in this discussion in semantic functions; rather, we are interested in ways in which speakers *use* terms. It is suggested that someone who states (1), *uses* "cat" *as* an occupant term. Similarly, someone who states (2), *uses* "cat" either *as* a descriptive term, or *as* a rigidified term, or *as* a role-term. And this is regardless of what the true semantics of "cat" is (if there is a fact of the matter in this respect at all).

We can get much clearer about the differences between the four terms by comparing their intensions. (For simplicity, we shall suppose a constant actual referent P+M.)

An occupant term "M" stands for "that which sortally M's (i.e., that which has the property M sortally)"; a role-term "M" stands for "that which contingently M's (i.e., that which contingently occupies the M-role)". Like descriptive terms, role terms and occupant terms designate that which fits their (constant) descriptive contents. Yet, unlike the case of descriptive terms, the descriptive contents of role terms and occupant terms involve *metaphysical* conditions: an occupant term "M" designates only things that are *sortally* M; a role term "M" designates only things that are *contingently* M. A descriptive "M", by contrast, designates *whatever* M's (i.e., has the property M) regardless of whether M is a sortal property or not.

In this discussion, we suppose that M is a manifest property (e.g., being arranged cat-wise). If the criteria of identity are material, then M is a contingent property. Since an occupant term "M" requires that M is a sortal property, relative to such material criteria, an occupant term "M" will have no designation. Yet a role term, as well as a descriptive term, will designate things that M. The intensions of "M" are thus:

Semantic function of "M"	Actual referent	Criteria of identity	W_1 P+M	W_2 Q+M	W_3 P+N	W_4 Q+N
Descriptive			+	+	-	-
Occupant	(P+M)	Material	-	-	-	-
Role			+	+	-	-

Table 4.1

If, by contrast, the criteria of identity are manifest, then M *is* a sortal property, and in this case an occupant term "M" will designate things that M, and now a role-term "M" will have no designation. Again, a descriptive term will also designate things that M:

Semantic function of "M"	Actual referent	Criteria of identity	W_1 P+M	W_2 Q+M	W_3 P+N	W_4 Q+N
Descriptive			+	+	-	-
Occupant	(P+M)	Manifest	+	+	-	-
Role			-	-	-	-

Table 4.2

Lastly, the intension of a role-term "M" and that of an occupant-term "M" are also distinct from the intension of a rigidified "M". For instance, if the criteria of identity are material, a rigidified "M" will designate the P+N stuff on counterfactual W_3 (Row 6 in Table 4.3 below), whereas both a role term "M" and a occupant term "M" will never do (Rows 2-5 below); if they designate at all, they only designate things that are M (the same applies to a descriptive term (Row 1)). In other words, the modal statement "the M might not have M'd" is true for a rigidified "M" but not for all other three "M"'s. Overall then, the intensions of the four types of "M" are the following:

	Semantic function of "M"	Actual referent	Criteria of identity	W_1 P+M	W_2 Q+M	W_3 P+N	W_4 Q+N
1.	Descriptive		(Whatever)	+	+	-	-
2.	Occupant		Material	-	-	-	-
3.			Manifest	+	+	-	-
4.	Role	(P+M)	Material	+	+	-	-
5.			Manifest	-	-	-	-
6.	Rigidified		Material	+	-	+	-
7.			Manifest	+	+	-	-

Table 4.3

Returning to our first problem, the proposed solution is as follows: The reason for the apparent contradiction between sentences (1) and (2), despite the fact that they both aim to express the same view, is that they use the term "cat" differently; specifically, (1) takes "cat" to be an occupant-term, whereas (2) does not (it is underdetermined whether "cat" in (2) is used as a descriptive term, or as a rigidified term, or as a role-term). Thus the contradiction between the two, at least on the pragmatic level of use, is only apparent. The two tokens of "cat" in the two sentences are used as different terms.

We can now utilise the two newly introduced pragmatic uses of terms to quickly solve the other two problems.

4.2.2.2   Second problem: Putnam's robot-cats scenario

Putnam says that were we to discover that cats were robots, the term "cat" would refer to robots, since, "cat" is a natural-kind term, and hence a rigidified term. Viz., Putnam claims that in such a case, "cat" would *semantically* refer to those robots. Maybe he is right. I do not know how to argue about the correct *semantic* function of terms like "cat". I don't even know if there is a fact of the matter as far as semantics are concerned. However, it is hardly debateable that, contrary to what New Theorists like Putnam prescribe, natural-kind terms are frequently *used*, i.e., pragmatically, as occupant-terms. E.g., "cat" is frequently used, pragmatically, to mean "that which *is*, i.e., sortally and not merely contingently, a cat". And since the robots in Putnam's scenario only contingently occupy the cat-role, they, relative to this occupant use, do not qualify as cats, and hence do not deserve the title. And it is for this reason that some people, namely, those who use "cat" as an occupant-term, will feel uncomfortable with Putnam's continuing to call them "cats", despite the discovery.

4.2.2.3   Third problem: the apparent vindicative nature of reduction

Some realms of existence are considered more credible than others. Typically, the physical realm is considered more credible than the mental realm, whose existence is doubted. Those who deny the existence of the realm in doubt are called *eliminativists*. Those who think that it exists may do so for two reasons. They may think that it exists independently of the credible realm; if they also acknowledge the existence of P's alongside such M's, they are called *dualists*. Alternatively, they may think that the realm in doubt is *reducible to* the credible realm, and that it is for this reason that it exists. Schematically, since M is nothing but P, and since P exists, M also exists. Traditionally, they are called *reductionists*, or, more specifically, *vindicative reductionists*. Thus, speaking more broadly about explanations in general, Lennon and Charles say,

> Vindicating the legitimacy of certain modes of explanation requires their reduction to phenomena whose intelligibility was considered less problematic or to explanations whose value was assured.[28]

Let us attend to the metaphysical level first, and then to the linguistic level.

On the metaphysical level, a reductionist ontology includes things that are, i.e. sortally, P's, some of which happen to occupy an M-role. E.g., it includes things that are

(sortally) physical states, some of which happen to occupy the pain-role but might not have occupied it. By contrast, a non-reductionist ontology (or pre-reductionist ontology, which is often the common ontology) includes things that *are*, i.e. sortally, M, and that are contingently constituted by P's. E.g., A non-reductionist ontology includes things that *are* pains (i.e. sortally), which happen to be realised in certain physical states, but that might have been, in principle, otherwise realised. Thus, all that reductionism admits into its ontology are P's that happen to be M-like. Yet the reduced things, the things that *are* M's (i.e., things that are *sortally* M), are deemed nominal, and are ultimately denied.

On the linguistic level, reductionists may use the term "M" either as an occupant-term or not (the latter case includes using "M" as a descriptive term; or as a rigidified term; or as a role-term).

Combining the metaphysical level with the linguistic level yields the following. A reductionist that takes "M" to be an occupant-term will assert (as in (2)) that, "strictly speaking, there are no M's; there are just P's that occupy the M-role". By contrast, a reductionist that does not take "M" to be an occupant-term will claim (as in (1)) that, "M's exist". Clearly, it is the latter reductionist who is called *vindicative*. However, for such a reductionist, "M" simply refers to P's, that contingently M (i.e., have the property M contingently). For the non-reductionists, by contrast, "M" refers to M's (i.e., things that sortally M) that contingently P (i.e., have the property P contingently). Thus, what the reductionist intends by "M" is very different from what the non-reductionist intends by the same term. The things that the non-reductionist intends are in fact denied, i.e., eliminated, by the reductionist. Hence, there is an important sense in which, from the non-reductionist point of view, reduction does not *vindicate* the existence of M's, but rather *eliminates* them. In other words, moving from the non-, or pre-, reductionist use of "M" in "M's exist" to the reductionist's use of "M" in "M's exist" involves a *shift in the reference* of "M".

4.2.2.4   Summary

Let us summarise our treatment of the three puzzles. We discussed the relation between different ways in which we take things to be. We considered two ways in which the world may be taken to be in a metaphysical respect, and four ways in which the semantic function of terms may be taken to be.

The two metaphysical views regarding cats were, (a) the reductionist view, according to which what we call "cat" are sortally collections of particles that are contingently arranged cat-wise; and (b) the non-reductionist view (so-called "common ontology"), according to which what we call "cat" *are* cats, i.e., things that have the property of being cat sortally, and that are contingently constituted by certain collections of particles.

The four pragmatic linguistic views were, (i) "cat" is descriptive; it thus means, "that which has the property of being a cat (whether sortally or not); (ii) "cat" is rigidified; it thus means, "that which is of the same kind of *these*"; (iii) "cat" is an occupant term; it thus means, "that which *is* a cat (i.e., has the property of being a cat sortally)"; and (iv) "cat" is a role term; it thus means, "that which contingently occupies the cat role." (The latter two were newly introduced.) The differences between the four functions are manifest in their different intensions (see Table 4.3).

The metaphysical views are independent of the linguistic views. E.g., each metaphysical view about cats is compatible with each of the four linguistic uses of "cat".

Reductionists believe that the objects we call "cats" contingently occupy the cat-role (be they collections of particles, or robots, or some unspecified physical entities). Thus, for those reductionists who use "cat" as an occupant term, the term "cat" fails to refer. Accordingly, they will claim that "cats do not exist". By contrast, any of the other three uses of "cat" does allow designation of such objects, and hence reductionists who use "cat" in one of these ways will be happy to assert that "cats exist". Incidentally, this includes New Theorists, who use "cat" as a rigidified term, and functionalists, who use "cat" as a descriptive term. The difference between the two, however, lies in their response to the counterfactual:

(3) Cats might not have occupied the cat-role

New Theorists will endorse it (Row 6 in Table 4.3), while functionalists will deny it (Row 4).

A role-term "cat" refers to things that contingently occupy the cat-role. An occupant term "cat" refers to things that have the property of being a cat sortally. Hence, the two terms stand for different types of objects. It follows that a statement "cats exist" that uses "cat" as a role-term, and the same statement that uses "cat" as an occupant term, assert the existence of different things. Consequently, they are ultimately different statements.

There are various possible combinations of a metaphysical view with a linguistic view. Many of these combinations will result in similar statements. Thus, a certain statement may underdetermine the linguistic view and the metaphysical view of the one who expressed the statement. It follows that even explicit statements like "cats exist!" underdetermine what the speaker takes the world to be like. The next section demonstrates this surprising fact.

4.2.2.4.1   What can we learn from someone's assertion that "cats exist"?

Someone declares,

(4) "Cats exist!"

What can we learn from this statement? One lesson that we can now draw from our discussion above is that as a matter of fact we can learn very little from such a statement, even about what the *speaker* asserts the case to be. The following possibilities are all consistent with (4).

Firstly, on the linguistic level, the speaker may take "cat" to be rigidified. Hence, "cat" simply designates these things (that are actually cat-like), whatever they are. In this case, therefore, she may have any metaphysical view about the world and still state (4). E.g., she may believe that the cat-like things we call "cats" are animals, or she may believe that they are robots; she may believe that they contingently occupy the cat-role, or that they are sortally cats. In other words, if she believes that "cat" is rigidified, then as long as things of "this" kind exist, whatever this kind is, she believes (4).

Secondly, she may believe that "cat" is a descriptive term, whose descriptive content is "that which occupies the cat-role (i.e., whether sortally or contingently)". Hence, that "cat" simply designates cat-like things, regardless of what these cat-like things are. Thus, as in the previous case, she may have any of the above metaphysical views about the world and still hold (4) true. In other words, if she believes that "cat" is descriptive, then as long as there are cat-like things, she also endorses (4).

Thirdly, she may, on the linguistic level, believe that "cat" is a role-term, i.e., that it means, "that which contingently occupies the cat-role", and on the metaphysical level, that the things which we call "cat" indeed contingently occupy their cat-role. Hence, that they deserve the name "cat", and that (4) is true.

And fourthly, she may believe that "cat" is an occupant term, i.e., that it stands for, "that which *is* a cat (i.e., sortally, not merely contingently)", and in addition that the things which we call "cat" indeed have the property of being a cat sortally. Hence, she will believe that *they* deserve the name "cat" and that (4) is true.

Overall then, the mere assertion that "cats exist", despite appearance, teaches us very little even about the *speaker's view* of the world. In particular, it determines neither her view of the semantics of "cat", nor her metaphysical view about what the world contains. All it teaches us is that she is willing to endorse (4).

Endnotes

---

[1] See also Section 1.3.2 and 3.1.

[2] Kripke (1972/1980, pp. 119-121)

[3] Excluding the odd cases discussed in Section 3.3.5

[4] The origins of this view are largely in Frege (1884/1974, § 47). Indeed, according to Frege, the sentence "all whales are mammals", like the sentence "all squares are rectangular", is about the relations between the *concepts*, "whale" and "mammal": when the second concept is included in the first, whatever falls under the first falls under the second. Hence, on Frege's view, such sentences are analytic (and hence *a priori* and necessary). The New Theory, clearly, has it by contrast that such sentences, although necessary, are by no means analytic nor *a priori*, but rather synthetic and *a posteriori*. Thus, following the insights of the New Theory, the status of such statements has been a subject of a considerable dispute.

[5] It should be reminded here though that LaPorte (2004) claims that contemporary biology tends to classify biological taxa according to historical evolutionary criteria. (See note 83 in Chapter 1.)

[6] Kripke (1972/1980, p. 138)

[7] Although Kripke insists that all natural-kind terms are rigidified, he nonetheless allows that the reference of some natural-kind terms is fixed not by a baptising act, but rather by a description. In our terminology, the claim is that although some natural-kind terms are directly-referring, others are reference-fixing. E.g., it may be that "water" was not stipulated to designate '*this* stuff' in all counterfactual worlds, but rather was stipulated to designate in all counterfactual worlds, that which is *watery* in the actual world. However, the important point to our discussion is that *some* natural-kind terms *are* held directly-referring and not reference-fixing; and in particular, that "fish" is such a directly-referring natural-kind term.

[8] Admittedly, fish is not a species but rather a higher biological category. In other words, being a fish is not a *sortal* property but an *essential* property. i.e., a property that is necessary yet not sufficient to its object (see Section 3.3.6). Nevertheless, rigidified terms can be stipulated to designate all and only members of a category higher than the species, i.e., all and only objects that share a certain *essential* property. Thus the "fish" example is still very much relevant here.

[9] Kripke (1977)

[10] Kripke (1972/1980, p. 138)

[11] Kripke (1972/1980, p. 136)

[12] As noted in Endnote 8 in this chapter, the fact that it is an essential property which is at stake here, and not essence, i.e., sortal property, makes no difference to our present analysis.

[13] Putnam (1962, p. 660). See also his (1977, p. 107). Indeed, much of what Kripke (1972/1980) says about three-legged tigers, automaton Nixon, etc., points in a similar direction.

[14] Putnam (1962, p. 660)

[15] Following our discussion of supervenience above (Section 3.4), it is interesting to explore the relation between reduction and supervenience. Supervenience was defined as follows: M supervenes on P iff there can be no M-difference without there being a P-difference (but not necessarily vice versa). Reduction requires, and hence entails, supervenience (i.e., supervenience is *necessary* for reduction): if M reduces to P then there can be no M-difference without there being a P-difference. It is less clear whether supervenience also entails reduction (i.e., whether supervenience is also *sufficient* for reduction). This depends on whether reduction means property *identity* (i.e., one-one correlation between M's and P's) or not; if it does, then supervenience does not suffice for reduction, because supervenience allows for multiple realisation (i.e., one-many correlation). See Kim (1984; 1990).

[16] Lewis (1994, reprinted in his 1999, p. 303). Jackson (1998a, p. 59) follows a similar line:

> Pr. 1   Temperature in gases = that which plays the temperature ('T') role in gases (conceptual claim)
>
> Pr. 2   That which plays the temperature role in gases = mean molecular kinetic energy. (Empirical discovery)
>
> Conc.   Temperature in gases = mean molecular kinetic energy. (Transitivity of '=').

However, Jackson indicates two ways in which the referring terms can be read: as rigidified and as descriptive. We shall come to this point soon below.

[17] For a classical exposition of causal-role functionalism see Smart (1959). Notable defenders of the view are Armstrong (1968), and Lewis (mainly 1972; 1994)

[18] Lewis (1983, p. 110)

[19] E.g., Chisholm (1973)

[20] It should be noted though that it is not the mere fact of occupying an M-role that renders M contingent; it is the fact that it is *a* P – i.e., something that is *sortally* a P, and which occupies the M-role – which makes M contingent. P may also be said to occupy the P-role, but that role, i.e., having the property P, would remain sortal (i.e., not contingent). For example, a collection of particles may be said to occupy the collection-of-particles role (i.e., being constituted by a collection of particles); yet in this case occupying the role would be a sortal property of the

collection. It is only when this collection occupies some other role, say the cat-role (i.e., being arranged cat-wise) that the role is thereby rendered contingent to the collection.

[21] In fact, as we have seen in the previous chapter (Section 3.3.6) a sortal property is that necessary and sufficient property that states what the object that instantiates it consists in. However, in view of the overwhelming overlap between sortal properties and necessary and sufficient properties, we shall treat them for the purpose of the present discussion as the same.

[22] Locke (1689, III iii 15, 16, 18)

[23] Locke (1689, II xxiii 3)

[24] Locke (1689, III vi 37)

[25] We can distinguish two types of nominal essences. Take the mereologist reductionist. She believes that cats are nominal essences. Now consider immaterial ghosts. Since the mereologist believes that there are no such things, she will claim that ghosts are nominal essences as well. Yet there is a clear difference between the two: whereas there are things that occupy a cat-role, there are no things that occupy an immaterial ghost role, simply because, according to the mereologist, there are no immaterial things. To generalise: a property is nominal if it is not sortal. A property is not sortal if it is, for instance, contingent; but also, if it is simply non-existent. These are two different ways in which a property may be held nominal.

[26] Putnam (1977, p. 107). Kripke (1972/1980) makes similar points about *discovering* that Nixon was an automaton, or that the Queen was not the daughter of George VI: we ought to still call him "Nixon" and her "Queen Elizabeth". (Recall that this type of possibility is crucially different from considering, *given* that Nixon is a person, the possibility that he might have been an automaton. See Sections 1.1.3, 1.5.1 and 2.3 (also Section 5.1.1.4 below) for reminders about the differences between the two types of possibilities.)

[27] Note that an occupant term is not equivalent to a sortal term. A sortal term is just a descriptive term, e.g., "that which has the property of being a cat (whether sortally or not)", whose descriptive content *turns out* to be linked with a sortal property. Thus, if being a cat is a sortal property, then a descriptive term "cat" *turns out* to be a sortal term. If, by contrast, being a cat is not a sortal property, then a descriptive "cat" *turns out* not to be a sortal term. In other words, a sortal term, being a descriptive term, involves no metaphysical conditions *as part of its descriptive content*. An occupant term, by contrast, does involve a metaphysical condition in its descriptive content. Consequently, every occupant term is a sortal term but not vice versa. However, a term that is *used as* an occupant term may be either sortal or not. Using "M" as an occupant term merely indicates that *the speaker believes* M to be a sortal property – a belief that need not correspond to reality. The next paragraph clarifies this point.

[28] Lennon and Charles (1992, p.1)

5 TWO SAMPLE APPLICATIONS OF THE THESIS

Overview

In the final chapter we put the thesis developed in this dissertation to work. It is applied to two case-studies from relatively recent philosophical literature: the Kripke-Lewis debate over the identity theory of mind, and the debate over the significance of Donnellan's referential/attributive distinction. Each of which is a much-discussed debate that hasn't been resolved for several decades. Applying our thesis to the two cases will generate novel accounts of the debates, and will thus serve to illustrate the importance of the thesis.

First case-study: the Kripke-Lewis debate over the identity theory of mind. Kripke argues against the functionalist identification of mental states with physical states, whereas Lewis defends it. Our analysis is based on exposing the intension of "pain". This analysis will reveal that the intension of "pain" as entailed by the view that Kripke attacks is different from the intension of "pain" as entailed by real functionalism. So Kripke fails to attack genuine functionalism. Interestingly, the same applies to Lewis's argument: the intension of "pain" as entailed by the view that Lewis attacks is different from the intension of "pain" as entailed by Kripke's original view. Thus, Lewis also fails to attack Kripke's genuine view. Moreover, our analysis will reveal that both parties in fact attack the same view. In addition, it will be shown that although Lewis's genuine view and Kripke's genuine view comprise different elements (namely different values for the variables in our formula), those elements happen to produce the same intension of "pain". Specifically, in both cases "pain" ends up designating, with respect to every counterfactual world, that which occupies the pain-role in that world. Such an intension is immune to the objections raised by both parties, and hence both genuine views remain unharmed by the objections.

Second case-study: the debate over the significance of Donnellan's distinction between attributively used definite descriptions and referentially used ones. Here, our analysis is based on exposing the intension of "Smith's murderer". In the first stage we

consider counterfactual worlds. Based on this consideration we conclude that three interpretations of Donnellan's distinctions – namely, in terms of singular vs. general propositions (which seems to be the standard interpretation); in terms of *de re/de dicto*; and in terms of rigidity – are fundamentally one and the same interpretation. It thus follows that Kripke's objection to the latter two interpretations is as effective against the first, standard interpretation. Next, we further introduce the distinction between designation in the actual world vs. designation in counterfactual worlds. This two-dimensionalist analysis reveals that (a) from an exegetical point of view, there are (at least) three consistent ways of interpreting Donnellan's distinction, i.e., all equally compatible with his original characterisation; and (b) from a philosophical point of view, that there are (at least) three important distinctions that apply to definite descriptions (regardless of which of them provides the correct interpretation to Donnellan's original distinction). In addition, such analysis provides a clear understanding of the "mechanism" by which Kripke's argument against the two interpretations operates.

5.1 The Kripke-Lewis Debate over the Identity Theory of Mind – A Critical Analysis

Introduction

Functionalism offers a unique way of identifying one set of alleged entities with another. In particular, it offers the identification of mental states with physical states. In the third lecture of his *Naming and Necessity*[1], Kripke argues directly against this doctrine, and in particular against the functionalist identification of pain with a physical state. David Lewis, who advocates functionalism, takes the opportunity to defend it in his "Reduction of Mind"[2], and at the same time to argue against Kripke. Specifically, Kripke accuses functionalism of allowing "pain" to designate some possible non-painful states, whereas Lewis accuses Kripke's view of allowing "pain" not to designate some possible painful states. Thus both objections accuse their opponent of having some unacceptable modal consequence.

In what follows I wish to argue that both objections are unsuccessful in attacking their opponent, and moreover, that they are both unsuccessful for the same reason. In particular, both objections successfully attack a certain view, but this view is different from their opponent's. It is a "straw-man" view. Surprisingly, both objections turn out to attack the same view. (Although each objection attacks it on different grounds.) This view is a "hybrid" view, crossbred between Lewis and Kripke's competing views, and is thus an additional view – neither Kripke's nor Lewis's. Hence each of the competing views, being different from this rejected view, remains unharmed by the attack launched against it.

Methodologically, this observation will be based on an analysis of the different views in terms of their semantic and metaphysical components, as detailed below. Based on this analysis, we shall determine the intensions of "pain" – i.e., the reference of "pain" with respect to different possible worlds[3] – that are entailed by the different views. Comparing these intensions will serve as a convenient means of comparing the views and assessing the objections.

The structure of the discussion is as follows. I will begin by presenting the debate by introducing the competing views, and the objections they raise against each other. I will then proceed to provide a comparative analysis of these views and of the objections.

This analysis will lead to the conclusion of this paper, namely that each of the two objections is unsuccessful in arguing against its opponent. I will close with an overview of this debate.

5.1.1 The debate

5.1.1.1 Functionalism

Functionalism offers a way of identifying a "suspected" set of entities with another set of "credible" entities. Consider genes – to borrow an example from Armstrong[4]. If we take the term "gene" to mean "that which occupies the causal role of the gene", i.e., something like "that which transmits hereditary characteristic etc.", then, by discovering that that role is actually occupied by the DNA molecule, we arrive at the identity "the gene is the DNA molecule". Similarly, if we understand a mental term like "pain" to mean "that which occupies the role of being painful", i.e., something like "that which causes winces and groans etc.", and we find that this role is occupied by some physical state C,[5] then we arrive at the identity "pain is C". If we further assume, like the functionalist does, that the world is ultimately occupied by physical entities, then applying the same procedure to all mental terms will result in an overall identification of the mental with the physical.

Moreover, the identity generated by functionalism is not symmetrical. Thus, pain, for example, is not only said to be "identical" to C, but it is also said to be "in fact", or "nothing over and above", C. Thus pain is *reduced to* C, and in general, the mental is *reduced to* the physical. Pain and C are of course types (that have many tokens), and thus the functionalist reduction applies to types and not merely to token.[6]

This, in short, is the functionalist identity theory of mind.

5.1.1.1.1 *The modal status of reductive identity statements*

Taking "pain" to stand for "that which occupies the pain role", whilst assuming that the occupant of this role (as well as of any other role) is a physical entity, entails that this role might have been occupied by states other than the actual one. E.g., although the pain-role is actually occupied by the type C, it might not have been occupied by this type. In other words, the identity statement "pain = C", if true, is contingently true; it is true of

the actual world, but false of some other possible worlds, namely, worlds in which the pain-role is not occupied by C.[7]

This point has implications for *rigidity*. A term is rigid iff it designates the same referent in all possible worlds; otherwise it is nonrigid. Since "pain", for the functionalist, designates different types of states in different possible worlds, it is thus rendered by this view a nonrigid designator.

5.1.1.2   Kripke's alternative

Contrary to the functionalist, Kripke thinks that "pain" *is* a rigid designator.

(A note about rigidity: As aforementioned, pain is a type and not a token. This makes the designation of "pain", if rigid, somewhat problematic (as we have witnessed in Section 2.2.4): what exactly does it designate rigidly? Is it the abstract type (or property), or is it the particular collection of tokens that fall under the type? (In the latter case, rigidity is best accounted for by designating, with respect to every possible world, tokens of *the same type as* that of the actual referents.) For the purpose of this discussion, we shall remain agnostic on this issue. E.g., we shall remain agnostic on whether "pain", if rigid, designates in all possible worlds the same type of state, or states of the same type. For want of a neutral term, we shall use either "type of state" or "states of a type" to indicate the disjunction of the two.)

Thus, "pain", if rigid, designates the same type of state in every possible world, and so does "C". It thus follows that if the two terms designate the same type of state in the actual world, they must also designate the same type of state in all possible worlds. (Similarly, if they designate different types of states in the actual world, they must also designate different types of states in all possible worlds). Hence the identity statement, if true of the actual world, must also be true of all possible worlds (and if false of the actual world, must also be false of all possible worlds). In short, the identity statement "pain = C", if true, is *necessarily* true, and not merely contingently true.

Thus a central issue which Lewis's functionalist view and Kripke's view are at odds about is the rigidity of "pain", and, accordingly, the question of the modal status of the identity statement "pain = C".

5.1.1.3 Kripke's objection to functionalism

In *Naming and Necessity*, Kripke argues directly against functionalism. The argument he puts forward is an *ad absurdum* argument – it rejects functionalism on the grounds that it leads to absurdity:

> A typical view is that *being a pain*, as a property of a physical state, is to be analysed in terms of the 'causal role' of the state,[8] in terms of the characteristic stimuli (e.g. pinpricks) which cause it and the characteristic behaviour it causes. … All I need to observe here is that the 'causal role' of the physical state is regarded by the theorists in question as a contingent property of the state, and thus it is supposed to be a contingent property of the state that it is a mental state at all, let alone that it is something as specific as pain. To repeat, this notion seems to me self-evidently absurd. It amounts to the view that the *very pain I now have* could have existed without being a mental state at all.[9]

The argument is simple: suppose, as the functionalist suggests, that "pain" designates some physical state of type C, i.e., that the actual occupant of the pain-role is C. It seems that all would agree, including the functionalist, that there is a world in which this C does not occupy the pain-role, i.e. the role of being painful. But this would entail that in that world *pain* does not occupy the pain-role, or, in Kripke's words, "that *that very sensation* could have existed without being a sensation." Furthermore, if what Kripke says is true of a particular token of pain, it is also true of *any* token of pain, and thus of the whole type. Moreover, a similar argument may apply to any functionalist reduction of types. So the attack is on the whole functionalist reductive enterprise.

Thus functionalism is accused by Kripke of entailing absurd outcomes, and hence of being unsustainable.

(Note however that on this basis Kripke does not reject *any* mental-physical identification in principle; his objection is directed to the specific sort of identification suggested by the functionalist.)

5.1.1.4 Lewis's objection to Kripke's view

As earlier remarked, a key disagreement between Kripke and Lewis is about the modal status of identity statements like "pain = C"; and, accordingly, about whether "pain" is a rigid designator or a nonrigid one. Lewis's argument against Kripke's view is thus directed against the rigidity of terms like "pain". Like Kripke's argument, Lewis's argument also aims to show that his opponent's view leads to unacceptable modal

consequences, and hence has to be abandoned. In his "Reduction of Mind", he reasons as follows,

> Kripke vigorously intuits that some names for mental states, in particular 'pain' are rigid designators: that is, it's not contingent what their referents are.[10] I myself intuit no such thing, so the non-rigidity imputed by causal-role analysis troubles me not at all.
>
> Here is an argument that 'pain' is not a rigid designator. Think of some occasion when you were in severe pain, unmistakable and unignorable. All will agree, except for some philosophers and faith healers, that there is a state that actually occupies the pain role (or near enough); that it is called 'pain'; and that you were in it on that occasion. For now, I assume nothing about the nature of this state, or about how it deserves its name. Now consider an unactualised situation in which it is some different state that occupies the pain role in place of the actual occupant; and in which you were in that different state; and which is otherwise as much like the actual situation as possible. Can you distinguish the actual situation from this unactualised alternative? I say not, or not without laborious investigation. But if 'pain' is a rigid designator, then the alternative situation is one in which you were not in pain, so you could distinguish the two very easily. So 'pain' is not a rigid designator.[11]

It is the last sentence that is aimed directly against Kripke. Suppose, again, that an actual occupant of the pain-role is some state of type C. We are asked to consider the counterfactual situation in which the occupant of the pain-role is not that C, but rather a state of a different type, say, D. Now if "pain" is rigid as Kripke claims it to be, then it designates, with respect to every possible world, the same type of state (or states of the same type[12]) as it designates in the actual world, namely, C. Hence, it does not designate the counterfactual state of type D. But this would mean that "pain" does *not* designate a state that is phenomenally indistinguishable from the actual state, or, in short, this would mean that "pain" does not designate a state that occupies the pain-role. And this result is unacceptable. Thus, concludes Lewis, "pain", contrary to Kripke's claim, cannot be a rigid designator, and is therefore nonrigid.

5.1.2 Analysis

So much for the presentation of the debate. Let us move on now to some analysis. In this part, I propose a reconstruction of the debate by revealing the elements composing the two views, and assessing the objections in the light of these elements.

The structure of this part is as follows. First, I propose that each view is best understood as being composed of two elements: a semantic one and a metaphysical one. The two views differ in both respects. This analysis will provide a clear comparison of the two views by specifying the exact respects in which they contrast. Subsequently,

determining the intensions of the term "pain" according to each view – i.e., what "pain" designates with respect to different possible worlds – will help to make these differences evident.

Next, the same analysis will serve to understand the objections as well; in particular it will reveal that in each case the metaphysical and semantical assumptions of the view that each objection actually attacks are different from the assumptions of the view that each objection claims to be attacking. In other words, both objections successfully attack a certain view, but this view is different from their opponent's one. The view that is in fact attacked is a mere "straw-man" view. Moreover, surprisingly, the two objections attack the *same* straw-man. Determining the intension of "pain" that is entailed by the attacked straw-man view as well will clearly illustrate its difference from the two genuine views.

Lastly, comparing the intensions of "pain" as entailed by each of the three views will serve as an explanation of the weakness of the attacked straw-man view, and, at the same time, of the strength of the two genuine views.

5.1.2.1   Functionalism analysed

The crux of functionalism lies in taking mental terms like "pain" to mean, "that which occupies the pain-role". Such reading has crucial implications – both semantic and metaphysical. On the semantic level, "pain" is thereby taken to be an abbreviation of a description, namely, "that which occupies the pain-role"; it thus designates, with respect to every possible world, that which fits this description in that world. In short, "pain", for the functionalist, is a descriptive term.[13]

On the metaphysical level, we have noted above that the distinction between role and occupant implies the (modal) contingency of the occupant to the role, i.e., the same role might be occupied by occupants of different types.[14] However, the independence implied by the role-occupant structure works both ways. i.e., the actual occupant might have occupied a different role. Thus, for instance, it is possible for the pain-role to be occupied by states of type other than C; and, also, it is possible for some C to occupy roles other than the pain-role. It follows, as Kripke observed above, that on the functionalist view occupying the pain-role turns out to be contingent to the occupants.

(Note that it is only the second of the above two possibilities – namely, the possibility that a certain state might not have occupied the role it actually occupies – that entails the contingency of the role to its bearer. By contrast, the mere possibility that the role may be occupied by different types of occupants does not in itself suffice to entail the contingency of the role. For example, the role of being a chemical element is occupied by different kinds of stuff, e.g., helium and hydrogen, and yet this role is essential (although not sufficient) to them; no kind of stuff can survive the loss of this property.[15])

It is the *combination* of these two implications of functionalism – namely, the semantic assumption that terms like "pain" are descriptive, *and* the metaphysical assumption that occupying the pain-role is contingent to the state that occupies this role – that entails the nonrigidity of terms like "pain", i.e., that terms like "pain" do not designate the same type of state in all possible worlds. Neither of these assumptions on its own can *guarantee* nonrigidity. Firstly, the mere semantic assumption of descriptiveness does not suffice for nonrigidity; this is because a descriptive term may, in principle, be rigid (albeit, *de facto* rigid). E.g., the term "the sum of 2 and 3" is descriptive – it designates, with respect to every possible world, that which fits its descriptive content – yet nonetheless it is also rigid; it designates the same entity, the number 5, in all possible worlds. Similarly, but somewhat more controversially, the term "stuff with atomic number 79" is also descriptive and, arguably, rigid – it designates gold in all possible worlds. Secondly, the mere metaphysical assumption of contingency does not guarantee nonrigidity either: a rigidified designator (i.e., *de jure* rigid[16]), e.g., the actualised description "the actual author of *Naming and Necessity*", is designed to designate, with respect to every possible world, the same referent that it designates in the actual world, namely, Kripke; it is thus rigid by stipulation. And this is regardless of Kripke's properties, and thus, *a fortiori*, regardless of the modal status of his properties. (*After* guaranteeing sameness of reference across worlds by using a rigidified designator, then, to be sure, the modal status of Kripke's properties will make a difference to *what* would count as the same as Kripke in other possible worlds and what would not. But this is only *after* sameness has already been secured by the rigidified term.[17])

Thus in sum, the functionalist view is composed of two elements: a semantic element, namely, that "pain" is a descriptive term, and a metaphysical element, namely that occupying the pain-role is contingent. It is the *combination* of the two elements which yields the nonrigidity of terms like "pain", that is implied by this view.

5.1.2.2 Kripke's view analysed

Like functionalism, Kripke's alternative view is also composed of a semantic element and a metaphysical one. On the semantic level, Kripke takes "pain", like other natural-kind terms, to be rigidified[18], i.e., each such term is designed to designate, with respect to every possible world, the same (type of) referent that it designates in the actual world.[19] Thus "pain", according to Kripke, is rigid by stipulation; its rigidity is guaranteed by the very semantic function of the term.

(It follows that unlike in the case of functionalism, the semantic element alone guarantees the rigidity of "pain". In the case of functionalism, by contrast, nonrigidity was a result of both the semantic *and* the metaphysical components of that view. However, as we shall soon witness, the force of Kripke's view relies nonetheless on the combination of *both* of its elements.)

On the metaphysical level, Kripke takes the property of being painful to be essential to its bearer. In his own words: "if something is pain it is essentially so".[20] So whatever has the property of being painful cannot survive the loss of this property. Moreover, being painful is not only essential, i.e., necessary, to its bearer, but it is also sufficient to it. As Kripke puts it:

> Pain … is picked out by an essential (indeed necessary *and sufficient*) property. For a sensation to be felt as pain *is* for it to be pain.[21]

What Kripke means by this, is that since being painful is a necessary and sufficient property of its bearer, it follows that it is the *criterion of identity* of its bearer, i.e., that something is pain iff it is painful. In terms of possible worlds, being necessary to its bearer means that if a state is painful in the actual world then it is painful in every possible world; being sufficient to its bearer means that, with respect to each possible world, whatever state is painful in that world, it is of the same type as the state that is painful in the actual world – they are both pains. Together, these two conditions entail that all and only painful states are pains, i.e., that something is pain iff it is painful.[22]

(Indeed, whereas functionalists talk about occupying the "causal pain-role", Kripke is talking about a "sensation" being "felt as" pain. Can we compare the two? I believe the answer should be yes. Both are talking about something like the property of occupying a pain-role, i.e., of being painful. The functionalist accounts for this property in terms of causes and effects, Kripke seems to have something more like qualia in mind.

However, it is clear that whereas functionalism takes occupying the pain-role to be contingent to the occupant (however we are to account for this property), Kripke takes this property to be essential and sufficient to the occupant (again, however, we are to account for this property). Accordingly, we shall henceforth use "occupying the pain-role" and "being painful" neutrally and interchangeably; i.e., whether it is to be accounted for in terms of causes and effects, in terms of qualia, or in other terms altogether.)

Overall then, the two views differ in both their semantic and their metaphysical components. In the semantic respect, whereas Lewis takes "pain" to be descriptive, i.e., "pain" designates, with respect to every possible world, that which fits the description "occupies the pain-role" in that world, Kripke takes "pain" to be rigidified, i.e., "pain" designates, with respect to every possible world, the same type of state that it designates in the actual world. On the metaphysical level, whereas Lewis believes that occupying the pain-role is contingent to its bearer, Kripke takes it to be essential (i.e., necessary) as well as sufficient. Table 5.1 below summarises this:

	Semantics	Ontology
Functionalism	"Pain" is descriptive; "pain" designates, with respect to each possible world, that which is painful in that world.	Occupying the pain-role is contingent
Kripke's view	"Pain is rigidified; "pain" designates, with respect to every possible world, the same type of state that it designates in the actual world.	Occupying the pain-role is essential (i.e., necessary) and also sufficient

Table 5.1

5.1.2.3   Kripke's objection analysed

To recall, Kripke's argument considers the counterfactual situation that is allegedly entailed by functionalism in which the actual occupant of the pain-role, say some C, does not occupy the pain-role. This, Kripke concludes, means that some pain is not painful, which is absurd.

On the metaphysical level, Kripke's scenario is in line with functionalism: by allowing the occupant of the pain-role to survive this role's loss, i.e., to stop being painful, being painful is rendered contingent – an element in the story that the

functionalist is happy to embrace. However, the functionalist should be less content with the semantic assumptions implicit in Kripke's scenario. Kripke's contention is that functionalism entails that pain might not have been painful. This is because functionalism allows that a state of type C, that is actually painful, might not have been painful. In possible-worlds semantics, this amounts to allowing a possible world in which C is not painful in that world. It thus follows that Kripke allows the (counterfactual) non-painful state of type C to deserve to be called "pain". In other words, Kripke takes "pain" to be a rigidified designator, i.e., to designate, with respect to every possible world, that which it designates in the actual world. (Had Kripke taken "pain" to be descriptive, i.e., to designate, with respect to each possible world, that which is painful *in that world*, "pain" would not designate the counterfactual non-painful C, and hence he would be barred from accusing functionalism for implying that "pain might not have been painful"). But according to functionalism, "pain" *is* a descriptive term, i.e., "pain" designates, with respect to each possible world, that which is painful *in that world*. Hence, for the functionalist, whereas the state in the actual world, the painful C, deserves to be called "pain", the state in the counterfactual world, the non-painful C, does *not* deserve to be called "pain", simply because it is not painful in that world; Just as Kripke does not deserve to be called "the author of *Naming and Necessity*" in worlds in which he did not author it (although he will still deserve to be called "the *actual* author of *Naming and Necessity*" in such worlds). Hence, on the functionalist view, the counterfactual situation is *not* one in which *pain* is not painful, but rather one in which something, which is *not* pain, is not painful. And this is far from being absurd.

Calling the state in the counterfactual situation "pain", as Kripke does, presupposes that "pain" is a rigidified term, i.e., that it simply designates, with respect to every possible world, the same (type of) state that it designates in the actual world. Thus, on this presupposition, since the actual referent, C, is called "pain", "pain" designates C in every possible world, including the non-painful C in the above counterfactual world. This semantic assumption is in fact just Kripke's own view on the semantics of "pain". But it is not the functionalist's semantic view. So, ultimately, the view that Kripke attacks is not the functionalist view. It is a "straw-man" view that shares the metaphysical element of functionalism, but not its semantic element.

To make the difference between functionalism and the straw-man view that Kripke attacks clearer, we can compare the intensions of "pain", i.e., what the term designates with respect to different possible worlds, as entailed by each view. We divide

possible worlds into the following four relevant types: W_1: the occupant of the pain-role is C (states of type C+P); W_2: the occupant of the pain-role is not C (states of type ¬C+P); W_3: C occupies another role instead of the pain-role (states of type C+¬P); and W_4: a world devoid of states C, and in which the pain-role has no occupant (states of type ¬C+¬P).[23] With respect to these possible worlds, the intensions of "pain" on the functionalist view and on the view that Kripke attacks clearly differ. Specifically, the functionalist's "pain" designates whatever is painful, and thus designates the type of state in W_1 and in W_2. By contrast, according to the view that Kripke attacks, "pain" designates the same type of state that it designates in the actual world, which is C+P; and since, in this scenario, P is considered contingent (i.e., not necessary) and C is considered necessary and sufficient, being the same as C+P amounts to just being C; thus "pain" designates the type of state in W_1 and in W_3. In total, the intensions of "pain" are thus the following:

	Semantics	Metaphysics	W_1 C+P	W_2 ¬C+P	W_3 C+¬P	W_4 ¬C+¬P
Functionalism	"Pain" is descriptive	P is contingent; C is essential	+	+	-	-
Straw-man functionalism attacked by Kripke	"Pain" is rigidified		+	-	+	-

Table 5.2

The two views thus have different designations in W_2 and W_3, and thereby ultimately different intensions.

Kripke's objection is based on the fact that "pain" allegedly designates the type of state in W_3, i.e., the C+¬P type of state (bottom row), which he takes to be absurd since this type of state is not painful. This, indeed, is a fair accusation. However, as is clear from Table 5.2 (top row), functionalism entails no such designation.

To conclude, the position that yields the absurd result is one that has the metaphysical element that being painful is a contingent property, and the semantic element that "pain" is a rigidified (i.e., *de jure* rigid) term. This is a hybrid position that has the metaphysical element of functionalism yet the semantics of Kripke's own view. Ultimately, it is not functionalism. Functionalism, as it is, was not shown to yield any absurdity, and thus remains unharmed. At least by this argument. Indeed, the functionalist may consistently accept Kripke's argument, i.e., that the straw-man view

that Kripke attacks leads to absurdity and hence needs to be abandoned, and at the same time continue advocating functionalism.

5.1.2.4   Lewis's objection analysed

Interestingly, an examination of Lewis's argument based on the analysis suggested here shows that his objection to Kripke is flawed in the same way that Kripke's objection is; it successfully rejects some view, but that view is not Kripke's. Specifically, the attacked view has Kripke's semantic element, but lacks Kripke's metaphysical element, which is replaced by the functionalist's metaphysical element. Let me demonstrate this.

To recall, Lewis's argument considers a counterfactual situation in which the pain-role is being occupied by a type of state other than the state C (which occupies the pain-role in the actual situation). Now if "pain" is rigid, says Lewis, it cannot designate the type of state in this counterfactual situation, $\neg$C, as this state is of different type from the actual C. But, since the type of state in the counterfactual situation, $\neg$C, is painful, this will entail that a painful state is deprived of being called "pain", a consequence that we should not accept. Thus Lewis concludes that "pain", contrary to Kripke's view, is a nonrigid designator.

On the semantic level, Lewis's scenario is in line with Kripke's view: "pain" is taken to be rigidified (i.e., *de jure* rigid) in that it is stipulated to designate, with respect to every possible world, the same type of state that it designates in the actual world. Given that the type of state in the actual world is a painful C, i.e., C+P, "pain" in this scenario is taken to designate whatever is identical to that C+P type of state. However, Lewis's premise that "pain" does not apply to the type of state in the counterfactual situation, namely, to the $\neg$C+P type of state, is at odds with Kripke's *metaphysical* assumptions. Let me explain this. Recall that according to Kripke, being painful is a property that is necessary and sufficient to its bearers. Hence, merely having that property in some counterfactual world guarantees that that type of state is identical to the one that is painful in the actual world. Hence, according to Kripke's metaphysical assumption, the type of state in the counterfactual scenario, ($\neg$C+P), by the mere fact of being painful, *is* identical to the actual type of state (C+P). It follows that if "pain" is a rigidified term, as is assumed in this case, it *does* apply to the counterfactual type of state $\neg$C+P as well. Thus, on Kripke's view, the counterfactual situation is not a case in which a painful state

is not called "pain", but rather a case in which a painful state *is* called "pain". Contrary to what Lewis claims. And this is perfectly acceptable.

Indeed, if we take Lewis's metaphysics on board, i.e., if we take being painful to be contingent, and being C to be essential, then the type of state in the counterfactual situation, ¬C+P, is rendered distinct from the actual type of state, C+P, and thus a rigidified "pain" cannot designate it. So Lewis's contention is in line with his own metaphysics, but not with Kripke's. Accordingly, the unacceptable outcome follows from Lewis's scenario only for a view that accepts *his* metaphysics rather than Kripke's. In other words, the view attacked by Lewis is not Kripke's view, but rather a hybrid view, which is composed of Kripke's semantics, namely, the (*de jure*) rigidity of "pain", together with Lewis's metaphysics, namely, the contingency of being painful.

Once again, revealing the intensions of "pain" that are implied by Kripke's view and by the view that Lewis attacks will clarify the difference between the two views. Both Kripke's "pain" and the "pain" in the view attacked by Lewis are rigidified terms, i.e., they designate, with respect to every possible world, that which is identical to the actual referent, namely, to the C+P type of state. However, since Kripke takes being painful, P, to be necessary and sufficient to its bearer, it follows that being the same as C+P is simply being P; by contrast, on the view that Lewis attacks, being the same as C+P amounts to being C, i.e., this view (implicitly) takes C to be necessary and sufficient (and P to be contingent). Thus, with respect to the same types of possible worlds as before, Kripke's "pain" designates the types of states in worlds W_1 and W_2; whereas "pain" on the view that Lewis attacks designates the types of states in worlds W_1 and W_3. The intensions of "pain" according to the two views are thus:

	Semantics	Metaphysics	W_1 C+P	W_2 ¬C+P	W_3 C+¬P	W_4 ¬C+¬P
Kripke's view		P is essential;	+	+	-	-
Straw-man Kripkeanism attacked by Lewis	"Pain" is rigidified	P is contingent; C is essential	+	-	+	-

Table 5.3

Again, the intensions of the two views clearly differ; they have different designations in W_2 and W_3.

Lewis's argument against Kripke is based on the alleged fact that "pain" fails to designate the type of state in w_2 (bottom row), i.e., the painful ¬C+P type of state. Yet, as Table 5.3 clearly demonstrates, Kripke's view in fact entails that "pain" does designate this type of state (upper row).

To recap, the position that yields the unacceptable result is a hybrid position that has the semantic element of Kripke's view, yet the metaphysical element of Lewis's functionalist view. Ultimately, it is not Kripke's view. It is, again, a straw-man view. Kripke's genuine view, as it is, was not shown to yield any unacceptable circumstances, and thus remains unharmed by Lewis's argument. Thus Kripke may consistently accept Lewis's claim that the straw-man view leads to unacceptable consequences and therefore has to be discarded, and at the same time cling to his own view.

5.1.2.5   Overall comparison

It might have struck you already that the two straw-man views attacked by the two arguments are in fact one and the same view; namely, it is a view composed of a semantic element that takes "pain" to be rigidified, as Kripke believes, combined with a metaphysical element that takes being painful to be contingent, as Lewis believes. As shown above, this view has the following intension:

	Semantics	Metaphysics	W_1 C+P	W_2 ¬C+P	W_3 C+¬P	W_4 ¬C+¬P
Straw-man functionalism, as well as… Straw-man Kripkeanism	"Pain" is rigidified	P is contingent; C is essential and sufficient	+	-	+	-

Table 5.4

Both arguments rightly indicate that this view carries unacceptable consequences, and hence that it is unsustainable. However, each argument points to different unacceptable consequences. Specifically, Kripke's argument highlights the fact that on that view, "pain" designates the type of state in W_3, (C+¬P), which entails that "pain" designates a non-painful type of state; Lewis, in his turn, points to the fact that on this view "pain" does not designate the type of state in W_2, (¬C+P), which entails that "pain" does not designate a painful type of state. Indeed, both these consequences are disturbing and thus strongly support rejection of that view.

So much for the weakness of the attacked view. Wherein lies the strength of the original genuine views held by Kripke and by Lewis?

To avoid Lewis's type of objection, a view has to entail that it would not be possible for "pain" not to designate a painful state, or, in short, a view has to entail that "pain" designates *all* painful states. To avoid Kripke's type of objection, a view has to entail that it would not be possible for "pain" to designate a non-painful state, or in short, "pain" has to designate *only* painful states. Thus, to avoid both objections, a view simply has to entail that "pain" designates *all and only* painful states. Technically speaking then, a resistant view should thus have the following intension:

	W_1 C + P	W_2 ¬C + P	W_3 C + ¬P	W_4 ¬C + ¬P
Resistant view	+	+	-	-

Table 5.5

But looking back at our analysis, it turns out that both Lewis's and Kripke's genuine views entail just this very same intension of "pain"!

	Semantics	**Metaphysics**	W_1 C + P	W_2 ¬C + P	W_3 C + ¬P	W_4 ¬C + ¬P
Lewis	Descriptive	Being painful is contingent	+	+	-	-
Kripke	Rigidified	Being painful is essential	+	+	-	-

Table 5.6

Thus by sharing this very same intension, Lewis and Kripke's views are equally resistant to the objections considered.

This is rather interesting as Lewis and Kripke are completely opposed with respect to both their semantical and metaphysical elements, and yet they yield the same intension of "pain" (as Table 5.6 shows). Generally, it thus follows that two views with different semantic and metaphysical elements may nevertheless entail the same intension. In other words, although the intension of a term in a theory is determined by the theory's semantic and metaphysical elements,[24] these elements are nonetheless underdetermined by the intension. In our case, this common intension of "pain" is one that designates all and only painful states.

However, there is nonetheless some indication of the difference between the two views. This difference is manifested in their different implications with regard to the

rigidity of "pain", and, accordingly, with regard to the modal status of the identity statement "pain = C". Since the functionalist takes being painful to be contingent whereas being C to be essential (and, indeed, sufficient), it turns out that the referents of "pain" across possible worlds, namely, P+C and P+¬C, are distinct, and hence that "pain" is nonrigid, and that the identity statement "pain = C" is accordingly, if true, contingent. By contrast, since Kripke takes being painful to be necessary and sufficient, it turns out that these same referents *are* identical, and hence that "pain" *is* a rigid designator, and the identity statement "pain = C" is, if true, necessary:

	Semantics	Metaphysics	W_1 C+P	W_2 ¬C+P	W_3 C+¬P	W_4 ¬C+¬P	Rigidity of "pain"
Kripke	Non-descriptive	Being painful is essential	+	+	-	-	Rigid
Lewis	Descriptive	Being painful is contingent	+	+	-	-	Nonrigid

Table 5.7

5.1.3   Conclusion

To sum up, we can look at the debate in the following way. It seems to be agreed on both sides that a good theory should have the consequence that the term "pain" designates all and only painful states, that is, in the actual world as well as in all possible worlds. In terms of intensions then, a desired view is one that entails the following intension of "pain":

W_1 C + P	W_2 ¬C + P	W_3 C + ¬P	W_4 ¬C + ¬P
+	+	-	-

Table 5.8

Lewis's view and Kripke's view differ both in their semantic and in their metaphysical elements. Nevertheless, both Kripke's and Lewis's views entail the above intension of "pain". So the criterion of complying with this intension underdetermines which of their theories is right.

The view that both Kripke and Lewis attack has a different intension:

W_1	W_2	W_3	W_4
C + P	¬C + P	C + ¬P	¬C + ¬P
+	-	+	-

Table 5.9

It allows both that a non-painful type of state is pain (W_3), and that a painful type of state is not pain (W_2). Kripke attacks this view for the first consequence; Lewis for the second. Each rightly concludes that this view is wrong. However, as the attacked view is neither Lewis's nor Kripke's, their own views remain unharmed by the attacks.

The intension that is implied by a certain view is the result of the combination of the semantics and metaphysics assumed by a view. The three views – Lewis's functionalism, Kripke's, and the attacked view –all differ in these assumptions. Both Lewis's and Kripke's views entail the same intension. It turns out then that although intension results from the combination of a semantic element and a metaphysical element, nonetheless different elements may yield the same intension. The difference between Kripke and Lewis's views is nonetheless manifested in the different implications of the views with respect to the rigidity of "pain".

The following table summarises this analysis:

	Semantics	Metaphysics	W_1 C+P	W_2 ¬C+P	W_3 C+¬P	W_4 ¬C+¬P	Rigidity
Kripke	Non-descriptive	Being painful is essential	+	+	-	-	Rigid
Lewis	Descriptive	Being painful is contingent	+	+	-	-	Non-rigid
Pseudo-Lewis and Pseudo-Kripke	Non-descriptive	Being painful is contingent	+	- Unacceptable because a painful state is not called "pain"	+ Unacceptable because a non-painful state is called "pain"	-	Rigid

Table 5.10

5.2 Donnellan's Referential-Attributive Distinction: An Argument against the Standard Interpretation and a Proposed Two Dimensionalist Analysis

Introduction

I discuss Donnellan's distinction between attributive and referential uses of definite descriptions.[25] I put forward two theses. The negative thesis is a challenge to the standard interpretation of Donnellan's distinction in terms of general vs. singular propositions. The positive thesis offers a two-dimensionalist analysis of the distinction and of the ways in which it may be interpreted.

The standard interpretation of Donnellan's distinction between attributive and referential uses of definite descriptions is in terms of general vs. singular propositions. In particular, on this interpretation, when "the F" is used *attributively* in a sentence of the form "the F is G", the sentence communicates a general, object-independent, proposition, [the x: Fx](Gx); whereas when "the F" is used *referentially* in a sentence "the F is G", the sentence communicates a singular, object-dependent proposition, (Ga). (It is widely agreed that an attributively used definite description not only *communicates*, i.e., pragmatically, a general proposition, but that it also *expresses*, i.e., semantically, that proposition. The controversy over the semantic significance of the distinction is about whether a referentially used definite description, in addition to communicating, i.e., pragmatically, a singular proposition, can also express, i.e., semantically, that singular proposition.) Given the context of the Russell-Strawson debate over the semantics of definite descriptions,[26] this interpretation seems highly plausible. Kripke objected to two other interpretations of Donnellan's distinction, namely, in terms of rigidity, and in terms of *de re/de dicto*.[27] It will be argued that Kripke's arguments are as effective against the above standard interpretation, and hence pose a real challenge to it. This extension of Kripke's arguments is based on introducing possible-worlds analysis.

Next, it is proposed that things become even clearer if we employ (in the spirit of two-dimensionalism) the distinction between worlds considered as actual and worlds considered as counterfactual. Analysing Donnellan's referential/attributive distinction in this light reveals that Donnellan's initial characterisation of his distinction is compatible with (at least) three different interpretations of it (one of which is the above standard

interpretation). It thus follows that Donnellan's characterisation of his distinction leaves it underdetermined which of the three interpretations is the correct one. I thus conclude that Donnellan's distinction is not well-defined. In addition, such analysis provides a clear understanding of the "mechanism" by which Kripke's argument operates.

5.2.1 Referential/attributive – Donnellan's cases

Donnellan introduces two fundamentally distinct cases in which the following sentence that includes the definite description "Smith's murderer" is asserted:

(1) Smith's murderer is insane.

The first case takes place in front of Smith's brutally mutilated body. Looking at the body, someone who does not know who the murderer is, exclaims sentence (1). The second case takes place on Jones's trial for Smith's murder. In this case, watching the rather odd behaviour of Jones, someone exclaims (1) again; it so happens that Jones did not in fact murder Smith. In the first case, says Donnellan, the speaker "states something about whoever or whatever is the so-and-so"; this is an *attributive* use of the definite description "Smith's murderer". In the second case the speaker "uses the description to enable his audience to pick out whom or what he is talking about and states something about that person or thing"; and this is a *referential* use of the definite description.

5.2.2 The Russell–Strawson debate and the standard interpretation

Donnellan's explicit aim in introducing his distinction was to contribute to the debate between Russell and Strawson over the semantics of definite descriptions. On Russell's account, definite descriptions of the form "the F" are to be analysed as quantificational expressions [the x: Fx], and hence they contribute to express object-independent, *general propositions* of the form [the x: Fx](Gx).[28] In other words, on Russell's view, the semantic value of a definite description is a rule for "picking out" the referent (namely, fitting the descriptive content of "the F"); it is not a concrete individual. Strawson objected to Russell's quantificational account. He claimed that there are cases, such as incomplete definite descriptions like "the table is covered with books", to which Russell's analysis fails to provide an adequate account. (Russell's account suggests a table that uniquely satisfies the definite description "the table", which is clearly not the case here.)[29] In these cases, the sentence seems to express an object-dependent, *singular proposition* (Ga); viz., in

such cases the semantic value of the definite description *is* a concrete individual. If this is sound, then such cases serve as counterexamples to Russell's quantificational account, which is thereby refuted. It is in the context of this debate that Donnellan proposed his distinction between the two uses of definite description: the attributive and the referential. The proposal was that Russell's analysis captures the attributive use but not the referential one, whereas Strawson's remarks apply to the referential use but not to the attributive one.

Given this context, it is but natural that the standard interpretation of Donnellan's distinction was formed along the following lines:[30]

Attributive use: If a definite description "the F" is used attributively in a sentence "the F is G" then the sentence communicates an object-independent, general, proposition.

Referential use: If a definite description "the F" is used referentially in a sentence "the F is G" then the sentence communicates an object-dependent, singular, proposition.

An ongoing debate concerning this distinction is about whether in addition to being able to *communicate* general and singular propositions, definite descriptions can also *express*, i.e., *semantically*, the two types of propositions.[31] Or, in short, whether in addition to being pragmatic, it is also semantic or not. In particular, it is controversial whether a definite description can semantically *express* a singular proposition, that is, whether the semantic value of a definite description may, in some cases, be a concrete individual.[32] Reimer puts it very clearly:

> It is uncontroversial that definite descriptions can be used either referentially or attributively – that they can be used to *communicate* either singular or general propositions. It is also (relatively) uncontroversial that, when used attributively, the proposition literally expressed by the sentence uttered is a general proposition. What *is* controversial is the claim that, when used referentially, the proposition *literally expressed* by the sentence uttered is a *singular* (rather than general) proposition.[33]

One outcome of interpreting referential definite descriptions in terms of contributing to singular propositions is that it allows, in principle, a definite description "the F" to apply to individuals that fail to F (as in fact happens in Donnellan's trial case. Note however that although the referent of a referentially used description *may* fail to fit the description, it, of course, *need* not do so). Some, understandably, find it highly problematic. E.g., Wiggins says,

> Donnellan's [account] … depends on the, for me, incredible idea that if I say 'the man drinking champagne is [G]' and the man I mean, although drinking water, is [G], then what I say is true.[34]

In particular, many of those who endorse the semantic significance thesis, i.e., that definite descriptions may have a referential *semantic* function, feel uncomfortable with this feature, and thus make the additional provision that the referent of a referentially used "the F" *is* indeed F. I.e., that in addition to being the object intended by the speaker, it *also* has to fit the descriptive content of "the F".[35] This clearly deviates from Donnellan's original definition of referential definite description, which explicitly allows such terms to designate an intended object even if that object fails to fit the description: "… using a definite description referentially, a speaker may say something true even though the description applies to nothing".

5.2.3 Kripke's rejection of the *de re/de dicto* interpretation and of the rigidity interpretation

Kripke's "Speaker Reference and Semantic Reference"[36] is best known for arguing against the semantic significance of Donnellan's distinction. The argument is inspired by Grice's ideas about *conversational implicature*.[37] In holding that "what … a speaker has *said* may be false, what he *meant* may be true",[38] Grice draws a distinction between what a speaker says and what a speaker means. In a similar spirit, Kripke argues that although in the trial case the *speaker reference* may be Jones (i.e., what is *meant*), nevertheless this is not the *semantic reference* (i.e., what is *said*); the semantic reference of "Smith's murderer" is just Smith's murderer, regardless of how the speaker chooses to use the term. Most anti-ambiguity arguments follow a similar line of reasoning.[39] However, before offering his anti-ambiguity argument, Kripke dedicates a few passages to rejecting a couple of interpretations of the attributive/referential distinction. This part, I will argue, implies a serious challenge to the standard interpretation in terms of expressing singular/general propositions – a challenge which I suspect was largely ignored by those engaged in the semantic debate over Donnellan's distinction.

The two interpretations of the attributive/referential distinction that Kripke rejects are, (a) in terms of *de re/de dicto*; and (b) in terms of rigidity. In particular, Kripke argues against the identification of attributively used definite descriptions with *de dicto* definite description, and also against the identification of attributively used definite descriptions with nonrigid definite description (by which, I believe, he means non-rigid*ified* definite description. I'll explain this soon). In each argument, Kripke provides a

powerful counterexample to make his case. Let me briefly introduce these counterexamples.

To argue against identifying attributively used definite descriptions with *de dicto* definite description, Kripke provides the following example:[40]

> (2) The police know concerning Smith's murderer, whoever he is, that he committed the murder; but they're not saying who he is.

The definite description "Smith's murderer" carries Donnellan's mark of the attributive: it has the "whoever is the so-and-so" mark. Thus, Kripke concludes that it is used attributively (by this Kripke assumes that the "whoever is the so-and-so" mark is not only necessary for the attributive use, but is in fact also sufficient for it). Nevertheless, claims Kripke, the description in this context is also *de re*. I take it that this amounts to saying that (2) should be analysed as follows:[41]

> (2') There is someone such that, this someone murdered Smith, and the police know *of that someone* – whoever he is - that he committed the murder; but they're not saying who he is.

In other words, the existential quantifier has a wide scope relative to the propositional attitude operator "know";[42] so we have a case of an attributively used definite description which is also *de re*. It thus follows that attributively used definite descriptions cannot be identified with *de dicto* definite descriptions; and, accordingly, that referentially used definite descriptions cannot be identified with *de re* definite descriptions (since the above example shows that some *de re* definite descriptions are used attributively).[43]

Let us look now at Kripke's argument against the interpretation of Donnellan's distinction in terms of rigidity.

The argument is based on objecting to the identification of attributively used definite descriptions with nonrigid definite descriptions. Kripke provides the following example:[44]

> (3) That bastard – the man who killed Smith, whoever he may be – is surely insane.

Here, Kripke, in Kaplan's style, uses "that" as a rigidifying device (it would be more explicit to use Kaplan's "dthat" – see Section 2.2.3), to ensure the rigid designation of the definite description. (Those who take demonstratives to require an act of demonstration, i.e., ostension, may object here that since there is no ostension involved in (3), the "that"

device is not a demonstrative one, and hence cannot be said to rigidify the definite description. However, this can be easily remedied by using another rigidifying device instead, namely, actualisation:

(3') The *actual* murderer of Smith – whoever he is – is surely insane.)

As in (2), according to Donnellan's mark of the "whoever is the so-and-so" clause, the definite description in (3) is used attributively (again, this requires Kripke's assumption above that the "whoever is the so-and-so" is *sufficient* for an attributive use). Yet it is also *de jure* rigid, i.e., rigidified; it is designed to designate the same individual (namely, the one who murdered Smith in the actual world) in all counterfactual worlds (at least in the case of (3') if not in (3)). It thus follows that an attributively used definite description cannot be identified with a definite description which is non *de jure* rigid (i.e., not rigidified[45]). Correspondingly, a referentially used definite description cannot be identified with a *de jure* rigid designator, since some *de jure* rigid designators are used attributively.

5.2.4   Extending Kripke's argument - against the standard interpretation

I would like to argue that the above arguments can be extended to apply to the standard interpretation as well. Thus, I will attempt to show that if Kripke's (3) (or, more clearly, our (3')) undermines the rigidity interpretation, then it equally undermines the standard interpretation in terms of general vs. singular propositions. And the same goes for (2) (and more clearly for (2')). My argument will rely on possible-worlds analysis.

In its possible-worlds version, the Russellian account of a definite description takes it to designate, with respect to each counterfactual world, that which fits the description in that world. Thus such a definite description may (and typically does) designate different referents in different counterfactual worlds. Hence the generality of the proposition. By contrast, a definite description that contributes to a singular proposition is such that it is designed to designate, with respect to every counterfactual world, the same referent that it designates in the actual world. Hence the singularity of the proposition.

In other words, a definite description that contributes to a general proposition is just a non-rigidified (i.e., descriptive) definite description – namely, it designates, with respect to each counterfactual world, that which fits the description in that world (whether this referent is identical to the referent in the actual world or not). Whereas a

definite description that contributes to a singular proposition is just a rigidified (i.e., *de jure* rigid) definite description – it is designed to designate, with respect to every counterfactual world, that which it designates in the actual world (whether it also fits the description in that world or not). This assimilation of definite descriptions that contribute to singular propositions with rigidified definite descriptions is nicely expressed in Wettstein (1991); talking about the referential use of "Smith's murderer" in Donnellan's trial case, he says:

> I am committed to the view that the singular proposition determined in the context in question is true in a possible world just in case Jones, the actual murderer, is insane in that world.[46]

Now (3') above purports to show that attributively used definite descriptions cannot be identified with rigidified definite descriptions. Given the present assimilation of rigidified definite descriptions with definite descriptions that contribute to singular propositions, it follows that attributively used definite descriptions cannot be identified with definite descriptions that contribute to a general proposition. Indeed, (3') seems to be just a case of a sentence that expresses a singular proposition, and which at the same time also involves an attributively used definite description. So if (3') suffices to refute the rigidity interpretation of Donnellan's distinction, it thereby suffices to refute the standard interpretation of that distinction.

I believe that similar considerations apply to (2). The intensional context that induces the *de re*/*de dicto* ambiguity in (2) is the propositional attitude 'know that'. Let us induce a similar ambiguity using a modal context instead of a propositional attitude context:[47]

(4) Necessarily, Smith's murderer (whoever that is) is insane.

This sentence is ambiguous between a *de re* reading, on which (4) amounts to

(4') There is someone who is Smith's murderer (whoever that is), and it is necessary that *that* someone is insane.

and a *de dicto* reading, on which (4) amounts to

(4") It is necessary that, someone murdered Smith's and (whoever that someone is) that someone is insane.

In (4') the existential quantifier has a wide scope relative to the necessity operator, whereas in (4') the existential quantifier has a narrow scope relative to the necessity

operator. The necessity operator suggests an easy reformulation of the two readings in terms of possible worlds. Thus on the *de dicto* reading, (4) expresses the proposition,

> (4*) With respect to every counterfactual world, the person who murdered Smith *in the actual world* (whoever that is) is insane in that counterfactual world.

whereas on the *de dicto* reading, (4) expresses the proposition

> (4**) With respect to each counterfactual world, the person who murdered Smith *in that counterfactual world* (whoever that is) is insane in that counterfactual world.

It is clear now that the *de re* reading of (4), namely (4*), takes it to express a singular proposition (whether this proposition is true or not).

However, (4*) has the attributive "whoever that is" mark, and hence (if this mark is indeed sufficient for being attributive,) "Smith's murderer" is used attributively. So (4*) is a case of an attributively used definite description that contributes to a singular proposition. It thus follows, again, that attributively used definite descriptions cannot be identified with definite descriptions that contribute to a general proposition.

In sum, we've discussed four distinctions that apply to definite descriptions: (a) between those that contribute to general propositions and those that contribute to singular propositions; (b) between non-rigidified ones and rigidified ones; (c) between *de re* ones and *de dicto* ones; and (d) between those used referentially and those used attributively. Using possible-worlds semantics, it was suggested that the first three distinctions amount to just one distinction, namely, between a definite description that designates, with respect to each counterfactual world, that which fits the description in that world, and between a definite description that is designed to designate, with respect to every counterfactual world, that which it designates in the actual world (respectively):[48]

Definite descriptions that designate, with respect to each counterfactual world, that which fits the description in that world	Definite descriptions that are designed to designate, with respect to every counterfactual world, that which they designate in the actual world
Definite descriptions that contribute to general propositions	Definite descriptions that contribute to singular propositions
Non-rigidified definite descriptions	Rigidified definite descriptions
De dicto definite descriptions	*De re* definite descriptions

Table 5.11

Consequently, the *intensions* of all definite descriptions on each column are identical: constant on the left column; not necessarily constant on the right column.

Now this distinction cross-cuts the fourth distinction between attributive and referential definite descriptions (if Kripke is right about the sufficiency of the "whoever is the so-and-so" mark for the attributive use). In particular, attributively used definite descriptions may be rigidified (either by a *de re* reading, or by actualisation or by Kaplan's dthat operator, etc.), and thus express singular propositions. Hence, the attributive cannot be defined in terms of general propositions (since some attributive definite descriptions contribute to singular propositions); likewise, the referential cannot be defined in terms of singular propositions (as some singular propositions involve attributive definite descriptions):

	Definite descriptions expressing general propositions = *De dicto* definite descriptions = Non-rigidified definite descriptions	**Definite descriptions expressing singular propositions** = *De re* definite descriptions = Rigidified definite descriptions
Attributively used definite descriptions	"Smith's murderer, whoever that is, is insane" (Mutilated body case)	"Smith's actual murderer, whoever that is, is insane"
Referentially used definite descriptions	?	"Smith's murderer is insane" (Trial case)

Table 5.12

5.2.5   A novel account: referential/attributive and two-dimensional semantics

In this last section, I would like to propose a novel two-dimensionalist account for Donnellan's distinction. This account also reveals the "mechanism" by which Kripke's above arguments against the two interpretations operates.

If our above possible-worlds analysis is correct, then (2) and (3) express singular propositions. But despite being singular propositions, the "whoever he is" clause still seems to indicate *some* sense of generality. I.e., there is some sense in which "the murderer, whoever he is" *is* different from simply "Brown" (supposing that Brown is the actual murderer). True, (2) and (3) show that this generality does not apply to counterfactual worlds, as the referent is the same in all those worlds. Whence the generality then?

The answer, I suggest, lies in the distinction between possibilities of the *counterfactual* vs. possibilities of the *actual*. Thus, although the generality in (2) and (3) does not apply to possible worlds considered as *counterfactual*, it does apply to possible worlds considered as *actual*. In particular, the generality involved in "Smith's murderer (whoever that is)" corresponds to the different possibilities with respect to the *actual* murderer. If this *actual* murderer is Jones, then the definite description in (2) and (3) designates Jones in every *counterfactual* world, yet if the actual murderer is Brown, then the same definite description designates Brown in every counterfactual world, etc. In sum, the intension of (an attributive) "Smith's murderer, whoever that is", when contributing to *singular* propositions (like in (2) and in (3) above) has the following structure:

Counterfactual worlds The actual world	W_1 Smith's murderer is Jones	W_2 Smith's murderer is Brown	W_3 Smith's murderer is Taylor	...
W_1 Smith's murderer is Jones	Jones	Jones	Jones	Jones
W_2 Smith's murderer is Brown	Brown	Brown	Brown	Brown
W_3 Smith's murderer is Taylor	Taylor	Taylor	Taylor	Taylor

Table 5.13

In other words, although the definite description "Smith's murderer (whoever that is)" in sentences like (2) and (3) designates the same individual in all *counterfactual* worlds (i.e., possible worlds considered as counterfactual), nevertheless, it designates different individuals in different *actual* worlds (i.e., possible worlds considered as actual). And hence the generality aspect.

This two-dimensionalist distinction, between possibilities of the actual vs. possibilities of the counterfactual is a powerful tool. Applying it to the referential/attributive distinction will generate a novel account of the latter distinction and its interpretations. Let me demonstrate this.

I believe the following to be an adequate summary of Donnellan's initial characterisation of the referential/attributive distinction. An attributively used definite

description designates that which fits the description, regardless of whether that object is intended by the speaker or not. Such designation entails (potentially) different referents in different possible worlds. Hence the generality aspect of the attributive use. A referentially used definite description, by contrast, designates the object intended by the speaker, regardless of whether that object fits the description or not. Such designation entails the same referent in all possible worlds. Hence the singularity aspect of the referential use.

But a definite description may, in principle, pick out its referent in either of these two ways (i.e., via fitting the description, or regardless of the description) in *each* of two types of possible worlds, namely, possible worlds considered as *actual*, and possible worlds considered as *counterfactual*. The two types of designation, together with the two types of possible worlds, thus give rise to four possible semantic functions:

	Designation in possible actual worlds	**Designation in possible counterfactual worlds**	
1	By fitting a descriptive content	By fitting a descriptive content	*Descriptive*
2	By fitting a descriptive content	Regardless of descriptive content	*Reference-fixing*
3	Regardless of descriptive content	By fitting a descriptive content	---
4	Regardless of descriptive content	Regardless of descriptive content	*Directly-referring*

Table 5.14

We called the three types of terms that seem to be used in English, *descriptive*; *reference-fixing*; and *directly-referring*, as detailed in Table 5.14.[49] (The third category was cancelled-out on grounds of being un-instantiated, and hence will play no part in our analysis.)

Donnellan's original characterisation of the attributive/referential distinction is not sensitive to the distinction between the two types of possible words. Thus, whereas his characterisation points to some general vs. singular *aspect* of designation in possible worlds, it does not tell us whether these are possible worlds considered as *actual* or as *counterfactual*, or both. It is thus open to various interpretations. Let us specify these.

One such interpretation takes the distinction to apply to designation in *counterfactual* worlds. Thus, an attributively used definite description designates, with respect to every counterfactual world, that which fits the description *in that counterfactual world*, whether it is the same referent as the actual one or not (since the third category, Row 3, is cancelled-out). This category thus includes non-rigidified terms, i.e., descriptive terms alone (Row 1). A referentially used definite description, by contrast, designates,

with respect to every counterfactual world, regardless of the descriptive content, i.e., it simply designates that which it designates in the actual world, whether it fits the description in that counterfactual world or not. This category thus includes rigidified terms, namely, reference-fixing definite descriptions (Row 2) as well as directly-referring definite descriptions (Row 4).[50] This seems to be the standard interpretation. (And, if our analysis above is sound, it is also the interpretation in terms of *de re/de dicto*, and the interpretation in terms of rigidity – all of which were claimed to amount, in principle, to the same interpretation.)

In light of the above two-dimensionalist analysis, we can now better understand Kripke's arguments against the rigidity interpretation and against the *de re/de dicto* interpretation of Donnellan's distinction (which were taken here to be as effective against the standard interpretation). One mark of the attributive is the "whatever is the so-and-so" clause. This mark guarantees some generality. The standard interpretation took it to apply to counterfactual worlds. I.e., on this interpretation, the attributive is identified with a general designation in *counterfactual* worlds. Kripke construct an example in which the generality of the "whatever is the so-and-so" is applied to *actual* (rather than to counterfactual) worlds, i.e., the term designates different referents in different *actual* worlds, whereas in counterfactual worlds there is singularity, i.e., the term designates the same referent in all counterfactual worlds (as shown in Table 5.13). He does that by using a reference-fixing definite description, i.e., a rigidified definite description (e.g., in (2), (3), and (4)). Thus such a definite description is attributive since it has the "whatever is the so-and-so" (it is general with respect to different *actual* worlds), yet it is at the same time singular (with respect to *counterfactual* worlds).

Thus our analysis reveals that the attributive *is* necessarily tied to *some* generality – *either* in counterfactual worlds *or* in actual worlds. What Kripke showed is that the attributive cannot be *conclusively* identified with generality in counterfactual worlds, i.e., with contribution to general propositions.

Another interpretation may take Donnellan's distinction to apply to designation in the *actual* world. On this interpretation, an attributively used definite description designates, with respect to every *actual* world, that which fits the description in that world, whether it is intended by the speaker or not. This category will thus include descriptive terms and reference-fixing terms (Rows 1 and 2). A referentially used definite description, by contrast, designates, with respect to every *actual* world, regardless of

descriptive content, i.e., it designates that which the speaker intends, whether it fits the description in that actual world or not. (As Row 3 is cancelled out,) this category will thus include directly-referring definite descriptions only (Row 4). This interpretation may be Kripke's interpretation, at least as it emerges from his arguments presented here.[51]

Yet another interpretation may take Donnellan's distinction to apply to designation in *both* the actual world *and* in counterfactual worlds. On this interpretation, an attributively used definite description designates, with respect to every possible world (whether actual *or* counterfactual), that which fits the description in that world, whether it is intended by the speaker or not, and whether it is the same as the actual referent or not. This category will thus include only descriptive terms (Row 1). A referentially used definite description, by contrast, designates, with respect to every possible world (whether actual or counterfactual) regardless of any descriptive content. I.e., with respect to every actual world, it designates that which the speaker intends, and then, with respect to every counterfactual world, it simply designates that same actual referent. This category will thus include directly-referring definite descriptions only (Row 4).[52] Thus on this interpretation, reference-fixing definite descriptions (like in (2) and (3)) fit in neither category. It follows that on this interpretation the referential/attributive distinction turns out not to be exhaustive. I believe this interpretation is no less plausible than the previous two, yet I'm aware of no one that explicitly endorses it.

Now which of these three interpretations is the correct one?

If the present analysis is correct, this cannot be answered, as Donnellan's initial formulation of this distinction is simply too loose. Put differently, each interpretation draws an important distinction between two types of definite descriptions. But since Donnellan's initial characterisation of his distinction was not sensitive to the difference between designation in actual worlds vs. designation in counterfactual worlds, it leaves it underdetermined which of the three distinctions (drawn by the three interpretations) is the one that is equivalent to Donnellan's original distinction.

Finally, regarding each distinction drawn by the different interpretations, it is a fair question whether it is a semantic one or not, i.e., whether it corresponds to an ambiguity in definite descriptions or not. And this is regardless of whether the distinction is indeed equivalent to the referential/attributive distinction or not.

Endnotes

---

[1] Kripke (1972)

[2] Lewis (1994)

[3] Kripke and Lewis are discussing possible worlds, without specifying whether these are considered as actual or as counterfactual. It is rather clear however that their discussion is concerned with counterfactual worlds. Yet, to conform to Kripke and Lewis's original discussion, I shall be using simply "possible worlds" as well here.

[4] Armstrong (1999, p. 86), who attributes the example to Brian Medlin.

[5] "C" for the legendary C-fibre firing, of course

[6] It is very important to note that gene and pain are *types* and not *tokens*. There are many tokens of gene, as there are many tokens of pain. And so are the DNA molecule and C. Therefore, if the occupants of a certain role are all of the same type, we get a one-one correlation between the suspected *type* and the credible *type*, and hence the reduction is in fact a type-type identity. If, however, there is more than one type of entities that occupies the role, e.g., the occupant of the pain-role in human-beings is C, whereas the occupant of the same role in Martians is D, then the correlation between this suspected type and the credible types is a one-many correlation, and the reduction is thus mere supervenience between the two types. I.e., the credible type determines the suspected type, but not vice versa. In other words, the suspected type is said to be *multiply realisable* in the credible types (see Section 3.4).

[7] As was mentioned earlier (Endnote 20 in Chapter 4), it is possible, in principle, that some role should be occupied by a certain type of occupant essentially, and not merely contingently, i.e., that it might not have been occupied by another occupant instead. In other words, the mere role-occupant relation does not, as such, necessarily entail the contingency of the occupant to the role. However, it is entirely obvious that the way in which *functionalists* use the role-occupant relation does imply such contingency. Indeed, they way they use it imply the *mutual* independence of occupants and roles. Lewis's quote below, for one, expresses this feature very clearly.

[8] Kripke's original footnote:

> For example, David Armstrong (1968); see the discussion review by Thomas Nagel, Philosophical Review 79 (1970), pp. 394-403; and David Lewis, 'An Argument for the Identity Theory', The Journal of Philosophy, pp. 17-25.

[9] Kripke (1972/1980, p. 147, *italics* in original)

[10] Lewis's original footnote: "Saul Kripke, *Naming and Necessity* (Blackwell, 1980) pp. 147-8".

[11] Lewis (1994, reprinted in 1999, p. 304)

[12] Choose your favourite view of designation of rigid natural-kind terms (see Section 5.1.1.2)

[13] Indeed, *at least* descriptive. It is more likely to be even a role-term (see Section 4.2.2), i.e., to mean, "that which *contingently* occupies the pain-role". In other words, it is likely to include the metaphysical condition as part of the description. However, for the purpose of this discussion, it suffices that the term is taken to be descriptive, so we need not commit to the stronger assumption that it is taken to be a role term.

[14] Section 5.1.1.1.1.

[15] The same holds for all other *secondary substances*, as Aristotle called them, e.g., being an animal. (See Section 3.3.4.).

[16] For the distinction between *de jure* and *de facto* rigidity see Section 1.3.1.

[17] See Section 1.3.2.

[18] Recall the difference between rigid and rigidified: all rigidified terms are rigid, yet not all rigid terms are rigidified; some rigid terms are descriptive. Rigidified rigid terms are rigid by stipulation, i.e., *de jure*, whereas descriptive rigid terms are not thus stipulated, they simply turn out to be rigid, i.e., *de facto* (See Section 1.3.1).

[19] Recall that the category of rigidified terms is concerned with designation in counterfactual worlds, and not so much with designation in the actual world. Yet there are two ways in which a rigidified term can pick out its referent *in the actual world*, namely, by fitting a descriptive content ("reference-fixing" terms), or directly, i.e., regardless of any descriptive content ("directly referring" terms). (See Section 2.1.) Kripke remains agnostic on whether natural-kind terms are reference-fixing or directly-referring (see Section 2.2.4).

[20] Kripke (1972/1980, p. 148)

[21] Kripke (1971/1999, endnote 18, p. 88, my *italics*)

[22] Indeed, Kripke seems to suggest that being painful is what pain in fact consists in. I believe it is appropriate to call *such* a necessary and sufficient property, i.e., one that specifies what an object consist in, a *sortal* property (see Section 3.3.5). Some necessary and sufficient properties don't specify what the object consists in; e.g., the property 'being the sum of 2+3' doesn't. However, it suffices for the purpose of the present argument that Kripke takes being painful to be necessary and sufficient, whether he also takes it to be a sortal property or not.

[23] For the sake of simplicity, let us ignore here worlds that contain more than one of these types, e.g., worlds in which there are both painful Cs, and non-painful Ds (as we did in all other examples throughout the dissertation. E.g., see Section 1.1.1 for example). The way to

accommodate such worlds in our analysis would be to replace the "+" and "-" signs in the table that apply to the whole world, with a specification of what is designated by the term in that world.

[24] As well as, of course, by the actual state of the world (as this thesis endeavours to maintain). However, as this factor is not at issue in this debate, we ignore it throughout.

[25] Donnellan (1966)

[26] Russell (1905; 1919); Strawson's attack (1950); Russell's rebuttal (1957).

[27] Kripke (1977)

[28] Russell (1905). Russell specific account of such a proposition is as the conjunction of the following three conditions: (1) there is at least one F (existence); (2) there is at most one F (uniqueness); and, (3) everything that is F is G (predication).

[29] Strawson (1950)

[30] E.g., Kaplan (1978); Wettstein (1981; 1991a; 1991b); Salmon (1982; 1991); Reimer (1989a; 1998b); Neale (1990); Nelson (1999); Feit (2003); Bontly (2005)

[31] Donnellan's own view on this matter is somewhat unclear. On the one hand, he explicitly states that the distinction is not semantic: "It does not seem at all attractive to suppose an ambiguity in the meaning of the words; it does not appear to be semantically ambiguous." (1966, p. 59); yet on the other hand, much of what he in fact says in his paper suggests to the contrary.

[32] Interestingly, there is an equivalent debate about indefinite descriptions; notably, Chastain (1975) and Devitt (2004) think that there is a referential/attributive ambiguity in this case, whereas Ludlow and Neale (1990) deny such ambiguity. See also King (1988).

[33] Reimer (1998a, p. 92)

[34] Wiggins (1975, p. 28, endnote 9)

[35] E.g., Reimer (1998) and Devitt (2004). Kaplan (1978) distinguishes his *demonstrative* use, from Donnellan's *referential* use just on this aspect:

> Donnellan and I disagree on how to bring the intended demonstrating into the picture. To put it crudely, Donnellan believes that for most purposes we should take the demonstratum to be the intended demonstratum. I believe that these are different notions that may well involve different objects.

Exceptions to this provision are Larson and Segal (1995, pp. 213, 340-1), who argue that Donnellan's referential use is semantic even without the additional proviso, i.e., that a definite description can semantically designate an object that in fact fails to fit the description.

[36] Kripke (1977)

[37] Grice (1967/1975)

[38] Grice (1969, p. 142)

[39] E.g., Salmon (1982, 1991), Bach (1987), and Neale (1990)

[40] Kripke (1977, p. 9-10)

[41] For a discussion of the *de re*/*de dicto* ambiguity, and its scope analysis, see Section 2.2.2.1.2.

[42] An (unreasonable) *de dicto* reading of this sentence would be,

> (2") the police know that, there is someone such that that individual is Smith's murderer
> – whoever that is, but they're not saying who that individual is.

This reading makes little sense, as it renders the use of "but" here inconsistent. I.e., since on this reading the police's knowledge does not involve a particular individual, they anyway can't tell who that individual is.

[43] A similar point is made by Kaplan (1978): "It is now clear that I can assert of the first child to be born in the 21st century that he will be bald". Thus, although the definite description "the first child to be born in the 21st century" refers, in 1978, to 'whoever is the so-and-so', it is of a particular child that it is asserted that he will be bald, i.e., it is a *de re* attribution.

[44] Kripke (1977, p. 11)

[45] The relevant candidate for defining attributive definite descriptions is not simply being a nonrigid definite description, but rather being a non-rigid*ified* definite description. I.e., a definite description that is not *stipulated* to be rigid. The two are not identical. A non-rigidified definite description can nonetheless be (*de facto*) rigid. E.g., "the sum of 2 and 3"; the term is not stipulated to be rigid by a rigidifying mechanism, yet it nonetheless *turns out* to be rigid. For a discussion of the distinction between rigid and rigidified (as well as between nonrigid and non-rigidified) see Section 1.3.1.

[46] Wettstein (1991, p. 57)

[47] Whereas possible-worlds semantics copes well with intensional contexts such as tense, necessity, indexicals and even obligation, propositional attitudes pose some well-known problems to such analysis. In particular, they violate substitution of co-intensional terms. E.g., while Smith may know that 17*7 is not a prime, he may not know that 119 is not a prime. (See Section ii in the Introduction.)

[48] In the *de re/de dicto* case, this applies at least to cases in which the definite description is embedded in modal context. It is less clear whether such an account applies to other intensional contexts, such as knowledge, as in (2). See Endnote 47.

[49] See Chapter 2, in particular Section 2.1, for an elaborate discussion of these semantic functions.

[50] Those who add the proviso that the actual referent has to fit the description thereby rule out directly-referring definite descriptions, and limit this category to reference-fixing definite descriptions alone. This, however, violates Donnellan's explicit characterisation of the referential (see end of Section 5.2.2.).

[51] Kripke claims that (2) and (3) are cases of attributively used definite descriptions; we have shown that in these cases, the definite description is reference-fixing, i.e., it designates, with respect to the actual world that which fits the description, and then designates, with respect to every counterfactual world, that same individual, whether it fits the description in that world or not (Row 2). Yet surely an attributively used definite description *may* also be descriptive (Row 1), and thus contribute to a general proposition (e.g. Donnellan's mutilated body case). It thus follows that the attributive use includes *both* reference-fixing definite descriptions and descriptive definite descriptions. This leaves referentially used definite descriptions the third available category, i.e., that of directly-referring terms (Row 4). However, this reasoning is not enough to conclusively ascribe to Kripke the above interpretation, even on the mere basis of his above arguments.

It should also be noted that Kripke's (1977, p.15) account of the referential use includes the additional feature that the speaker *believes* that the intended referent fits the description. However, this requirement surely cannot be true, as it is at odds with Donnellan's (1966) explicit rejection of such a requirement: "it is possible for a definite description to be used referentially where the speaker believes that nothing fits the description."

[52] Note that in addition there is a combined option, i.e., a definite description may designate, with respect to each possible world, that which is *both* intended by the speaker, *and* also fits the description, in that world. This is in fact Devitt's (2004) explicit view about referential definite descriptions:

> [A referential] 'The murder is insane' is equivalent either to 'That is a murderer and insane' or to Lepore and Ludwig's '[the x: x = that and x is a murderer](x is insane)'.

However, again, that would violate Donnellan's own characterisation of the referential, which explicitly allows the referent of a referentially used definite description to fail to fit the description.

REFERENCES

- Adams, R. Merrihew, (1979) "Primitive Thisness and Primitive Identity" *Journal of Philosophy*, 76: 5-26

- Anderson, C. A. (1984) "General intensional logic" In D. Gabbay and F. Guenthner (eds.), *Handbook of Philosophical Logic*, D. Reidel, Volume II, Chapter II.7, pp. 355–385

- Aristotle, *Categories.* Translated with notes by J. L. Ackrill (1963), Oxford: Clarendon Press

- Armstrong, D., (1968) *A Materialistic Theory of the Mind.* London: RKP

- ____, (1989) *Universals*, Boulder, CO: Westview Press

- ____, (1997) *A World of States of Affairs*, Cambridge: Cambridge University Press

- ____, (1999) *The Mind Body Problem: An Opinionated Introduction* Oxford: Westview Press

- ____, (2004) *Truth and Truthmakers*, Cambridge: Cambridge University Press

- Ayer, A. J., (1936) *Language, Truth, and Logic*, London: Gollancz

- Ayers, Michael, (1990) *Locke*, vol. 2, London: Routledge

- Bach, Kent, (1987) *Thought and Reference*, Oxford: Clarendon Press

- ____, (1992a) "Intentions and Demonstrations", *Analysis* 52: 140-146

- ____, (1992b) "Paving the Road to Reference", *Philosophical Studies* 67: 295-300

- ____, (online article) "Comparing Frege and Russell", http://userwww.sfsu.edu/~kbach/FregeRus.html

- Barcan Marcus, Ruth, (1961) "Modalities and Intensional Languages", *Synthese,* 13: 303-22

- Bayart, A., (1958) "Correction de la Logique Modale du Premier et du Second Ordre *S*5 [Soundness of First and Second Order *S*5 Modal Logic]", *Logique et Analyse* 1: 28–44

- ____, (1959) "Quasi-Ad´equation de la Logique Modale du Second Ordre *S*5 et Ad´equation de la Logique Modale du Premier Ordre *S*5 [Quasi-Completeness of Second-Order *S*5 Modal Logic and Completeness of First Order *S*5 Modal Logic]", *Logique et Analyse* 2: 99–121

- Bickle, John (1998) *Psychoneural Reduction: The New Wave.* Cambridge, MA: MIT Press

- Borg, Emma, (2000) "Complex Demonstratives", *Philosophical Studies* 97: 229-249

- Braun, David, (1994) "Structured Characters and Complex Demonstratives" *Philosophical Studies* 74: 193-219

- ____, (2001) "Indexicals" in *The Stanford Encyclopedia of Philosophy* Edward N. Zalta (ed.), URL = http://www.seop.leeds.ac.uk/entries/indexicals

- Bressan, A., (1972) *A General Interpreted Modal Calculus*, New Haven: Yale University Press

- Broad, C. D. (1933) *Examination of McTaggart's Philosophy: Volume I.* Cambridge: Cambridge University Press

- Brock, Stuart. (2004) "The Ubiquitous Problem of Empty Names," *Journal of Philosophy*, 101: 277–98

- Burke, Michael, (1994) "Preserving the Principle of One Object to a Place: A Novel Account of the Relations among Objects, Sorts, Sortals and Persistence Conditions", *Philosophy and Phenomenological Research*, 54: 591-624, Reprinted in Rea, M., (ed.) (1997), *Material Constitution*, Lanham MD: Rowman and Littlefield: 236-269

- Caplan, Ben (2005) "Against Widescopism" *Philosophical Studies*, 125: 167–190

- Carnap, Rudolf, (1946) "Modalities and Quantification" *Journal of Symbolic Logic* 11: 33–64

- ____, (1947) *Meaning and Necessity*, Chicago: University of Chicago Press, Enlarged edition 1956

- Carter, William. R., (1989) "How to Change Your Mind", *Canadian Journal of Philosophy* 19: 1-14

- Chalmers, David J. (1996) *The Conscious Mind.* Oxford: Oxford University Press

- ____, (2002) "On Sense and Intension", *Philosophical Perspectives* 16:135–182

- ____, (2006) "Two-Dimensional Semantics" in Lepore E., and Smith B., (eds.), *The Oxford Handbook of Philosophy of Language* Oxford: Oxford University Press: 574-606

- Chandler, Hugh S., (1976) "Plantinga on the Contingently Possible" *Analysis*, 36: 106-109

- Chastain, Charles, (1975) "Reference and Context" in K. Gunderson (ed.), *Minnesota Studies in the Philosophy of Science, Volume VII: Language, Mind, and Knowledge* Minneapolis: University of Minnesota Press: 194-269

- Chisholm, Roderick M., (1967) "Identity through Possible Worlds: Some Questions", *Noûs*, 1: 1–8

- ____, (1973) "Parts as Essential to their Wholes" *Review of Metaphysics*, 26: 581-603

- ____, (1976) *Person and Object: A Metaphysical Study.* George Allen & Unwin Ltd., London, ch.3

- Church, Alonzo, (1944) *Introduction to Mathematical Logic, Part I.* Princeton University Press. Revised and enlarged, 1956

- ____, (1946) "A formulation of the logic of sense and denotation (abstract)". *The Journal of Symbolic Logic XI*, 31

- ____, (1951) "A formulation of the logic of sense and denotation" In P. Henle (Ed.), *Structure, Method and Meaning.* New York: The Liberal Arts Press, pp 3–24

- Cocchiarella, Nino (1986) *Logical Investigations of Predication Theory and the Problem of Universals.* Napoli: Bibliopolis

- Cook, Monte (1980) "If 'Cat' is a Rigid Designator, What Does it Designate?" *Philosophical Studies*, 37: 61–4

- Copeland, B. Jack, (2006) "Meredith, Prior, and the History of Possible Worlds Semantics" *Synthese* 150: 373–397

- Cresswell, Max J., (1985) *Structured Meanings*, MIT Press, Cambridge, MA

- Davidson, Donald, (1970) "Mental Events" in D. Davidson, (ed.) *Essays on Actions and Events.* Oxford: Clarendon Press (1980): 207-225

- Davies, Martin. & Humberstone, I.L (1981) "Two notions of necessity" *Philosophical Studies* 58: 1-30

- Della Rocca, Michael (2002) "Essentialism versus Essentialism" in Tamar Gendler and John Hawthorne (eds.): *Conceivability and Possibility.* New York: Oxford University Press, pp. 223-252

- Devitt, Michael (1981a) *Designation*, New York: Columbia University Press

- ____, (1981b) "Donnellan's Distinction" *Midwest Studies in Philosophy* 6: 511-524

- ____, (2004) "The Case for Referential Descriptions" in Bezuidenhout, A., and Reimer, M., eds., *Descriptions and Beyond*, Oxford: Oxford University Press, pp. 280-305

- ____, (2005) "Rigid Application" *Philosophical Studies* 125: 139 –165

- Devitt, Michael, and Sterelny, Kim (1999) *Language and Reality: An Introduction to the Philosophy of Language*, 2nd edition. Cambridge: MIT Press

- Donagan, Alan (1963) "Universals and Metaphysical Realism" *Monist* 47: 211–246

- Donnellan, Keith, (1966) "Reference and Definite Descriptions", *Philosophical Review* 75: 281-304, reprinted in A. P. Martinich (ed.), The Philosophy of Language (Oxford: Oxford University Press, 2001), pp. 247-259

- Dummett, Michael, (1973) *Frege: Philosophy of Language*, New York: Harper and Raw

- _____, (1981) *The Interpretation of Frege's Philosophy*, London: Duckworth

- Evans, Gareth, (1977) "Reference and Contingency" *The Monist* 62: 161-89

- Everett, Anthony (2005) "Recent Defences of Descriptivism," *Mind & Language*, 20: 103–39

- Feldman, Fred, (1992) *Confrontations with the Reaper*, New York: Oxford University Press

- Fine, Kit, (1994) "Essence and Modality" in James Tomberlin, ed., *Philosophical Perspectives, 8: Logic and Language*, (Atascadero, CA: Ridgeview Publishing), pp. 1-16

- Fitting, Melvin, (2006) "Intensional Logic" in *The Stanford Encyclopedia of Philosophy* Edward N. Zalta (ed.), URL = http://www.seop.leeds.ac.uk/entries/logic-intensional/index.html

- Fodor, Jerry, (1974) "Special Sciences: Or the Disunity of Science as a Working Hypothesis" *Synthese*, 28: 97-115

- Forbes, Graham, (1985) *The Metaphysics of Modality*, Oxford: Oxford University Press

- Frege, Gottlob (1884) *Die Grundlagen der Arithmetic: eine logisch-philosophische Untersuchung zum Begriff der Zahl*, Breslau: w. Koebner; translated by J. L. Austin as *The Foundations of Arithmetic: A Logic-Mathematical Enquiry into the Concept of Number*, Oxford: Blackwell, second revised edition, 1974

- _____, (1892) "Uber sinn und bedeutung". *Z. Philos. Philos. Kritik*, 100:373-394. *reprinted as:* "On Sense and Nominatum", In Feigl H. and Sellars W., eds., (1947) *Readings in Philosophical Analysis*, pp. 85-102, Appleton-Century-Croft, New York

- _____, (1918) "Der Gedanke" *Beiträge zur Philosophie des Deutschen Idealismus*, vol. I, pp. 58-77. Reprinted as "Thoughts" In Frege (1984) (ed. B. McGuiness, trans. P. Geach and R.H. Stoothoff), *Collected Papers on Mathematics, Logic, and Philosophy*, pp. 351-72, Oxford: Blackwell

- Galois, Andre, (1998) "De re/de dicto" in E. Craig (Ed.), *Routledge Encyclopedia of Philosophy*. London: Routledge, http://www.rep.routledge.com/article/X009

- Garrett, Brian, (1998) *Personal Identity and Self-Consciousness*, London: Routledge

- Geach, Peter, (1962) *Reference and Generality: An Examination of some Medieval and Modern Theories*, Ithaca: Cornell University Press

- Gibbard, Allan, (1975) "Contingent Identity" *Journal of Philosophical Logic*, 4: 187–221

- Grandy, Richard E., (2006) "Sortals" *The Stanford Encyclopedia of Philosophy*, Edward N. Zalta (ed.), URL = http://plato.stanford.edu/entries/sortals

- Grice, H. Paul (1975) "Logic and conversation" in P. Cole and J. Morgan (eds.) *Syntax and Semantics* 3: 41-58. [originally delivered as William James lectures at Harvard University in 1967]

- _____, (1969) "Vacuous Names" in D. Davidson and J. Hintikka (eds.), *Words and Objections*, D. Reidel, pp. 118-145

- Hobbes, Thomas, (1839) *The English Works of Thomas Hobbes* Vol.1, Concerning Body, ed. William Molesworth. London: Bohn

- Horgan, Terrance, (1982) "Supervenience and Microphysics" *Pacific Philosophical Quarterly* 63: 29-43

- Hudson, Hud, (2001) *A Materialist Metaphysics of the Human Person*, Ithaca: Cornell

- Hunter, David (2005) "Soames and Widescopism," *Philosophical Studies*, 123: 231 – 241

- Hyde, Dominic (2005) "Sorites Paradox" in *The Stanford Encyclopedia of Philosophy*, Edward N. Zalta (ed.), URL = http://plato.stanford.edu/entries/sorites-paradox/

- Jackson, Frank C. (1998a) *From Metaphysics to Ethics: A Defence of Conceptual Analysis*, Oxford: Oxford University Press

- _____, (1998b) "Reference and Description Revisited" In J. Tomberlin (ed.), *Philosophical Perspectives 12: Language, Mind, and Ontology*, Oxford: Blackwell, pp. 201-218

- Johnston, Mark, (1987) "Human Beings", *Journal of Philosophy* 84: 59-83

- Kaplan, David, (1967) "Transworld Heir Lines", in Loux, M., (ed.), (1979) *The Possible and the Actual*, Ithaca, NY: Cornell University Press, pp. 88–109

- _____, (1975) "How to Russell a Frege-Church", Journal of Philosophy, 72: 716-29

- _____, (1977) "Demonstratives", in J. Almog, J. Perry and H. Wettstein (eds.), (1989) *Themes from Kaplan*, Oxford: Oxford University Press, pp. 481–563

- _____, (1978) "Dthat", in Cole, P. (ed.), *Syntax and Semantics* 9 (New York: Academic Press), pp. 221-253, reprinted in A. P. Martinich (ed.), The Philosophy of Language (Oxford: Oxford University Press, 2001), pp. 325-338

- _____, (1979) "On the logic of demonstratives", in P. French, T. Uehling, Jr., & H. Wettstein (Eds.), *Contemporary Perspectives in the Philosophy of Language*. Minneapolis: U. of Minnesota Press, pp. 401-412

- _____, (1989) "Afterthoughts", in J. Almog, J. Perry and H. Wettstein (eds.) *Themes from Kaplan*, Oxford: Oxford University Press, pp. 565–614

- Kim, Jaegwon, (1984) "Concepts of Supervenience" in Kim (ed.) *Supervenience and Mind: Selected Philosophical Essays*, Cambridge: Cambridge University Press 1993, pp. 53-78

- ____, (1987) "'Strong' and 'Global' Supervenience Revisited" *Philosophy and Phenomenological Research*, 48, 2: 315-326

- ____, (1988) "Supervenience for Multiple Domains" *Philosophical Topics*, Vol. 16: 129-150

- ____, (1990) "Supervenience as a Philosophical Concept," reprinted in Kim (ed.) *Supervenience and Mind: Selected Philosophical Essays*, Cambridge: Cambridge University Press 1993, pp.131-160

- King, Jeffrey C., (1988) "Are Indefinite Descriptions Ambiguous?" *Philosophical Studies* 53: 417-40

- ____, (2001) *Complex Demonstratives: A Quantificational Approach*, Cambridge, MA: MIT Press

- Kripke, Saul, (1959) "A Completeness Theorem in Modal Logic" *Journal of Symbolic Logic* 24: 1–14

- ____, (1971) "Identity and Necessity" in Milton K. Munitz, ed., *Identity and Individuation,* New York University Press, pp. 135–164. Reprinted in *Metaphysics: An Anthology*. Edited by Jaegwon Kim, and Ernest Sosa. Malden, MA: Blackwell Publishers, 1999, pp. 72-84

- _____, (1972) "Naming and Necessity", in D. Davidson and G. Harman, eds., *Semantics of Natural language*. Reidel, Dordrecht, pp. 253-355; Reprinted in Kripke, Saul (1980) *Naming and Necessity*, Cambridge, MA: Harvard University Press. [Originally delivered as three talks in Princeton University in January 1970]

- ____, (1977) "Speaker Reference and Semantic Reference" in *Midwest Studies in Philosophy* 2: 255-76

- Kroon, Frederick, (1987) "Causal Descriptivism" *Australasian Journal of Philosophy* 65: 1-17

- LaPorte, Joseph (2004) *Natural-kinds and Conceptual Change.* New York: Cambridge University Press

- ____, (2006a) "Rigid Designation" *The Stanford Encyclopedia of Philosophy* Edward N. Zalta (ed.), URL = http://www.seop.leeds.ac.uk/entries/rigid-designators

- ____, (2006b) "Rigid designators for properties" *Philosophical Studies* 130: 321–336

- Larson, Richard, and Segal, Gabriel, (1995) *Knowledge of Meaning*, Cambridge: MA: MIT Press

- Lennon, Kathleen, and Charles, David, (1992) *Reduction, Explanation and Realism*, Clarendon Press: Oxford

- Lepore, Ernest, and Ludwig, Kirk. (2000) "The Semantics and Pragmatics of Complex Demonstratives" *Mind*, 109: 199-240

- Lewis, David, (1966) "An Argument for the Identity Theory", *The Journal of Philosophy*, 63: 17-25

- ____, (1971) "Counterparts of Persons and their Bodies", *Journal of Philosophy* 68: 203-211; Reprinted in David Lewis (1983), *Philosophical Papers*, Vol. I. Oxford: Oxford University Press, pp. 47-54

- ____, (1972) "Psychophysical and Theoretical Identifications", *Australasian Journal of Philosophy*, 50: 249-258

- ____, (1973) *Counterfactuals*, Oxford: Blackwell

- ____, (1976) "Survival and Identity", in *The Identities of Persons*, A. Rorty (ed.), Berkeley: California, and reprinted in his (1983) *Philosophical Papers* vol. I, Oxford University Press

- ____, (1983) "Mad Pain and Martian Pain" in Lewis D. *Philosophical Papers, Vol.* 1, Oxford, Oxford University Press

- ____, (1984) "Putnam's Paradox" *Australasian Journal of Philosophy* 62: 221-236

- ____, (1986) *On the Plurality of Worlds*, Oxford, Blackwell

- ____, (1994) "Reduction of Mind" in Guttenplan (ed.) *A Companion to the Philosophy of Mind* (Blackwell). Reprinted in Lewis, D., (1999) *Papers in Metaphysics and Epistemology* (Cambridge University Press), pp. 291-324

- Linsky, Leonard. (1963) "Reference and Referents", In Charles E. Caton, ed., *Philosophy and ordinary language*. Urbana: University of Illinois Press

- Linsky, Bernard (1984) "General Terms as Designators" *Pacific Philosophical Quarterly*, 65: 259–76

- Loar, Brian, (1976) "The Semantics of Singular Terms" *Philosophical Studies* 30: 353-377

- Locke, John, (1689) *An Essay Concerning Human Understanding*, ed. P. Nidditch (1975), Oxford: Clarendon Press

- Lombard, L., (1986) *Events: A Metaphysical Study*, London: Routledge

- Lopez de Sa, Dan, (*Forthcoming* 2007) "Rigidity, General terms, and Trivialization" *Proceedings of the Aristotelian Society*

- Loux, Michael, (1998) *Metaphysics: A Contemporary Introduction,* 2nd ed. London: Routledge

- Lowe, Jonathan, (1989), *Kinds of Being: A Study of Individuation, Identity and the Logic of Sortal Terms* (Oxford: Blackwell)

- ____, (1997) "Objects and Criteria of Identity", in Hale and Wright (eds.) (1997) *A Companion to the Philosophy of Language* (Oxford: Blackwell), pp. 613–33

- Ludlow, Peter, (2005) "Descriptions" *The Stanford Encyclopedia of Philosophy* Edward N. Zalta (ed.), URL = http://www.seop.leeds.ac.uk/entries/descriptions

- Ludlow, Peter, and Neale, Stephen (1991) "Indefinite Descriptions: In Defense of Russell" *Linguistics and Philosophy*, 14: 171-202

- Mackie, David, (1999) "Personal Identity and Dead People", *Philosophical Studies* 95: 219-242

- Mackie, Penelope, (2006) "Transworld Identity" *The Stanford Encyclopedia of Philosophy* Edward N. Zalta (ed.), URL = http://plato.stanford.edu/entries/identity-transworld/

- Marti, Genoveva (1998) "Sense and reference" in E. Craig (Ed.), *Routledge Encyclopedia of Philosophy*. London: Routledge

- ____, (2004) "Rigidity and general terms" *Proceedings of the Aristotelian Society*, pp. 104-131

- McDowell, John, (1977) "On the Sense and Reference of Proper Names" *Mind* 86: 159-185

- McGinn, Colin, (1981) "The Mechanism of Reference", Synthese 49: 157-186

- McKay, Thomas, and Nelson, Michael, (2005) "The *De re/De dicto* Distinction: A Supplement to Propositional Attitude reports" *The Stanford Encyclopedia of Philosophy* Edward N. Zalta (ed.), URL = http://www.seop.leeds.ac.uk/entries/prop-attitude-reports/index.html

- McKeon, R., (ed.) (1929) *Selections from Medieval Philosophers*, 2 Vols. (New York: Scribners)

- McLaughlin, Brian, (1996) "Supervenience" in Borchert, ed., *Encyclopedia of Philosophy Supplement*. Macmillan, 558-560

- McLaughlin, Brian, and Bennett, Karen (2005) "Supervenience" in *The Stanford Encyclopedia of Philosophy* Edward N. Zalta (ed.), URL = http://plato.stanford.edu/entries/supervenience/

- Meredith, C., and Prior, A., (1956) "Interpretations of Different Modal Logics in the 'Property Calculus'" (mimeograph), University of Canterbury Philosophy Department, reprinted in Copeland, B.J., (ed.) (1996) *Logic and Reality: Essays on the Legacy of Arthur Prior*, Clarendon Press, Oxford

- Mertz, D. W., (1996) *Moderate Realism and its Logic*, New Haven: Yale

- Mill, John Stuart (1867) *A System of Logic,* London: Longmans. Definitive 8th edition. 1949 printed in London: Longmans, Green and Company

- Molière, Jean-Baptiste Poquelin, (1673) *Le Malade imaginaire*, translated by Martin Sorrell (1995) *The Hypochondriac* (Nick Hern Books)

- Montague, R., (1960) "On the nature of certain philosophical entities" *The Monist 53*: 159–194

- ____, (1970) "Pragmatics and intensional logic" *Synthèse 22*: 68–94

- Nagel, Thomas, (1986) *The View from Nowhere*, Oxford University Press

- Neale, Stephen, (1990) *Descriptions*, (Cambridge MA: MIT Press)

- ____, (1993) "Term Limits" in J. Tomberlin (ed.), *Philosophical Perspectives 7: Language and Logic*, Atascadero, CA: Ridgeview, pp. 89–124

- Nelson, Michael, (2002) "Descriptivism Defended" *Noûs* 36: 3, 408-35

- Noonan, Harold, (1976) "Dummett on Abstract Objects", *Analysis* 36: 49–54

- ____, (1978) "Count Nouns and Mass Nouns", *Analysis* 38: 167–72

- ____, (1980) *Objects and Identity*, The Hague: Martinus Nijhoff

- ____, (1989) *Personal Identity*, London: Routledge

- ____, (1991) "Indeterminate Identity, Contingent Identity and Abelardian Predicates" *The Philosophical Quarterly,* 41: 183-193

- ____, (2006) "Identity", *The Stanford Encyclopedia of Philosophy*, Edward N. Zalta (ed.), URL=http://plato.stanford.edu/entries/identity/

- Nozick, Robert, (1981) *Philosophical Explanations*, Cambridge, Harvard University Press

- Olson, Eric, (1997) *The Human Animal: Personal Identity Without Psychology*, Oxford University Press

- ____, (2002) "Personal Identity" in *The Stanford Encyclopedia of Philosophy* Edward N. Zalta (ed.), URL = http://plato.stanford.edu/entries/identity-personal

- Parfit, Derek, (1971) "Personal Identity", *Philosophical Review* 80: 3-27

- ____, (1984) *Reasons and Persons*. Oxford: Oxford University Press

- Partee, Barbara, (1970) "Opacity, Coreference, and Pronouns" in D. Davidson and G. Harman, eds., *Semantics of Natural language*. Reidel, Dordrecht, pp. 359-385

- Pelman, Alik (2006) "A Blueprint for a Calculator of Intensions" in Bennett B. and Fellbaum C., *Formal Ontology in Information Systems*, Amsterdam: IOS Press, pp. 193-203

- Perry, John R., (1972) "Can the Self Divide?" *Journal of Philosophy* 69: 463-488

- ____, (1977) "Frege on Demonstratives" *Philosophical Review* 86: 474-97

- ____, (1998) "Semantics, Possible Worlds" in E. Craig (Ed.), *Routledge Encyclopedia of Philosophy*. London: Routledge. Retrieved from http://www.rep.routledge.com/article/U039

- Plantinga, Alvin, (1974) *The Nature of Necessity*, Oxford: Oxford University Press

- ____, (1985) "Self Profile" in J. Tomberlin and P. Van Inwagen (eds.), Alvin Plantinga, Dortrecht: D. Reidel, pp. 3-97

- Plutarch, *Life of Theseus*

- Putnam, Hillary, (1962) "It Ain't Necessarily So" *Journal of Philosophy*, LIX: 658-671

- ____, (1967) "Psychological Predicates" In W.H. Capitan and D.D. Merrill (eds.), *Art, Mind, and Religion*. Pittsburgh: University of Pittsburgh Press, pp. 37-48

- ____, (1975) "The Meaning of 'Meaning'" in H. Putnam, (ed.), *Mind, Language, and Reality*, Cambridge University Press, London

- ____, (1977) "Is Semantics Possible" in Schwartz, *Naming, Necessity and Natural-kinds*, (Cornell Uni. Press, Ithaca and London)

- Quine Willard V., (1953) "On What There Is" *From A Logical Point of View*. Cambridge. Mass: Harvard U. Press

- ____, (1956) "Quantifiers and Propositional Attitudes", *Journal of Philosophy* 53: 177-187

- ____, (1960) *Word and Object* Cambridge MA: MIT Press

- ____, (1961) "Reply to Professor Marcus" *Synthese* 20, Reprinted in Quine, (1976) Ways of Paradox, rev. enlarged edition (Cambridge, Mass.: Harvard University Press), pp. 177-184

- ____, (1963a) "Reference and modality" in *From a Logical Point of View* (second ed.), New York: Harper Torchbooks, pp. 139–159

- ____, (1963b) "Identity, Ostension and Hypostasis" in *From a Logical Point of View* (second ed.), New York: Harper Torchbooks, pp. 65-79

- Rea, Michael, (ed.) (1997) *Material Constitution*, Lanham MD: Rowman and Littlefield

- Reimer, Marga, (1998a) "Donnellan's Distinction/Kripke's Test", *Analysis* 58: 89-100

- ____, (1998b) "The Wettstein/Salmon Debate: Critique and Resolution" *Pacific Philosophical Quarterly* 79: 130-151

- ____, (2003) "Reference" *The Stanford Encyclopedia of Philosophy* Edward N. Zalta (ed.), URL = http://plato.stanford.edu/entries/reference

- Richard, Mark, (1993) "Articulated Terms" *Philosophical Perspectives* 7: 207-230

- Russell, Bertrand, (1903) *The Principles of Mathematics*, Cambridge: At the University Press.

- ____, (1905) "On Denoting" *Mind* 14: 479-93

- ____, (1908) "Mathematical Logic as Based on the Theory of Types" *American Journal of Mathematics*, 30: 222-262

- ____, (1910) "Knowledge by Acquaintance and Knowledge by Description" *Proceedings of the Aristotelian Society*, 11: 108-128. Reprinted in Russell, Bertrand, (1917) *Mysticism and Logic,* London: George Allen and Unwin

- ____, (1912) *The Problems of Philosophy.* London: Home University Library

- ____, (1919) "Descriptions" in *Introduction to Mathematical Philosophy*, London: George Allen and Unwin Publishers Ltd

- ____, (1957) "Mr Strawson on Referring" in *Mind* 66: 385-389

- Russell, Bertrand, and Whitehead Alfred North, (1910, 1912, 1913) *Principia Mathematica*, 3 vols, Cambridge: Cambridge University Press

- Salmon, Nathan, (1979) "How *Not* to Derive Essentialism from the Theory of Reference" *The Journal of Philosophy*, 76: 703–25

- ____, (1981) *Reference and Essence* (Princeton: Princeton Uni. Press)

- ____, (1982) "Assertion and Incomplete Definite Descriptions", *Philosophical-Studies* 42: 37-46

- ____, (1986) *Frege's Puzzle* (MIT Press)

- ____, (1991) "The Pragmatic Fallacy", *Philosophical Studies*: 83-97

- Searle, John, (1983) *Intentionality*, Cambridge: Cambridge University Press

- Schwartz, Stephen P., (2002) "Kinds, General Terms, and Rigidity: A Reply to LaPorte" *Philosophical Studies*, 109: 265–77

- Shiffer, Stephen, (1978) "The Basis of Reference" *Erkenntnis* 13: 171-206

- Shoemaker, Sydney, (1970) "Persons and Their Pasts", *American Philosophical Quarterly* 7: 269-285

- ____, (1984) "Personal Identity: A Materialist's Account", in Shoemaker and Swinburne, *Personal Identity*, Oxford: Blackwell

- ____, (1997) "Self and Substance", in *Philosophical Perspectives* 11, J. Tomberlin (ed.): 283-319

- _____, (1999) "Self, Body, and Coincidence", *Proceedings of the Aristotelian Society*, Supplementary Volume 73: 287-306

- Sidelle, Alan (1992) "Rigidity, Ontology, and Semantic Structure" *The Journal of Philosophy*, Vol. 89, No. 8: 410-43

- Sider, Ted, (1996) "All the world's a stage" *Australasian Journal of Philosophy* 74: 433-53

- ____, (1999) "Global Supervenience and Identity Across Times and Worlds" *Philosophy and Phenomenological Research* 59: 913-37

- Smart, J.J.C., (1959) "Sensations and Brain Processes", *Philosophical Review*, 68: 141-156

- Smiley, T., (1957) "Modal Logic", lecture handout, Department of Philosophy, University of Cambridge

- Smith, A., D., (1984) "Rigidity and Scope" Mind, XCII, 369: 177-93

- Snowdon, Paul, (1990) "Persons, Animals, and Ourselves", in *The Person and the Human Mind*, C. Gill. (ed.), Oxford: Clarendon Press

- Soames, Scott (1989) "Direct Reference and Propositional Attitudes" in Almog, J., J. Perry, and H. Wettstein, eds. (1989) *Themes from Kaplan*, New York: Oxford University Press, pp. 393-419

- ____, (2002) *Beyond Rigidity: The Unfinished Semantic Agenda of Naming and Necessity*, New York: Oxford University Press

- Sosa, David, (2001) "Rigidity in the Scope of Russell's Theory" *Noûs*, 35: 1–38

- Stalnaker, Robert C., (1972) "Pragmatics" in Davidson and Harman 1972, pp. 380–97. Reprinted in Martinich (1990), pp. 176–86

- ____, (1978): "Assertion", in P. Cole (ed.), *Syntax and Semantics, Volume 9: Pragmatics*: Academic Press*, pp. 315–332

- ____, (1986) "Counterparts and Identity", in French, P., Uehling, T., and Wettstein, H., (eds.), 1986, *Midwest Studies in Philosophy XI: Studies in Essentialism*, Minneapolis: University of Minnesota Press, pp. 121–40

- ____, (1996) "Varieties of Supervenience" *Philosophical Perspectives* 10: 221-41

- Stanley, Jason (1997) "Names and Rigid Designation" in Bob Hale and Crispin Wright (eds.), *A Companion to the Philosophy of Language*. Oxford: Blackwell, pp. 555–85

- Stanley, Jason, and Szabo, Zoltan G., (2000) "On Quantifier Domain Restriction" *Mind and Language*, 15: 219-61

- Stevenson, Leslie, (1972) "Relative Identity and Leibniz's Law", *Philosophical Quarterly*, 22

- Steward, Helen, (1990) "Identity Statements and the Necessary A Posteriori", *Journal of Philosophy* 87: 385–98

- Strawson Peter F., (1950) "On Referring", *Mind*, 59: 320-44

- ____, (1959) *Individuals* (London: Methuen)

- ____, (1974) *Subject and Predicate in Logic and Grammar*, London: Methuen

- Swoyer, Chris (2000) "properties" in *The Stanford Encyclopedia of Philosophy* Edward N. Zalta (ed.), URL = http://plato.stanford.edu/entries/properties/#instantiation

- Thomson, Judith J., (1997) "People and Their Bodies", in *Reading Parfit*, J. Dancy (ed.), Oxford: Blackwell

- Unger, Peter, (1980) "The Problem of the Many" in P. French, T. Uehling & H. Wettstein (eds.), *Midwest studies in Philosophy*, Vol. 5: *Studies in Epistemology* (Minneapolis, Uni. Of Minnesota Press), pp. 411-467

- ____, (1990) *Identity, Consciousness, and Value*, New York: Oxford University Press

- ____, (2000) "The Survival of the Sentient", in *Philosophical Perspectives* 11, J. Tomberlin (ed.), Malden, MA: Blackwell

- van Inwagen, Peter, (1985) "Plantinga on Trans-world Identity", in *Alvin Plantinga*, J. Tomberlin and P. van Inwagen (eds.), Dordrecht: Reidel, pp. 101–20

- ____, (1997) "Materialism and the Psychological-Continuity Account of Personal Identity", in *Philosophical Perspectives* 11, J. Tomberlin (ed.), and reprinted in his *Ontology, Identity, and Modality*, Cambridge University Press, 2001

- Wettstein, Howard, (1981) "Demonstrative Reference and Definite Descriptions" *Philosophical Studies* 40: 241-257

- ____, (1983) "The Semantic Significance of the Referential-Attributive Distinction", *Philosophical Studies* 44: 187-196

- Wiggins, David, (1967) *Identity and Spatio-temporal Continuity*, Oxford: Blackwell

- ____, (1975) "Identity, designation, essentialism, and physicalism" *Philosophia* 5: 1-30

- ____, (1980) *Sameness and Substance*, Oxford: Blackwell

- _____, (1984) "The Sense and Reference of Predicates: A Running Repair to Frege's Doctrine and a Plea for the Copula" in C. Wright (ed.) (1984) *Frege: Tradition and Influence*, Oxford: Blackwell

- _____, (2001) *Sameness and Substance Renewed*, Cambridge: Cambridge University Press

- Williams, Bernard, (1956-7) "Personal Identity and Individuation", *Proceedings of the Aristotelian Society* 57, and reprinted in his *Problems of the Self*, Cambridge University Press, 1973

- _____, (1970) "The Self and the Future", *Philosophical Review* 59, and reprinted in his *Problems of the Self*, Cambridge University Press, 1973

- Williams, D., C., (1953) "The Elements of Being I" *Review of Metaphysics* 7: 3-18

- William of Sherwood (1240's) *Introductiones in Logicam*, in M. Grabmann (ed.) (1937), "Die Introductiones in logicam des Wilhelm von Shyreswood", Munich: *Sitzungberichte der Akademie der Wissenschaften*, Philosophisch-historische Klasse 10, pp. 30–106; tr. N. Kretzmann (1966) *William of Sherwood's 'Introduction to Logic'*, Minneapolis: University of Minnesota Press

- Wittgenstein, Ludwig, (1921) *Tractatus Logico-Philosophicus*, Routledge and Kegan Paul

- _____, (1956) [1978], *Remarks on the Foundations of Mathematics*, Revised Edition, Oxford: Basil Blackwell, G.H. von Wright, R. Rhees and G.E.M. Anscombe (eds.); translated by G.E.M Anscombe

- Zalta, Edward N., (1988) *Intensional Logic and the Metaphysics of Intentionality*. Cambridge, MA: MIT Press