In the preceding chapter, we saw that in order to arrive at the ‘new’ view on names, we had to supplement the notion of rigid designation with the assumption of direct reference. Only then is the picture of names as rigid and non-descriptive fully described. It also became apparent that this view on proper names has to be accompanied by assumptions about the identity of individuals. These assumptions are at the core of the dispute between descriptivist theories of proper names and the group of new theories that includes causal theories and theories of direct reference.

An analysis of modal statements, and of the possible-world frameworks that underlie particular approaches, should provide us with a good basis on which to evaluate various positions to the issue of identity of individuals. In this chapter, we shall pay a lot of attention to the transworld identity of individuals but we should keep in mind that the lessons thus learned also have a direct bearing on what is assumed about the identity of individuals within one world, for example the actual one.

I shall present and analyse the possible-world frameworks of David Lewis, Saul Kripke, and Robert Stalnaker, in this order. I chose this particular sequence of presentation because it will allow us to progress from the most realist view on possible worlds to the
least realist one, and – as I will try to show – from one least motivated by speakers’ intuitions about natural language to the one that seems to capture them best. We shall focus on the ontological commitments of each of these frameworks, especially those that pertain to the identity of individuals. We shall try to highlight the connections between the way a particular theory builds its possible-world framework, the stance it takes with respect to the issue of transworld identity of individuals, and the predictions it makes concerning the modal status of particular kinds of propositions involving individuals.

In the course of this inquiry, we shall focus on a number of questions, and try to determine the relevant answers with respect to each possible-world framework as we encounter it: What is the author’s own perception of the place of modal semantics, that is, what does the author think he models? What presuppositions does his particular position on haecceitism and essentialism entail, and how are those presuppositions argued for? Could this possible-world framework function without essentialist assumptions? How does a particular modal framework co-determine a theory of content, and what is the epistemic status of the terms in which reference is specified? Do Lewis’s, Kripke’s, and Stalnaker’s approaches to possible worlds address the same issues?

3.1 Lewis’s Possible-world Framework

In the following sections, we shall examine Lewis’s notion of a possible world and his theory of counterparts, and then look at the motivation underlying the two, as well as their mutual relations. We shall pay especially close attention to those parts of his theory that have a bearing on the identity of individuals, i.e., some implications of the theory of counterparts, and the position with respect to essentialism it entails. Finally, we shall assess the plausibility of Lewis’s proposal in the context of the broader tasks of a semantic theory of modal statements, focusing on the process of evaluation of modal statements and the role of modal statements in communication.
3.1. Lewis’s Possible-world Framework

3.1.1 Lewis: Let’s Be Realistic

There are two basic approaches to the notion of possible worlds. One can either be a possibilist, and hold that there are such things as possible worlds, which are entities in their own right, irreducible to anything else, or one can develop a theory that does not require a commitment to the existence of possible worlds. This view is advocated by various actualist and paraphrastic approaches. Theorists who adopt the paraphrastic approach believe that while one may use the terminology of possible worlds, it is, strictly speaking, just a façon de parler, and the apparent reference to possible entities can and should be paraphrased away. Actualists, on the other hand, try to construct possible worlds from actual entities of some kind. Their approach is to try and find some actual entities that are analogous to possible worlds and can, therefore, serve as possibilia.\footnote{This classification is loosely based on Lycan’s overview in Lycan, 1979, 285.}

Our excursion into the ways of building possible worlds starts with possibilism and its most prominent advocate, David Lewis. He summarises his doctrine of possible worlds as adherence to the following theses:\footnote{My overview is based on Stalnaker’s overview in Stalnaker, 1979, 227, which in turn is based on Lewis, 1973, 84-91.}

(1) **Possible worlds exist.** They are just as real as the actual world. They do not actually exist, since to actually exist is to exist in the actual world.

(2) **Other possible worlds are things of the same kind as the actual world – “I and my surroundings.”**\footnote{Lewis, 1979a, 184.} They differ not “in kind but only in what goes on in them. Our actual world is only one world among others. We call it alone ‘actual’ not because it differs in kind from all the rest but because it is the world we inhabit.”\footnote{Lewis, 1979a, 184.}

(3) **The indexical analysis of ‘actual’ is the correct analysis.** “The inhabitants of other worlds may truly call their own
worlds actual if they mean by ‘actual’ what we do; for the meaning we give to ‘actual’ is such that it refers at any world \( i \) to that world \( i \) itself. ‘Actual’ is an indexical, like ‘I’ or ‘here’, or ‘now’: it depends for its reference on the circumstances of utterance, to wit the world where the utterance is located.”

(4) Possible worlds cannot be reduced to anything more basic. Lewis says: “I emphatically do not identify possible worlds in any way with respectable linguistic entities; I take them to be respectable entities in their own right. When I profess realism about possible worlds, I mean to be taken literally. Possible worlds are what they are and not some other thing.”

This is the basic doctrine, which was introduced in Lewis’s ‘Possible Worlds’. How does Lewis motivate this position? The answer to this question comes in two parts: firstly, there is the folk-psychological motivation which Lewis is overtly trying to offer, and secondly, there are theoretical considerations that lead him to this view.

Lewis opens the exposition of his views by saying that “it is uncontroversially true that things might be otherwise than they are.” Equally uncontroversially, so he continues, we could then say that “there are many ways things could have been.” And concludes that we could then call them ‘possible worlds’.

Lewis is trying to coax us into accepting that we have believed in possible worlds all along because ‘ways things could have been’ are a part of the folk ontology. We should, however, ask ourselves

---

5Lewis, 1979a, 184.
6Lewis, 1979a, 185.
7Lewis, 1979a.
8All quotations in this paragraph are from Lewis, 1979a, 182.
9It seems clear from the context of the article (Lewis, 1979a) that ‘folk ontology’ is to be understood as being a part of folk psychology. A folk ontology is the sum of kinds of entities whose existence seems implied by our every-day ways of speaking and dealing with the world.
whether the alleged existence of something within a folk ontology is a good enough reason for adopting it into a semantic theory. After all, the existence of ‘sakes’ – as in ‘for God’s sake’, ‘for my sake’, etc. – is implied by the folk ontology too, and yet we would consider a claim that ‘sakes’ exist quite eccentric. While folk psychology is a reasonably good guide in our everyday reasoning, folk ontology is not necessarily a good basis for a semantic theory. Even so, let us concede for the argument’s sake that we do believe in the existence of the ‘ways things might have been’. But still, nothing in folk ontology implies that we should think of them as existing in the same way in which the actual world exists. Folk ontology seems silent on this issue, and Lewis still owes us an argument why we should adopt this view.

Perhaps then thesis (1) is motivated not by an observation of natural language, but by a different kind of consideration: a conviction that modal notions should be analysed in terms of possible worlds because every other explanation turns out to be circular in the end. Let us assume for the time being that ‘possibly S’ means that S is a consistent sentence. But what is consistency? Rephrasing Lewis, we could say that a consistent sentence is one that could be true (or, equivalently, one that is not necessarily false), the explanation is circular. If a consistent sentence is one that comes out true under some assignment of extensions to its non-logical vocabulary, the explanation is incorrect, because some assignments of extensions are impossible. For example, an assignment that would have the extensions of ‘sheep’ and ‘pig’ overlap, is not possible. If, however, we say that a sentence is consistent if true under some possible assignment of extensions, the explanation is once again circular. Lewis argues that if we analyse modal notions as quantifiers over possible worlds but assume that possible worlds are some kind of ‘respectable’ linguistic entities, for example maximal consistent sets of sentences of some language, or maximal sets of atomic sentences, the theory turns out to be either circular or incorrect, depending on how we explain

\[\text{10For example Stalnaker, 1979 defends the indexical analysis of actuality while rejecting Lewis's full-blown realism about possible worlds.}\]

\[\text{11Lewis, 1979a, 183.}\]
consistency.\textsuperscript{12} That is why he concludes that possible worlds are entities in their own right, irreducible to anything more primitive. And this, I believe, is the primary consideration that motivates Lewis’s thesis (1).

Lewis establishes his ontology in thesis (1), where he claims that possible worlds are just as real as our world, and in thesis (4), which follows from (1), where he states that they are entities in their own right. While thesis (3) does not deal with ontology explicitly, it has an impact on it. It establishes a distinction between actual existence and non-actual existence.\textsuperscript{13} As we shall see later, Lewis’s apparatus has largely been directed towards an analysis of counterfactuals. This, I believe, can help us understand the very realist approach to possible worlds: If we think of possible worlds rather as if they were counterfactual situations, and view counterfactual situations as situations which would have been actual, had things developed otherwise, then thinking of possible situations as quite on a par with actual situations seems not quite counterintuitive. And taking from there the step into thinking about possible worlds as being of the same kind as the actual worlds is then a natural consequence of this line of thinking.

Now that we have familiarised ourselves with the basics of Lewis’s theory of possible worlds, we shall look at how Lewis deals with identity of individuals in possible worlds.

### 3.1.2 Adopt a Counterpart!

We shall now introduce Lewis’s theory of counterparts, and then look at how it relates back to Lewis’s notion of possible world. In the preceding section, we saw that once we start thinking of possible worlds as being just as real as the actual world, the conclusion that they cannot be reduced to anything else seems to follow. In this section, we shall try to show that the theory of counterparts is a

\textsuperscript{12}Lewis, 1979a is largely devoted to debunking less-than-realist ways of constructing possible worlds. The arguments sketched here are found in Lewis, 1979a, 183.

\textsuperscript{13}We shall treat this distinction in greater detail later.
natural consequence of the realist approach to possible worlds. At this point, we shall not try to argue for or against realism about possible worlds, but rather just investigate how its various parts are interconnected.

Every possible-world framework takes some view on the issue of identity of individuals. Some theories adopt the notion of transworld individuals, that is, the view that one and the same individual can exist in more than one possible world. Other theories reject that view, claiming that strictly speaking, this cannot be the case. Lewis, who advocates the theory of counterparts, belongs to the latter group.

Lewis outlines the treatment of individuals within his possible-world framework in his theory of counterparts, which he defines by the following theses:\textsuperscript{14}

(a) Nothing is in anything except a world.

(b) Nothing is in two worlds.

(c) Whatever is a counterpart is in a world.

(d) Whatever has a counterpart is in a world.

(e) Nothing is a counterpart of anything else in its world.

(f) Anything in a world is a counterpart of itself.

(g) Some world contains all and only actual things.

(h) Something is actual.

The world mentioned in (g) is unique. That is because if something is actual (claim (h)), nothing is in two worlds (b), and we assume the indexical analysis of ‘actual’, then everything that is in the same world as the actual entity is actual (everything in that world is actual), and from the actual entity’s point of view nothing that does

\textsuperscript{14}Lewis, 1979b, 111.
not share the same world is actual (the world in which the actual entity is then contains all actual things).

Crucial to Lewis’s treatment of identity of individuals in various possible worlds is the notion of a counterpart relation. Where in a framework opting for transworld individuals I may have different properties in different possible worlds, in Lewis’s framework, thesis (b) says that I stay put in one world, my actual world, and have counterparts in non-actual worlds that resemble me in various ways. But my counterparts are not really me.

The counterpart relation is a relation of similarity, and similarity is here understood rather informally. Lewis sees that this may cause problems, admitting that the counterpart relation

\[ \ldots \text{is problematic in the way all relations of similarity are: it is the resultant of similarities and dissimilarities in a multitude of respects weighted by the importances of the various respects and by the degrees of similarities.}^{15} \]

Let us, however, put the problems inherent in the notion of similarity aside, and see what else can be said about the counterpart relation. It has a number of important properties.\(^{16}\) It is nontransitive: If something, \(x_1\), resembles me more closely than anything else in a world \(w_1\), and something else, \(x_2\), resembles \(x_1\) more closely than anything else in its world \(w_2\) does, then \(x_2\) is a counterpart of \(x_1\), but not necessarily my counterpart. There might be something in \(w_2\) that resembles me more closely.

The counterpart relation is nonsymmetric: Suppose there is something, \(x_1\), in a world \(w_1\) that is a blend of my twin sister Marie and me. Suppose also, that it resembles Marie more closely than it resembles me. It may well be the thing that resembles me in that world most closely, and is therefore my counterpart. On the other hand, because it resembles Marie more closely than me, its counterpart in the actual world will be Marie, not I.

Something can have more than one counterpart in another possible world: In a world where there are, for example, two persons

\(^{15}\text{Lewis, 1979a, 112.}\)
\(^{16}\text{Which Lewis outlines in Lewis, 1979b, 113.}\)
who resemble me equally closely, both of them are my counterparts. This works also the other way around: there could be a world where Marie and I have one common counterpart, say Marianne, because it holds for both of us that there is nothing in that world that resembles either of us more closely.

As we noted, the counterpart relation is a relation of similarity, and there is little one can say about it in general terms. In particular models, however, we can specify what kind of similarity we are interested in. That is why in a model where the relation is in some way specified, there can be something in a possible world, $w_1$, that does not have any counterpart in another possible world, $w_2$, and, by the same principle, there can be some something in some possible world, $w_2$, that is not a counterpart of anything in a possible world $w_1$.

Surprising as it is, the counterpart theory is a natural outgrowth of Lewis’s conception of possible worlds. We shall try to show this by examining the tenets of Lewis’s conception of possible worlds, and focusing on their consequences for the identity of individuals.

We know from thesis (2), (see p. 63) that possible worlds are of the same kind as the actual world, and from thesis (4) that they are not reducible to anything more primitive. This implies, among other things, that if something is a wooden desk in the actual world, some of its counterparts might be wooden desks, and that if I am a person of flesh and blood in this world, then in another possible world my counterpart can still be a person of flesh and blood, and not some kind of shadow of myself (as Quine was once inclined to object). Possible worlds are not some sort of shadows of the actual world; they are concrete to the same degree our world is. (They are also abstract to the same degree, but that is not the point.)

\footnote{Wyman’s overpopulated universe is in many ways unlovely. It offends the aesthetic sense of us who have a taste for desert landscapes, but this is not the worst of it. Wyman’s slum of possibles is a breeding ground for disorderly elements. Take, for instance, the possible fat man in that doorway; and again, the possible bald man in that doorway. Are they the same possible man, or two possible men? How do we decide? How many possible men are there in that doorway?” Quine, 1961a, 4.}
that if possible worlds have the ontological status Lewis claims they have, they cannot overlap. Here is how we can show it.

Suppose there is an individual, say Anna, in the intersection of the set of individuals of \( w_1 \) and that of \( w_2 \). Anna is then both in \( w_1 \) and in \( w_2 \), and by thesis (3) she calls both \( w_1 \) and \( w_2 \) her actual world. But facts in \( w_1 \) and \( w_2 \) differ, for otherwise they would be the same world. So something is and is not a fact for Anna, which is absurd. Therefore, an individual cannot belong to the set of individuals of two distinct worlds.\(^{18}\)

Because possible worlds cannot overlap, they cannot share individuals. If I exist in the actual world, then I cannot at the same time exist in another world because that world would be my actual world. If I am a certain spatiotemporal entity in one world, I cannot be the same entity in another world. The best we can do is to say that in some possible world there is someone a lot like me, my counterpart. And that is what we set out to show – the counterpart theory is a consequence of Lewis’s theses (2), (3), and (4) describing possible worlds. The importance of this observation lies in its pointing out that it is not possible to adopt a strongly realist view of possible worlds, and a notion of a transworld individual at the same time.

In the next section, we shall turn our attention to the indexical theory of actuality and some reasoning that seems to make it more plausible than it may seem at first sight.

3.1.3 The Indexical Theory of Actuality, Natural Remedies, and the Man in the Street

According to the indexical theory of actuality, the actuality of the actual world consists in its being our world, the world in which this

\(^{18}\)On a slightly different note, if we discovered that in some far away galaxy there is a world just like ours (except that, perhaps, what seems to be water on that planet is composed of XYZ rather than \( \text{H}_2\text{O} \), it would still be a discovery about the actual world. Possible worlds are not far-away planets. They are not like Putnam’s Twin Earth, which we treat in section 3.3.5.
is being written. Actuality is a property a world possesses not absolutely but relatively, in relation to its inhabitants. This may seem to run counter to our intuition that the actual is, absolutely considered, more real than the merely possible. After all, we care more about actual events than about possible ones, more about actual people rather than possible ones, and more about actual train schedules rather than just possible ones. We aim at exploring and making predictions about the actual world rather than about possible worlds. We seem prejudiced in favour of the actual. This is reflected in our attitudes. When I say ‘I intend to meet you tomorrow at 5 p.m.’, I am expressing my intention of meeting you in the actual world, not just in any world, for although it certainly is possible that we meet in one or another possible world, what I want is for us to meet in the actual one.

If we do, indeed, think that the actual world is a place fundamentally unlike any other, we may be disinclined to adopt the indexical theory of actuality, and inclined to replace it with one more intuitively plausible. What would such a theory look like? It would probably hold that actuality is a property of the actual world, and that it distinguishes it from all other possible worlds. It is a property the actual world possesses absolutely rather than just in relation to its inhabitants.

This view would take seriously our certainty of our own actuality; it would take into account our prejudice in favour of the actual. It may seem to be a reasonable alternative to the indexical theory. But how does it account for non-actual possible worlds? Presumably, non-actual possible worlds could have been actual, that is, are possibly actual. That seems to be the basis of reasoning with counterfactuals. It also implies, however, that for any non-actual possible world $w_1$, there is some possible world $w_2$ in which the world $w_1$ is actual. But then $w_1$ and $w_2$ must be one and the same. So $w_1$ is actual in $w_1$. In effect, each possible world is actual in itself.

How then does the actual world differ from other possible worlds

---

19 This much is in the basic tenets of Lewis's doctrine of possible worlds, p. 63.
with respect to the actuality property? It has the actuality property actually, and not just possibly. Yet we just saw that every possible world is actual in itself. So how is the actual world different from other worlds? It is actually actual as opposed to just possibly actual (i.e., actual for its actual inhabitants as opposed to being actual to its possible inhabitants) — but that is the view we wanted to oppose. It thus seems that the ‘intuitive’ theory of actuality leads to much the same view on actuality as the indexical theory did.

As I mentioned before, an analysis of the functioning of counterfactuals is an important part of Lewis’s work, and it can help us understand the reasons behind his adoption of strongly realist possible worlds. It is important to note that when we want to explain our use of counterfactuals, we look not only at the truth conditions of counterfactual statements (e.g., ‘This car could have killed me.’), but also at why they express states of affairs that concern us. Counterfactual states of affairs, or possible worlds, concern us because they could have been the case, they could have been actual. Once we take this insight seriously, it seems hard to hold that actuality is an absolute property of the actual world.

It seems that our dissatisfaction with both the ‘intuitive’ and the indexical theory of actuality stems from an incompatibility of the intuitions which drive our thinking about actuality and possible worlds. We can try to satisfy our ‘prejudice in favour of the actual’, and run then into problems when explaining why possibilities concern us, or we can start from taking possibilities seriously, and end up dissatisfied with the conclusion that there is nothing inherently special about the actuality of our world. Lewis has a coherent proposal regarding the status of actuality, and unless we come up with a solution that would do justice to both sides of our intuitions, it cannot easily be dismissed.

But even if we consider Lewis’s approach to actuality with the seriousness it deserves, we should ask ourselves whether Lewis’s ‘actual world’ is anything like what the proverbial man in the street would think it is. In particular, can we all inhabit the same Lewisian

---

actual world? It follows from thesis (3) (p. 63) that I, the writer, inhabit the actual world. I and my surroundings are actual by definition. The same holds for every speaker or writer. But does it follow that my actual world and another speaker’s actual world are the same? The answer is not quite straightforward.

One could try to support a positive answer by pointing to the language Lewis uses – he consistently refers to ‘our world’, ‘the world we inhabit’.\(^{21}\) This, however, is not a very conclusive kind of evidence. We can arrive at a better-grounded answer if we look at Lewis’s theory in more detail.

A Lewisian possible world has parts,\(^{22}\) namely possible individuals. If two things are a part of the same world, we call them ‘worldmates’. Being parts of the same world, worldmates are spatiotemporarily related. It works also the other way around – what unifies a world are the spatiotemporal relations of its parts.\(^ {23}\) If things are related in space and time, they are a part of the same world. If something in a world \(w_1\) were to interact with a thing in a world \(w_2\), then the two worlds would have to be identical. This is in part a consequence of thesis (b), introduced on p. 67, which states that nothing can be a part of two worlds. Because distinct possible worlds are not spatiotemporarily related, an event in one world cannot cause an event in another world. There is no transworld causation. And it seems indisputable that other speakers and I are related to one another in space and time, and we do causally interact. Therefore, we inhabit the same actual world. At least in this sense then the world we all inhabit is the actual world of the man in the street.

\(^{21}\)Lewis, 1979a, 184, my italics.

\(^{22}\)The following presentation follows Lewis, 1986, 69-81.

\(^{23}\)Lewis here assumes that spatio-temporal continuity defines the identity of a world. It is an independent assumption that cannot be derived from any of the theses that were used to define possible worlds.
3.1.4 Essentialism: A Matter of Choice?

Just as possible worlds are alternatives to entire worlds, so some parts of those worlds are alternative possibilities for individuals. And just as quantification over possible worlds can be restricted by various accessibility relations, so quantification over possible individuals can be restricted by various counterpart relations. We can restrict both of these relations for particular purposes in various ways. We could restrict the possible worlds over which we wish to quantify, and consider only the nomologically or historically accessible, the epistemically accessible, or for example only the doxastically accessible ones. We could impose similar restrictions on the counterpart relation. We may choose to consider only the worlds that obey the same laws of physics as our world (or the same legal standards, for that matter), and similarly we can consider only counterparts who have the same physiology as we do, or have the same number of toes.

Lewis points out the generality of his framework, saying that

\[ \ldots \text{sometimes one is expected to take a position, once and for all, about what is or isn’t possible } \text{de re} \text{ for an individual.} \]

I would suggest instead that the restricting of modalities by accessibility or counterpart relations, like the restricting of quantifiers generally, is a very fluid sort of affair: inconstant, somewhat indeterminate, and subject to instant change in response to contextual pressure. Not anything goes, but a great deal does.\(^\text{24}\)

Given this approach, Lewis could accommodate a variety of essentialist positions, including Kripke’s. He could restrict the counterpart relation for humans, for example, and consider as a counterpart of a person only something that is human, and that has the ancestry it does in the actual world. He could do the same for material objects and restrict the counterpart relation to only those individuals that have the same material composition. But doing so absolutely, saying that we can always consider only such counterparts, would be an uncongenial move within his framework because Lewis is not an essentialist in the same sense in which for example Kripke is.

\(^{24}\text{Lewis, 1986, 8.}\)
Lewis aims at providing a general possible-world framework, more flexible than any of the frameworks we shall consider later. It can be used for all kinds of purposes in response to different kinds of contexts. The counterpart relation is a relation of similarity. Everything is like anything else in some respect, and for particular purposes we choose to consider particular similarities, and, accordingly, adopt various restrictions on the counterpart relation. But it is important to note that no particular restriction on either the kind of possible worlds we wish to consider, or on the counterpart relation, can be motivated from the nature of the framework itself.

3.1.5 A Battle of Individuals: Transworld Versus Worldbound

As we mentioned before, Lewis addresses in his counterpart theory one of the most interesting and difficult questions of modal semantics: Can an individual exist in more than one possible world? The two basic answers on the market are, not surprisingly, yes, an individual can exist across possible worlds, that is, there are transworld individuals; and no, an individual can only exist in one possible world, that is, individuals are worldbound. And we have shown that Lewis is an advocate of worldbound individuals.

In this section, we shall consider one well-known argument against transworld individuals. It does not come from Lewis’s writing – for Lewis, worldbound individuals are a natural consequence of his ontology, and he sees little need to provide independent support for his position on individuals. Other theorists, however, have proposed both direct and indirect arguments for and against worldbound individuals.

The argument we shall consider here was proposed by Roderick Chisholm.\textsuperscript{25} It is directed against the notion of transworld individuals, and uses reasoning about gradual changes. The principle is simple: Imagine an entity in the actual world, alter its description slightly, adjust the description of other entities in that world.

\textsuperscript{25}I adapted my version from Chisholm, 1979.
to accommodate this alteration, and then ask yourself whether the entity in the world we thus reached is identical with the entity in the world we started in. We start, for example, with the actual William Shakespeare, allow him to live for 53 years instead of 52 years, and accommodate other descriptions in this world, so that Anne Hathaway was married to a man who lived to be 53, etc. Thus we arrive at a description of another world. Let us call the actual world \( w_1 \), and the world we arrived at \( w_2 \). Is the Shakespeare in \( w_1 \) the same man as in \( w_2 \)? One could object that any identity of the Shakespeare of \( w_1 \) with the Shakespeare of \( w_2 \) is incompatible with the thesis of indiscernibility of identicals. How could the Shakespeare who lived 53 years be identical with a Shakespeare who lived 52 years?

We might, however, see this as parallel with a different question: How could the Shakespeare who got married at the age of 18 be the same man who wrote *The Winter's Tale*, if the former is young and the latter old? One could perhaps say that it is not true that the old Shakespeare has properties that make him distinct from the young Shakespeare, that it is rather the case that Shakespeare had the property of being young when he got married, and the property of being somewhat advanced in age when he wrote *The Winter's Tale*. His properties, though different, are not incompatible. And this holds for the different possible worlds, too: Shakespeare can consistently have the property of *living for 52 years* in \( w_1 \) and *living for 53 years* in \( w_2 \). We could thus assume that the actual Shakespeare is identical with the Shakespeare of the world where he lived 53 years.

Now let us now suppose that we arrived at \( w_2 \) by not only altering Shakespeare’s age, but also introducing alterations to our –actual-world – description of Francis Bacon, Lord Verulam and Viscount St. Albans. In \( w_2 \), Sir Francis Bacon lived to be 64 instead of 65, his wife was married to a man who died at the age of 64 etc. Now let us move to a world \( w_3 \) where Shakespeare lived to be 54 and Sir Francis 63 years old, while, again, accommodating these changes in the rest of that world. Moving thus from one possible world into the next, we arrive at a world in which Shakespeare died at the age of 65, and Sir Francis at the age of 53. In this world, Shakespeare and Sir Francis
3.1. Lewis's Possible-world Framework

have, so to speak, exchanged their ages. Let us then move into yet other possible worlds and exchange further properties, so that in $w_k$, Sir Francis authored *Hamlet*, *Romeo and Juliet*, and *King Lear*, as well as the rest of the plays commonly ascribed to Shakespeare in the actual world, and Shakespeare is the Lord Chancellor to James I, the author of *Novum Organum*, etc. Finally, let us move into a world where the two exchange their names and titles. Proceeding in this way, we finally arrive at a world $w_n$ which is like $w_1$, the actual world, except that the Shakespeare of $w_n$ can be traced back to Sir Francis of $w_1$ and vice versa. Should we now say that the Shakespeare of $w_n$ is identical with Sir Francis of $w_1$? In other words, is there such an $x$ so that $x$ is Sir Francis Bacon in $w_1$ and Shakespeare in $w_n$? How should we decide?

Assume that we answer in the affirmative: there is such an $x$ such that it is Shakespeare in $w_n$ and Sir Francis in $w_1$. But if this is the case, how are we to tell the two worlds apart? Should we say that though different, these worlds are indiscernible from one another? The two Shakespeares could be seen as discernible because one has the property of being Sir Francis while the other does not. The option of distinguishing worlds by essential properties of some of their individuals shows that there can be a sense of ‘indiscernible’ on which ‘indiscernibles are identical’ tells us more than ‘identicals are identical’.

If $w_1$ and $w_n$ are two different possible worlds, then there could be infinitely many other possible worlds as difficult to distinguish from one another as $w_1$ and $w_n$ are. Why do we assume that the Shakespeare of $w_1$ is identical with the Sir Francis of $w_n$? We made this conclusion possible once we conceded that an individual can be found in more than one possible world. It seemed perfectly reasonable to assume that Shakespeare retains his identity through small changes, such as were involved in the transition between $w_1$ and $w_n$. These transitions can be as gradual and slight as one likes, but once we allow identity to be transitive, and accept that an individual can exist in more than one possible world, we seem to take the first step on quite a slippery slope.

Is there a way of retaining identity through possible worlds while
blocking the extreme consequences such as we just encountered? Adoption of non-trivial essential properties could be a solution. But if essential properties were to help us with our problem, we would have to be able to specify which properties are essential for a given individual. Only that way could we decide which transitions from one possible world to another lead to a ‘loss of identity’. However, Chisholm objects, there seems to be no procedure for deciding which properties are essential and which accidental of a particular individual. Properties can be necessary under a certain description of an object, but, according to Chisholm, we do not have any meaningful procedure that would deliver necessary properties in the required, non-trivial sense, in isolation from a particular context.

We can summarise Chisholm’s conclusion as follows: Once we allow for any property of an individual to be changed without loss of identity, there is no acceptable way of stopping the process. Ultimately, we reach the counterintuitive result that, so to say, something can look like an apple, smell like apple, taste like apple, and yet be an avocado. This result could be avoided only by either positing essential properties that are unique to each individual – but, as Chisholm claims, we do not have a good way of deciding which properties should play this role – or by preventing the whole slippery slope altogether. The latter option would amount to banning all changes of properties on pain of loss of identity. This step is in fact equivalent to claiming that all properties are essential of an individual. And if all properties are essential to an individual, that includes relational ones as well, and lo and behold, we have just concluded that individuals are worldbound.

The problem with making all properties essential is that modal logic was developed to analyse modal statements, and the predictions a framework of worldbound individuals makes with respect to those statements do not seem to match with our intuitions. Some of us may that think the statement ‘If the Twin Towers had not been destroyed in 2001, Bush would not be a president now’ is true. But whether this is true is not the point. The point is rather that most of

\[26\text{Chisholm, 1979, 85.}\]
us understand this statement as being about the actual George W. Bush, the president of the USA. It is reasonable to see this statement as a claim that had the Towers not been destroyed, things would have gone otherwise for this very individual, George W. Bush (not to mention many other actual-world individuals whose lives would have been different). This reading, however, seems to require George W. Bush to exist in more than one possible world, and that – the possibility of a transworld individual – is what both Chisholm and Lewis argue against.

One may be tempted to point out that if what is needed to stop the argument using gradual changes is a property unique and essential of each individual, we have such properties. They are called ‘haecceities’. So why could we not use them? The problem with haecceities is that they could not help decide which properties are essential and which are contingent because haecceities are not descriptive. The argument Chisholm makes could be made even if we assumed that Shakespeare and Sir Francis both retain their respective haecceities.

Chisholm27 thus seems to prove his point – by allowing the identity of an entity to be preserved through changes, we open the door to some counterintuitive results. Chisholm suggests that in order to prevent these results we should ban all changes in an individual’s properties on pain of loss of identity. In the following section we shall see that the problem with his argument might be that it just proves too much.

3.1.6 Can We Survive a Change?

Lewis28 suggests that we should think of possible worlds as analogical to moments in time. And as a matter of fact, Chisholm’s argument is parallel to the well-known puzzle of the ship of Theseus, which deals with the issue of identity of individuals and change over time.

We shall now briefly review the Theseus puzzle, and then see which

---

28 Lewis, 1979a, 184.
of its lessons could be applied to Lewis's position and Chisholm's argument.

This puzzle has been around since antiquity, but for our purposes we shall consider two somewhat modernised versions.

According to the first scenario, Theseus sailed away with a complete supply of new parts as his cargo. While at sea, he gradually replaced each part of his ship with a new one, and threw the old parts overboard. Did he return to the harbour on the same ship on which he left? Of course it had changed, but was it the same ship?

The second scenario is just like the first, except that Theseus was followed in another boat by a Scavenger who picked up all the pieces as Theseus threw them overboard, and used them to build a boat. The Scavenger then reached the port in a ship composed of the selfsame parts that the ship of Theseus had been composed of when it left the port. He docked his ship next to the ship on which Theseus arrived. Which of the two ships was then the ship of Theseus?

What are our options? Let $A$ be the ship Theseus started his voyage on, $B$ the ship he finished it on, and $C$ the ship the Scavenger finished his trip on. If we were to make a sameness of parts a necessary condition of identity, then $A$ would be identical with $C$. That would imply that Theseus changed ships during his voyage, because he started in $A$ and ended in $B$. But we know from the story that Theseus stayed on one ship during his whole journey. Alternatively,

---

29The story first surfaces in print in Plutarch (Vita Thesei, 22-23) in the following form: “The ship wherein Theseus and the youth of Athens returned had thirty oars, and it was preserved by the Athenians down even to the time of Demetrius Phalereus, for they took away the old planks as they decayed, putting new and stronger timber in their place, insomuch that this ship became a standing example among the philosophers, for the logical question of things that grow; one side holding that the ship remained the same, and the other contending that it was not the same.” This reference was found at http://faculty.washington.edu/smcohen/320/theseus.html in October 2003.

30The presentation of these two versions was inspired by http://faculty.washington.edu/smcohen/320/theseus.html, obtained in October 2003.
we could claim that since he did not change ships, $A$ must be identical with $B$, although they have no parts in common. That would also imply that $C$ is not identical with $A$ although it has all of its parts. Both suggestions seem equally counterintuitive.

One might suggest that spatiotemporal continuity is a better criterion of identity than identity of parts. Using this criterion, Theseus arrived at port in the same ship in which he left the port, and $A$ is therefore identical with $B$. On the other hand, there are cases when objects are taken apart, and then put together without, at least on the face of it, losing their identity. If I send a piece of antique furniture for restoration, the piece will be taken apart, exist for a while only as disjoint pieces of wood, and yet I would be very upset if the restorer informed me that the piece of furniture I left with him is gone. I would sue him. Yet, the spatiotemporal continuity that of piece of furniture has been interrupted.

The lesson seems to be that sometimes the criteria we normally use to determine identity break down, and that clear-cut universal criteria are vulnerable to counterexamples. Yet we usually seem to agree that things and individuals can retain their identity through some changes. The problem of identity through change is real, and cannot be brushed aside by claiming that objects do not persist through change.

Lewis is reluctant to accept any form of essentialism. He objects to a distinction between necessary and contingent properties because he thinks it is arbitrary. Yet by adopting worldbound individuals, he makes in effect all properties essential. As we already stated, in Lewis’s framework, any change in an individual leads to a change of identity. This framework is extremely general – by adopting various kinds of restrictions on the possible worlds under consideration and on the counterpart relation,, Lewis can model various kinds of modalities (for example, reasoning with logically possible worlds, physically possible worlds, epistemic alternatives). Yet underlying this flexibility, it remains the case that, strictly speaking, individuals are worldbound, and any non-actual scenario, regardless of how small the divergence from the actual world, creates a new possible world and a new possible individual. The Delia Smith who bakes
her apple-pie to perfection and the Delia Smith who burns it are, in this framework, seen as two distinct individuals.

Returning to Chisholm’s example: yes, there are real problems that arise in connection with transworld identity of individuals. But these problems are not specific to working in one or another possible-world framework – they are the old problems of persistence through change in a new guise. A move to block the persistence of identity through change is as unacceptable when dealing with transworld identity as it is when dealing with persistence through time. Lewis offers a principled solution to puzzles like that of the ship of Theseus but his solution is too strong. Perhaps the questions of identity and persistence through change are cannot be reduced to black and white. Perhaps they come by their very nature shades of grey.\(^{31}\)

### 3.1.7 Looking Into Possible Worlds

Perhaps the problem of transworld identity arises out of a particular way of looking at – or into – possible worlds. Maybe it is a by-product of looking at possible worlds as if they were somehow out there, a view supported to some degree by the way Lewis defines possible worlds and the counterpart relation.

We saw that on Lewis’s view, possible worlds are just as real as our world is, and at least some of them are inhabited by individuals of flesh and blood. At least some of these individuals bear a special relation to individuals in the actual world – they are their counterparts. But how do we determine what is the counterpart of, for example, my dog Mambo, in a given counterfactual world? Intuitively, we should find in that possible world the individual (or individuals) that most closely resemble the actual dog. Kaplan describes this process\(^{32}\) as looking through a ‘Jules Verne-o-scope.’ But suppose that in the possible world we are interested in, dogs have evolved to look just like humans. Which of those human-like individuals is my dog’s counterpart? To find out, we take the Jules Verne-o-scope,

---

\(^{31}\)We shall return to the question of identity in section 5.5 and 5.6 and explain somewhat cryptic remark concerning the colour grey.

\(^{32}\)Kaplan, 1978, 93.
examine the human-like dogs, and decide which of them is resembles Mambo most closely. As strange as it sounds, on Lewis’s view, given that the counterpart relation is a relation of similarity, the problem of locating counterparts is real.

A number of authors, Kripke\(^{33}\) and Richards\(^{34}\) among others, argue that locating the counterparts is a pseudoproblem, generated by the ‘telescope’ view itself, and not by anything in the nature of modality. Kripke points out that possible worlds are our stipulation, rather than something we discover, some far-away lands. Richards raises a related point: How is it possible for us to know anything about those possible worlds given they are ‘out there’, independently of our mental activity, causally and spatiotemporarily inaccessible to us? He says

[Lewis’s] truth-conditions are such that, for any given modal statement, it is impossible in general to determine whether they are met and hence whether the statement is true. There is, however, a certain measure of agreement between people about the truth-value of certain modal statements. Insofar as there is agreement one must assume that if it is not cat-echized into the populace without any understanding of any truth-conditions for these modal statements, then there is some other account of truth-conditions for these modal statements, and these truth-conditions are such that we may with some degree of confidence determine whether or not they are met.\(^{35}\)

If Lewis intends his possible-world framework to give an account of the meaning of modal sentences, and it seems he does, then possible worlds should be such that they can be used in giving an intuitively plausible explanation of why modal sentences function the way they do. In interpreting modal sentences we exhibit a degree of knowledge of what is possible and what is impossible, and that knowledge is

\(^{33}\)Kripke, 1980, 44.

\(^{34}\)Richards, 1975.

\(^{35}\)Richards, 1975, 109. We should understand Richards’s ‘truth-conditions’ in a Davidsonian way, where a sentence’s truth-conditions form the core of its compositional meaning.
based on our acquaintance with the actual world. Why then should we posit strongly possibilist Lewisian possible worlds at all? Lewis would probably answer that we set out trying to explain our modal notions and our explanans should be independent of our explanandum. Saying that our intuitive understanding of modal statements in any way determines what possible worlds can be, is putting the cart before the horse. This may well be a correct objection, but still, in view of the difficulties Lewis’s conception faces, we should perhaps try to find a new equilibrium between explaining modal notions, and giving an intuitively plausible analysis of counterfactuals. Actually, we shall see that extreme possibilism causes problems even in Lewis’s analysis of counterfactuals. Let us look at it now.

3.1.8 Counterfactuals and Worldbound Individuals

Lewis provides a well-known analysis of counterfactual statements. It is supposed to be one of the strong aspects of his theory. A statement like ‘Had I known how long the tram is going to take, I would have taken the bicycle’ is not explicitly modal, but it is not a claim about how things have actually turned out either. In general, statements of the form ‘If it were (had been) the case that \( p \), it would be (would have been) the case that \( q \)’ are about how things could have been, and we can think of them as modal.

Unlike other modal statements, counterfactuals do not require quantification over all possible worlds. In fact, the number of possible worlds that have to be considered is in most cases quite limited. My statement ‘Had I known how long the tram is going to take, I would have taken the bicycle’, should not be taken to mean that in any world where I know how long the tram takes I take my bicycle. In some worlds, it might be snowing too hard to ride a bicycle. Lewis would say that my claim concerns only those possible worlds that are quite ‘close’ to the actual world. My statement is supposed to

\[ ^{36} \text{The following argument is adapted from Lewis, 1973, 33-34.} \]

\[ ^{37} \text{Lewis, 1986, 20-23.} \]
be evaluated in those worlds that are as much like the actual world as is compatible with me knowing how long the tram takes, and it says that in those worlds, I use the bicycle.

On the face of it, this analysis seems attractive. The problem is that our description is not accurate. Strictly speaking, Lewis would analyse the counterfactual statements as being not about me and the bicycle, but as about my counterpart and the bicycle’s counterpart. And this does not seem to accommodate our pre-philosophical intuitions about counterfactuals. The statement above could have been followed by an exclamation ‘I should have used the bicycle, and that’s what I’ll do the next time!’ Our understanding of counterfactuals can lead to changes in our behaviour, and result in joy or regret. If, however, counterfactual statements are about our counterparts, and not about us, why should we care?\textsuperscript{38} As Loux says,\textsuperscript{39} “on Lewis’s view, things could have indeed gone otherwise, but they could not have been different for any of the individuals existing in our world.”

\textbf{Worldbound Individuals and Proper Names}

The worldboundedness of individuals has dramatic consequences for the semantics of proper names. As we just noted, nothing could have been different for actual individuals. If we stay strictly within Lewis’s framework, we cannot make good sense of the notion of a rigid designator – it collapses. Lewis does not offer us a useful kind of framework for dealing with issues of reference of proper names. His framework cannot model the tests used to tell for or against theories of proper names: Would Moses still be Moses had he not done anything ascribed to him in the Bible? By Lewis’s light, the answer is trivial. It will always be negative. Lewis offers us no guidance here. His framework seems to be of little if any use when it comes to the semantics of proper names. This framework is, as we admitted, very flexible, but its generality, and Lewis’s unwillingness to adopt any sort of essentialism, also leads to limitations. The main limitation we have encountered is the worldboundedness of

\textsuperscript{38}Kripke, 1971 voices a similar concern.

\textsuperscript{39}Loux, 1979, 42.
individuals, which results in the inapplicability of rigid designation.

One could, of course, brush this kind of criticism aside as dogmatic, and claim that the counterpart relation is basically a Lewisian equivalent of identity, and that thanks to the generality of his framework, we can restrict it in any way we wish. We could, for example, restrict it so as to model an actual speaker’s intuitions about potentialities of individuals as they are reflected in her judgements about the truth and falsity of modal statements. This would stop the counterintuitive examples in which we look with a ‘Jules-Verne-o-scope’ for my dog’s counterparts among humanlike creatures. But even if the counterpart relation was restricted in an intuitively plausible way, the worldboundedness of individuals would be unchanged. It is directly a result of the extremely realist position Lewis adopts. And that seems to be at the core of Lewis’s approach to possible worlds. A change in Lewis’s conception of individuals would result in far-reaching changes to the whole framework. We have seen all through the preceding sections that Lewis’s theory has an amazingly strong internal coherence. It is impossible to remove the more objectionable parts and preserve the parts that are attractive. Worldbound individuals are not attractive but they cannot be removed without transforming the whole framework beyond recognition.

3.1.9 Conclusion

We have now presented and analysed Lewis’s views on possible worlds and possible individuals. In the first two subsections, we reviewed the basic tenets of Lewis’s theory of possible worlds and the theory of counterparts. We focused on reconstructing the mutual relations between the theses of each of these theories, as well as Lewis’s motivation for adopting them. It is now hopefully clear in what way are both the theory of counterparts and the indexical analysis of actuality a natural outgrowth of Lewis’s position on possible worlds.

In particular, we have shown that one can arrive from theses (2), (3) and (4) of the theory of possible worlds at the conclusion that possible worlds cannot overlap or share individuals, and that no
individual can therefore be in more than one world. Then we only needed to add two very intuitive premises (which Lewis lists in the tenets of the theory of counterparts), namely that every individual is in a world, and that something is actual, in order to derive the whole of the theory of counterparts from the theory of possible worlds.

In section 3.1.3 I tried to show why we should not be too quick to dismiss the indexical analysis of actuality. I proposed a hypothetical ‘naive’ or ‘intuitive’ theory of actuality, developed it a little to see how it would deal with possible worlds, and concluded that such a theory quickly looses its intuitive plausibility once it is put to work. It may be difficult to propose an analysis of actuality that would work as well as Lewis’s indexical analysis does.

Because the counterpart relation is a relation of similarity, not of identity, and everything is like anything else in some respect, there is no reason for Lewis to adopt any sort of essentialism into the general form of his theory. One can restrict the counterpart relation in various ways depending on a particular purpose, but these restrictions are not motivated from within the framework – they are a consequence of the use to which we put the possible-world framework (we can, for example, consider only possible worlds that work in accordance with the laws of physics that hold in the actual world, and restrict the counterpart relation accordingly). Because the counterpart relation is not about identity, Lewis does not need to say anything about criteria of transworld identity.

Finally, we went through a number of objections to Lewis’s theory. They all targeted consequences of the counterpart theory. We saw that as it is, Lewis’s theory implies a somewhat unusual form of essentialism, according to which all properties are necessary. We also saw that Lewis’s framework is not well suited for analysing semantic properties of proper names.

Part of our goal was to demonstrate the internal cohesion of Lewis’s framework. This we did, and we arrived at the conclusion that to remove the theory of counterparts, which does seem to have rather unpleasant consequences, we would have to alter the very core of Lewis’s theory.
3.2 Kripke’s Approach to the Semantics of Modal Statements

In the preceding sections we analysed the problems connected with a strongly realist approach to possible worlds. In the following sections, we focus on a proposal that is markedly less realist about possible worlds, and does not support the notion of worldbound individuals. In this subchapter, we shall reconstruct, as best we can, Kripke’s approach to possible worlds, his stance with respect to haecceitism and essentialism, and investigate the role various kinds of essentialism play in his overall plan. Clarifying Kripke’s position with respect to these issues should help us evaluate his proposal concerning the semantics of proper names and the metaphysics of modality.

Saul Kripke’s first notable achievement was to provide a general framework that treats various systems of modal logic as variations within a common framework.\(^{40}\) Using the notions of a model, accessibility, validity in a model, and a possible world, Kripke gave a unified semantics to the four main modal systems (system M or K, and systems B, S4 and S5). Where there had been a plethora of seemingly unrelated systems, he brought order and unity.\(^ {41}\) He then went on to apply some of the insights from his work on formal systems to natural language, and it is his work on proper names\(^ {42}\) that we shall focus on right now. A substantial part of his work on the semantics of natural language is focused on the semantics of modal statements that involve individuals. In this analysis, he frequently uses possible worlds; in fact, his famous definitions of necessity, rigidity, and essential property are all given in terms of truth in possible worlds. It is then rather surprising that when we turn to his texts, we find that very little is explicitly said about the notion of a ‘possible world’ in the context of natural language anal-

---

\(^{40}\) Kripke, 1963.

\(^{41}\) He was not the only one. The work of Richard Montague, Stig Kanger, Ruth Barcan, Dana Scott, and Jaako Hintikka has to be mentioned in this context as well.

\(^{42}\) Kripke, 1971 and Kripke, 1980.
ysis. This may be a consequence of Kripke’s formal roots, but in our present inquiry, a formal definition of a possible world leaves many questions open. Filling in the picture of Kripke’s notion of possible worlds as it is used in his natural language analysis, that will be our next immediate task.

3.2.1 Stipulating Possible Worlds

It is not difficult to identify some views on possible worlds which Kripke does not entertain. We have already briefly mentioned his criticism of what he calls ‘the telescope view’. In essence, Kripke criticises Lewis for treating possible worlds as if they were disconnected from the actual world, which he makes clear by pointing out that on a Lewisian approach,

...one thinks, in this picture, of a possible world as if it were like a foreign country. One looks upon it as an observer. Maybe Nixon has moved to the other country and maybe he hasn’t, but one is given only qualities. One can observe all his qualities, but, of course, one doesn’t observe that someone is Nixon...So we had better have a better way of telling in terms of properties when we run into the same thing as we saw before; we had better have a way of telling, when we come across one of these other possible worlds, who was Nixon.44

In Kripke’s view, the problem results from Lewis’s insistence that possible worlds be given by qualitative descriptions. It is the kind of description we have of a Lewisian possible world that makes us feel like observers in a foreign country, or – in Kripke’s idiom – as if we were ‘viewing through a telescope.’45 On a Lewisian view, we

---

43In section 3.1.7.
44Kripke, 1980, 43-44.
45Kripke, 1980, 44. Actually, the telescope metaphor may have originated with David Kaplan Kaplan (1979c). ‘Transworld Heir Lines’, though published after Kripke’s 1971 Naming and Necessity lectures, was presented as a lecture in 1967. In both the lecture and the paper, Kaplan repeatedly uses the notion of a ‘Jules Verne-o-scope’.
know what things look like in a possible world, but not what they are. This, Kripke goes on to say, not just fails to reflect the way we intuitively interpret counterfactuals—it breeds unnecessary and misleading problems concerning the identity of individuals. This was a view we arrived at in our analysis of Lewis’s framework as well. But let us look now at what Kripke’s response to the problem is.

According to Kripke, we do not know how to give sufficient and necessary qualitative conditions for identity of material objects or persons even within one world, and yet in Lewis’s framework the qualitative criteria are the only way of relating counterparts in different possible worlds.

Kripke suggests that intuitively we interpret a counterfactual like ‘Nixon could have lost the election’ by bringing in a possible world where Nixon lost.

‘Possible worlds’ are stipulated, not discovered by powerful telescopes. There is no reason why we cannot stipulate that, in talking about what would have happened to Nixon in a certain counterfactual situation, we are talking about what would have happened to him.

It is a crucial part of Kripke’s proposal that one be able to stipulate, as a part of a description of a possible world, which individuals are involved in it, and that this can be done not by giving a qualitative description, but in some—yet to be specified—more direct way. Let us try to specify it now.

---

46See for example Kripke, 1980, 43: “Mathematics is the only case I really know of where [adequate necessary and sufficient conditions for identity] are given even within a possible world, to tell the truth. I don’t know of any such conditions for identity of material objects over time, or for people. Everyone knows what a problem this is. But, let’s forget about that. What seems to be more objectionable is that it depends on the wrong way of looking at what a possible world is.”

47Kripke, 1980, 44.
3.2. Kripke’s Approach to the Semantics of Modal Statements

3.2.2 Two Versions of Haecceitism

We have just seen that in Kripke’s view, it is admissible to specify a possible world by stipulating which individuals it contains. Actual-world individuals can figure directly in a description stipulating a non-actual possible world, and can, consequently, exist in more than one world. We have noted that individuals do not have to be specified only qualitatively but can also, for example, be specified using ostension or by their name. We can therefore make within Kripke’s framework good sense of an individual retaining its identity while its properties change, and that holds both in the actual world and in the non-actual ones.

In our analysis of the identity puzzles concerning the ship of Theseus, and the case of Shakespeare and Bacon, we saw that a Lewisian response was to ban all changes of properties of individuals, which in effect made all of their properties essential. Kripke opposes this view. According to him, we can meaningfully ask whether certain statements concerning an actual-world individual would hold of that very same individual in a counterfactual situation where that individual’s properties have changed.

Kripke admits that for most kinds of entities, we do not have available descriptions that would provide the necessary and sufficient criteria of their identity. We should then assume that there is some presumably non-descriptive fact of the matter in virtue of which we can tell whether an individual in a counterfactual situation is identical to the actual-world individual we are interested in. Nathan Salmon convincingly argues\(^{48}\) that a stipulation of possible worlds in terms of actual-world individuals requires at least the adoption of haecceitism.\(^{49}\) Let us have a look at why this should be the case:

Let us start by reminding ourselves of what we mean by haecceitism. Two passages from Kaplan’s work are usually brought forward to define it:

\(^{48}\)Salmon, 1986.

\(^{49}\)It is also compatible with adoption of even more ambitious essentialist doctrines, as we shall show shortly.
out reference to common attributes or behavior - whether this is the same individual in another possible world, that individuals can be extended in logical space (i.e., through possible worlds)... and that a common ‘thisness’ may underlie extreme dissimilarity or distinct thisnesses may underlie great resemblance, I call Haecceitism.

Haecceitism holds that we can meaningfully speak of a thing itself - without reference... to individuating concepts (other than being this thing), defining qualities, essential attributes, or any other of the paraphernalia that enable us to distinguish one thing from another. It may be that each thing has essential attributes with which it is vested at all times and in each possible world in which it exists. But that is an issue posterior to whether things have trans-world being.50

In my view, these passages characterise two related but distinct doctrines.51 The first passage outlines a sort of non-qualitative essentialism, according to which an individual’s identity across possible worlds is warranted by a primitive thisness. The second passage suggests something more modest: that it is possible to refer to an individual without taking recourse to any particular means of identifying it. This view, at least on the face of it, is not essentialist. For the time being, I shall call the first view full-blown haecceitism, and the second modest haecceitism.

In Kripke’s work, we find an analogical bifurcation.52 On the one hand, Kripke seems to endorse a modest, non-essentialist haecceitism. The following passage suggests as much:

Philosophers... have asked, are there objects behind the bundle of qualities, or is the object nothing but the bundle? Neither is the case; this table is wooden, brown, in the room,

---

50Kaplan, 1975, 722-723.
51Salmon, in Salmon (1986), discusses various versions of haecceitism. My treatment of this topic is influenced by his views but I use the analysis for a different purpose.
52This issue is connected to Kripke’s endorsement of essentialism but is a separate one.
etc. It has all these properties and is not a thing without properties, behind them; but it should not therefore be identified with the set, or ‘bundle’, of its properties, nor with the subset of its essential properties. 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.\textsuperscript{53}

Adopting a somewhat Wittgensteinian tone, Kripke is trying to ‘dissolve’ an apparent problem, and to show that an object is neither a ‘bundle of properties’ nor anything behind it. On the other hand, in his more essentialist mood, he says that “(roughly) being a table seems to be an essential property of the table.”\textsuperscript{54}

In order to see whether we can reconcile these two views, we have to look at the broader context of Kripke’s work. By reconstructing the role haecceitism is supposed to play, we can draw conclusions as to which version would fit the bill.

**Haecceitism? Yes, but which one?**

As we saw in Kripke’s criticism of Lewis, being able to describe possible worlds by stipulating non-descriptively which individuals are involved in them is supposed to result in a more intuitive view of possible worlds. Adoption of transworld individuals will help us leave behind the wrong - Lewisian – picture, according to which possible worlds are like ‘foreign countries’. Both the modest and the full-blown version of haecceitism could do this job. We might hold – as in full-blown haecceitism – that an haecceity is an essential property unique to each individual, or – as in the modest version – that haecceitism amounts to making it possible to pick out an individual without recourse to a qualitative description, but does not amount to a commitment to an underlying essential property, haecceity. In order to make his notion of transworld individual feasible,
Kripke needs at least the modest haecceitism but his views seem compatible with full-blown haecceitism as well.

The issue of haecceitism is closely connected with the issue of identity of individuals across possible worlds – something like haecceitism is needed for the notion of transworld individuals to work. And the adoption of transworld individuals is crucial for the workings of rigid designation, because – as we know – rigid designators are supposed to denote the same referent, the same individual, in all possible worlds. We can thus look at the notion of rigid designation and try to find out which version of haecceitism it necessitates.

In order to make the concept of rigid designation work, we have to be able to distinguish between the situations where there is a particular individual, say Bob, in some possible world, and those situations where this only seems to be the case. In Kripke’s worlds, not everything that looks like Richard Nixon is Richard Nixon: if, for example, Richard Nixon* were, in a non-actual world, an automaton fantastically resembling the actual-world Richard Nixon, who is a human, Richard Nixon and Richard Nixon* cannot be the same individual. We can be presented with a qualitatively defined world that contains an individual closely resembling the actual-world Richard Nixon, and be asked whether such a world indeed contains Richard Nixon. And that is why even in Kripkean possible worlds we still need cross-world identity criteria. They are needed to play the role of truth-warrants, to fix the truth-values of sentences concerning individuals (e.g., Richard Nixon), thus giving the question whether Richard Nixon is in the domain of individuals of a particular possible world a determinate answer.

Kripke proposes a number of essential properties that introduce some necessary conditions on an entity’s identity. Some of the essential properties take the form of conditions on the constituent parts of an entity. These properties provide some necessary but

55Kripke, 1980, 48.
56For example originating from a particular hunk of matter, and having the parents an individual in fact has.
57Being made of a particular hunk of matter, or being made up of atoms that have a particular atomic number are clear examples of giving identity criteria in
not sufficient criteria of identity for some kinds of entities. This, Kripke says, just reflects our state of knowledge: we do not know yet what the sufficient criteria of identity for most kinds of entities should look like.\textsuperscript{58} Specific essential properties of the kind mentioned here cannot therefore consistently do the job of telling apart those possible worlds where an individual occurs from those where it only may seem so. And this is where haecceitism comes in. In fact, one could see it as a tool designed to do just that. Haecceitism provides the identity criteria we needed to put flesh on the concept of a transworld individual, and, indirectly, the notion of rigid designation.

Is it the modest or the full-blown haecceitism that can play this role? What we need for rigid designation to work, is necessary and sufficient criteria of cross-world identity, something an individual has in every world where it exists. Kripke proposes that in every possible world in which it exists, an entity has the property of \textit{being the very entity it is}.\textsuperscript{59} This is not the property of self-identity, which trivially applies to every object. On the contrary, this property is unique to each entity: for every entity $x$, only $x$ has the property of \textit{being $x$}.

As commonly understood,\textsuperscript{60} a property is essential to an entity when the entity cannot fail to have it if it is to exist. It follows then that the property we have outlined in the last paragraph is an essential one. That is why we can conclude that the haecceitism Kripke needs has an essentialist import, and a weaker, modest version will not do. Kripke needs not only haecceitism, but haecceities as well.

\textsuperscript{58}Compare the passage from Kripke, 1980, 43, quoted in a footnote on p. 90.

\textsuperscript{59}This is, I believe, how we should read the passage in Kripke, 1980, 114, footnote 57, where he says that “(roughly) \textit{being a table} seems to be an essential property of the table.”

\textsuperscript{60}See p. 31.
3.2.3 Haecceities and Reduction

What kind of property is a Kripkean haecceity? Should we understand it as a primitive or can it be further analysed? In part, we can find a Kripkean answer in what was already stated: given our state of knowledge, and the preconditions of rigid designation, haecceities are the best we can do at the moment. Since this is co-determined by our current epistemic situation, it could change. Kripke seems to think that with respect to at least some kinds of entities, science can provide necessary identity criteria in terms of conditions on the constituent parts of those entities. In other cases, Kripke admits that the very conceptual possibility of a reductive analysis is an open question:

Although the statement that England fought Germany in 1943 perhaps cannot be reduced to any statement about individuals, nevertheless in some sense it is not a fact ‘over and above’ the collection of all facts about persons, and their behavior over history... Similarly, perhaps, facts about material objects are not facts ‘over and above’ facts about their constituent molecules... In each case we seek criteria of identity across possible worlds for certain particulars in terms of those for other, more ‘basic’ particulars. If statements about nations (or tribes) are not reducible to those about other more ‘basic’ constituents, if there is some ‘open texture’ in the relationship between them, we can hardly expect to give hard and fast identity criteria...

Whether statements about one kind of entities are reducible to statements about another kind of entities is, at least in some cases, an open question. Its solution depends on there being bridging laws between those kinds of entities. And even when an exhaustive de-

---

61 For example in Kripke, 1980, 44, we read that “…characteristic theoretical identifications like ‘Heat is the motion of molecules’, are not contingent truths but necessary truths, and here of course I don’t mean just physically necessary, but necessary in the highest degree – whatever that means.”

62 Kripke, 1980, 50.

63 This is not all that needs to be said on the role of science within Kripke’s framework, but we shall return to this subject shortly.
scription of an object in terms of its constituent parts is available, it may be more practical to speak in terms of nations rather than individuals when discussing history, or in terms of tables and chairs instead of molecules when ordering furniture.

In principle, Kripke might say, science may provide us one day with necessary criteria of identity for all sort of entities. But for the time being, haecceities do the job where we do not have such criteria, and are often handier even when we do have them.

Now that we saw that something like haecceities is needed to make rigid designation work, we have a somewhat better idea about the assumptions that need to be made to establish a referring relation between a designator and its referent in various possible worlds. The next question we shall try to answer is what happens in the possible worlds where the referent does not exist.

3.2.4 Persistence and Obstinacy

We have now dealt with issues connected with determining whether a referent exists in a particular possible world. In this section, we shall focus on the referring relation, in particular on the question of what happens with reference in those possible worlds where the referent does not exist.\footnote{We have given a preliminary characterisation of the notions we shall use in this section already in section 1.7.3.}

The best known definition of rigid designation tells us that a rigid designator designates the same object in every possible world,\footnote{Kripke, 1980, 48.} In another definition we are told that a rigid designator designates the same object in every possible world in which that object exists.\footnote{Kripke, 1980, 49. We introduced this definition of persistently rigid designator on p. 26.} And finally, in Kripke’s letter to Kaplan, we read that “a designator \( D \) of an object \( x \) is rigid, if it designates \( x \) with respect to all possible worlds where \( x \) exists, and never designates an object other than \( x \) with respect to any possible world.”\footnote{Kaplan, 1989b, 569, my italics. We used this definition earlier, on p. 47.} These three definitions are
clearly not equivalent. Let us have a closer look at their differences. 

We can safely assume that a rigid designator designates the same object at least in every world where that object exists. All three definitions make this claim. What is not clear is what happens in the worlds where the designatum does not exist.

To account for various positions with respect to this issue, Kaplan and Salmon\textsuperscript{68} introduce some helpful distinctions. Firstly, there are rigid designators that designate the same thing in every possible world where that thing exists, and nothing in those worlds where it does not. These are called \textit{persistent} designators. Secondly, there are rigid designators that refer to the same thing in every possible world regardless of whether their designatum exists there or not. These are the \textit{obstinate designators}. There is yet another kind of rigid designator: \textit{Strongly rigid designators} rigidly refer to something necessarily existent. These designators are, by definition, both persistent and obstinate. Kripke cites mathematical descriptions (such as ‘the smallest prime’) as examples of strongly rigid designators.\textsuperscript{69}

The question we want to answer is: what kind of rigid designators are proper names? In the present context, we are not interested in the semantics of names of mathematical entities. We shall therefore leave them aside. Having done that, it seems quite clear that the proper names we are interested in refer to contingently existing entities. The particulars that exists in the actual world might have failed to exist, and there might have been more entities than there actually are. Had my parents never met, I would not have existed. Had they met earlier than they in fact did, I could have had an older brother. The difference between my older brother and me is that I contingently exist, while he, equally contingently, does not.

Having established that the referents we are interested in are contingently existent entities, we can conclude that proper names

\textsuperscript{68}Kaplan, 1989a, and Salmon, 1982, 32-40. Salmon elaborates on a distinction proposed by Kaplan.

\textsuperscript{69}Kripke introduces this notion in Kripke, 1980, 48. He also further distinguishes contingently existent and contingently non-existent entities from necessarily non-existing ones, like Sherlock Holmes, in Kripke, 1972a, but that is not relevant to our topic here.
are not strongly rigid designators. It remains to be seen, however, whether they are persistently or obstinately rigid.

When investigating reference in various possible worlds, it is sometimes useful to use the analogy with alternate timelines. The sentence ‘Anna Pilatova is dead’ is false now but true any time after my death. If the name ‘Anna Pilatova’ did not denote anything after my demise, that is, if it were a persistent designator, the sentence above could not be true after my death. It seems thus plausible to say that the name ‘Anna Pilatova’ does denote someone with respect to the 22nd century, namely me. It is because the term has a denotation in that situation that the sentence ‘Anna Pilatova is dead’ is true with respect to this future time. We could use the same reasoning when evaluating the sentence ‘It might have been the case that Anna Pilatova was never born’. Here, again, the sentence is true because there are possible worlds in which I do not exist, and in those possible worlds the name still denotes me, the actual person.

This leads to the conclusion that proper names are obstinately rigid designators. A proper name $N$ primarily designates an actual-world entity $x$. Having its reference thus fixed, $N$ shall refer to the same $x$ also in non-actual possible worlds. In those possible worlds where $x$ does not exist, negative existential statements involving $N$ can still turn out to be true. This is an interesting observation because reference is often seen as a relation between a designator and its designatum, and it is in that form that reference is usually discussed in the literature. The cases where the designatum does not exist and reference to it is still successful make us re-think this very intuitive picture of reference.

Reference is especially complicated in the case of entities that are contingently non-existent in the actual world. Their reference can only be introduced by means of description (as in ‘George, my older brother’ or ‘Jacques, the last Frenchman’), and while the proper names thus introduced are rigid, the identity of the referents is not as clear as the identity of entities that exist in the actual world.

\footnote{The following two examples, as well as the point they argue for, are adapted from Salmon, 1982, 37-39.}
For example, it might turn out that the description, by which the contingently non-existent referent was introduced, fails to pick out a unique entity. It seems that in such a case there would be no fact of the matter as to who the referent is. Seemingly contingently non-existent entities could also at closer inspection turn out to be necessarily non-existent (as is claimed for fictional entities). But that depends on the treatment of fictional entities one adopts.\footnote{Kripke clarifies his view of fictional entities in Kripke, 1980, 24, where he says: “So it is said that there might have been unicorns. And this is an example of something that I think is not the case. I think that even if archaeologists or geologists were to discover tomorrow some fossils conclusively showing the existence of animals in the past satisfying everything we know about unicorns from the myth of the unicorn, that would not show that there were unicorns.”} Reference to contingently existing entities differs from the reference to contingently non-existent entities because of the difference in our access to them. Our prejudice in favour of the actual may turn out to be rather a statement of sober recognition of the perils we can encounter in the realms of the merely possible.

### 3.2.5 Possible Worlds and Imagination

The case of the unicorn draws our attention to the limits of stipulating possible worlds by the means of specifying which entities occur in them. The fact that we can describe possible worlds this way should not make us think that anything we can imagine is genuinely possible. In Kripke’s view, possible worlds are not created by our imagination: what is and is not possible is given by the modal properties of actual-world entities. We already know that in Kripke’s view, necessity and contingency apply not just to statements, but to properties as well.\footnote{Kripke, 1980, 41.}

Modal properties are described in terms of possible worlds, so that, for example, an object has a property necessarily if it has that property in every possible world where it exists. Possible worlds were introduced to model modal properties of actual-world entities, and that is why it comes as no surprise that modal properties of...
actual-world entities place certain conditions on what goes on in these possible worlds. We can thus turn Kripke’s analysis of necessity around and say that only those worlds are possible where the properties which are necessary in the actual world hold.

There are many kinds of necessity, which result in different kinds of commitments, and the necessity that Kripke deals with results in metaphysical commitments.\textsuperscript{73} Kripkean necessity is about \textit{how things are} irrespective of our knowledge, language, or context. It may well be the case that there are many more necessary properties than we shall ever know, and that these unknown properties place restrictions on possible worlds just as much as those properties we think we have already identified.\textsuperscript{74}

Therefore, while we can imagine worlds where some properties that are necessarily in the actual world, e.g., some laws of physics, do not hold, Kripke would say that these are not really possible worlds. They are just figments of our imagination. We may call them impossible possible worlds or Oscar and Felix, but they have no place in Kripke’s ontology. In Kripke’s world, only those worlds are possible where everything that is necessary in the actual world holds. In this sense, Kripke is an actualist: what is possible is determined by the modal properties of the actual world. Necessary statements thus seem to play an important role in building Kripke’s possible world framework. It is time we paid closer attention to them.

\textbf{3.2.6 What is \textit{a priori}?
}

As we just noted, a world’s possibility is co-determined by modal properties of actual-world entities. Necessary statements play an

\textsuperscript{73}The robustness of Kripke’s concept of necessity is illustrated for example in Kripke, 1980, 142: “Any necessary truth, whether \textit{a priori} or \textit{a posteriori}, could not have turned out otherwise.”

\textsuperscript{74}On the other hand, Kripke says that “A possible world is \textit{given by the descriptive conditions we associate with it.}” Kripke, 1980, 44, italics in the original. This seems to clash with the view which I argue Kripke adopts. The way we should see it is that in this passage Kripke emphasises the difference between Lewis’s approach and his view, which he later explains in more detail.
important role here: for a world to be genuinely possible, all necessary statements must have the same truth-value in that world as they have in the actual world. Traditionally, philosophers have considered only those statements to be necessary which were known a priori. These statements were then said to be true everywhere if true at all. On this traditional view, necessary statements were those that could be known by reasoning alone. The subject matter of these a priori necessary statements were certain ideas, not things in the world, which were seen as contingent. It is well known that Kripke rejects this approach, and argues that a posteriori necessary statements are not only possible, but can be known as well.

The claims concerning a posteriori necessary statements are novel, controversial, and essential to Kripke’s enterprise. Before we start looking at particular statements Kripke claims are a posteriori and necessary, we shall examine what he means by a priori and a posteriori, and how he argues for the separation of the epistemic and the metaphysical distinctions.

Kripke presents his clearest and least presupposition-loaded argument for the separation of the epistemic and the metaphysical distinction, as well as for the existence of a posteriori necessary statements, using an example from mathematics. Let us briefly review it. Consider Goldbach’s conjecture. It says that every even number greater than 2 is a sum of two primes. We do not know whether this conjecture is true. None of us therefore has any a priori knowledge in this respect. When someone finds a proof of Goldbach’s conjecture and we come to believe it, it will be to us a new piece of information. It will be a posteriori evidence. Now notice, so Kripke urges, that, regardless of our ignorance, if the conjecture is true it is necessarily true, and if it is false it is necessarily false because the truth-value of a mathematical statement is not contingent. Our

\footnote{For example Duns Scotus has been interpreted as saying just that.}
\footnote{This had a lot to do with the religious assumptions made by medieval and early modern thinkers.}
\footnote{We have briefly introduced this subject in section 1.7.3. Here we assume the content of that section, and expand it.}
\footnote{The argument is given in Kripke, 1980, 38-38.}
3.2. Kripke’s Approach to the Semantics of Modal Statements

ignorance of a particular statement’s truth-value has no bearing on its truth-value. This is the gist of Kripke’s argument.

Before we go on evaluating this argument, we have to take a closer look at Kripke’s use of the terms a priori and a posteriori. This is because – as we shall see – Kripke uses those terms in a rather non-standard way.

When introducing the notion of a priori, Kripke quotes Kant’s definition of it, which says that a priori truths can be known independently of experience. Commenting on this definition, Kripke says that if something can be known a priori, we should ask ourselves for whom it is possible to know it in that way. In other words, Kripke endorses the view that a prioricity is relative to the knower. This leads him to say that

\[\ldots\text{it might be best therefore, instead of using the phrase ‘a priori truth’, to the extent that one uses it at all, to stick to the question of whether a particular person or knower knows something } a \text{ priori \ or \ believes it true on basis of } a \text{ priori evidence.}\]

We have seen above that Kripke thinks of the notion of a priori as describing the relation between a piece of information and a knower. In this quotation, however, he applies the notion to evidence. The obvious question is: What counts as a priori evidence? Kripke elucidates the concept in the following example. A person who works with a computer knows that the computer can answer whether a particular number is prime. No person has calculated that this particular number is prime, but the machine gave us the answer. If we then believe that this particular number is indeed prime, we believe it on the basis of our knowledge of the capacities of the computer. It seems therefore that we believe it on a posteriori grounds. Nonetheless, someone who made the requisite calculations himself could believe a priori that the number in question is prime. But

\text{\textsuperscript{79}}Kripke, 1980, 34.

\text{\textsuperscript{80}}Kripke, 1980, 35.

\text{\textsuperscript{81}}Kripke, 1980, 35.
that does not mean that person is necessarily right - one can make mistakes in a priori reasoning.

Something can be known, or at least rationally believed, \textit{a priori}, without being quite certain. You’ve read a proof in the math book; and, though you think it’s correct, maybe you’ve made a mistake... You’ve made a computation, perhaps with an error.\textsuperscript{82}

It seems that according to Kripke, a belief is a priori if it is based on a priori evidence, and evidence is a priori if it is available to the agent without recourse to external fact checking. This notion of a prioricity is inherently context-relative. Something that is a priori for one person need not be so for another one. This notion is very different from that which was used by Kant.

The underlying reason for the divergence between the traditional notion of a priori and Kripke’s notion may well lie in Kripke’s approach to analyticity, and indirectly, his approach to necessity. In the Kantian tradition, a statement is analytic if its truth can be determined by analysis of the terms involved alone. The necessity associated with a prioricity is thus intended to be primarily of an epistemic, but secondarily also of a semantic kind, derived from the meaning of terms involved in a sentence. The semantic kind of necessity, i.e., analyticity, was famously attacked by Quine\textsuperscript{83} and has largely fallen into disrepute. Kripke does not use that notion. But in the traditional Kantian picture, there was a connection between analyticity and the a priori: a statement was analytically true if its truth could be known just by analysis of the meaning of the concepts involved, and the knowledge thus derived was then a priori. This sort of ‘semantic necessity’ is not available to Kripke, who ascribes necessity to facts about the world, and the statements that express them.

In the Kantian picture, a competent speaker can figure out that some statements are necessary in virtue of his command of language. In the Kripkean picture, there is no parallel to this: different

\textsuperscript{82}Kripke, 1980, 39.
\textsuperscript{83}In Quine, 1961b.
agents know different things about the world, and that is why a priori knowledge is speaker-relative. The shift of meaning of ‘a priori’ between Kant and Kripke is to a large extent a result of the shift of interest from semantics to metaphysics. In this light, it is hardly surprising that Kripke’s notion of a priori is rather thin. Because it is not connected to necessity, it does not guarantee knowledge, and because it is not connected to a competence that is generally shared between agents (i.e., language), it is speaker-relative.

3.2.7 Natural Kinds and Haecceities

When dealing with Kripke’s approach to necessity and the a priori, one cannot avoid an overview of the kinds of statements, which, as Kripke claims, can a priori be known to be necessarily true if true at all. And any such overview has to include Kripke’s analysis of natural kind terms. Though natural kind terms are not a subject of our primary interest, the connection between haecceities and reference is, which more than justifies our little excursion into natural kinds.\(^{84}\)

We shall not speculate about what he would have or should have said had he elaborated more on the topic of natural kinds. In particular, we shall not assume that any gaps one may find in Kripke’s views can be filled by Putnam’s views.\(^{85}\)

\(^{84}\)We shall deal with natural kinds in a rather cursory manner. However, many of the authors whose work on proper names we mentioned or used in our explorations have also written on natural kinds. It was mainly the work of Kripke (1980), and Putnam (1975a) that started the debate on natural kinds, but important contribution were made also by Wiggins (1980), Burge (1973), McGinn (1976), and Salmon (1982), as well as all those whose articles were collected in French, Uehling, and Wettstein (1986).

\(^{85}\)Kripke says that the views on natural kinds and substances he entertains in \textit{Naming and Necessity} (1980) have many points of contact with Putnam’s pre-1963 writings, but also that “there are some divergences between Putnam’s approach and mine.” (Kripke, 1980, 122, footnote 62).
Kinds Are Introduced

Kripke deals with natural kinds mainly in his *Naming and Necessity*,\(^\text{86}\) focusing above all on the issues of necessary and contingent properties. His aim is to refute the view that observable properties usually associated with natural kind individuals and samples – e.g., *being tawny yellow with blackish transverse stripes and white belly* in the case of tigers – are necessary of them. We could, Kripke says, imagine circumstances where none of these commonly associated properties apply to a particular individual, and yet these would still be circumstances where that individual still belongs to the kind that is thus characterised. The reverse also holds: something might have all the identifying characteristics of some natural kind, yet form a separate kind.\(^\text{87}\) Therefore, Kripke concludes, “possession of most of these properties [by which we originally identified a kind] need not be a necessary condition for membership in the kind, nor need it be a sufficient condition.”\(^\text{88}\) Contrary to Frege’s, Russell’s, and even Mill’s views, natural kind terms are not descriptive. According to Kripke, they are much more like singular terms than has been thought, and his treatment of them is largely parallel to his treatment of proper names.

In Kripke’s view, proper names are attached to individuals by a hypothetical ‘baptism’, where a description, ostension or both can be used. Natural kinds also undergo a sort of ‘baptism’ where a definition and/or ostension is used,\(^\text{89}\) as in “Gold is the substance

\(^\text{86}\)Especially in Lecture III of Kripke, 1980.
\(^\text{87}\)See Kripke, 1980, 119, and the following: “Even though we don’t know the internal structure of tigers, we suppose - and let us suppose that we are right - that tigers form a certain species or natural kind. We can then imagine that there should be a creature which, though having all the external appearance of tigers, differs from them internally enough so that we should say that it is not the same kind of thing. We can imagine it without knowing anything about this internal structure. We can say in advance that we use the term ‘tiger’ to designate a species, and that anything not of this species, even though it looks like a tiger, is not in fact a tiger.” (Kripke, 1980, 121)
\(^\text{88}\)Kripke, 1980, 121.
\(^\text{89}\)We can, just like in the case of proper names, imagine circumstances where a natural kind is baptised using a definition only. For example, a number of
3.2. Kripke’s Approach to the Semantics of Modal Statements

instantiated by the items over here, or at any rate, by almost all of them.”

This sort of definition, which is used to introduce a natural kind term, does not express a necessary truth because the natural kind would have existed even if any particular items did not. Kripke says that

\[
\text{...in general, terms for natural kinds (e.g., animal, vegetable, and chemical kinds) get their reference fixed in this way [by the above mentioned sort of definition]; the substance is defined as the kind instantiated by (almost all of) a given sample.}\]

The reference of natural kind terms is rather similar to the reference of proper names. Natural kind terms are not descriptive: their referents may fail to have any of the properties commonly attributed to them, and something may have all of those properties and not be the intended referent. The descriptions used in fixing the referent also do not form a part of their meaning. In Kripke’s view, natural kind terms are rigid designators that designate directly.

**On Being A Natural Kind**

Natural kind terms are unlike proper names in that the question of the ontological status of their referents is somewhat more complicated. Everyone has some idea of what an individual is but few people outside of philosophy have ever heard about natural kinds. And it is difficult to see what ontological status Kripke intends natural kinds to have but he does give us some clues by saying that “the original concept of cat is: *that kind of thing*, where the kind can be identified by paradigmatic instances.” He also suggests that a natural kind can be identified by its instances but not with them – a change in their number does not amount to any change in the kind itself. Natural kinds must therefore be to some degree ontologically

unstable chemical elements high in the periodic table were defined by their atomic number years before they were first synthesised.

90Kripke, 1980, 135.
92Kripke, 1980, 122.
independent of their instances, which means that they are abstract entities.

What all instances of a natural kind have to share is their essential properties. In Kripke’s view,\(^93\) it is the task of science to discover what those essential properties are. We can, however, try to reconstruct a minimal picture of the semantics of natural kind terms without making any claims about the essential properties of particular kinds.

In general, we can assume – in parallel with individuals – that natural kinds have haecceities, which could be expressed as *being this sort of thing.*\(^94\) We can further suppose that instances of natural kinds have also individual haecceities (of the form *being this very thing* or *being this very individual*). Indeed, unless we assume that we have identified the essential properties for every natural kind, it seems that we need to assume natural kind haecceities in order to make rigid designation work. The reasoning behind this is the same we used for proper names. In the case of natural kinds that are exemplified by individuals (like ‘tiger’) or samples (like ‘gold’), if we want to be able to refer rigidly to an individual or a sample, we have to assume also individual haecceities.\(^95\)

Kripke does not say anything explicitly about the relation between individual haecceities and the haecceities of kinds. However, we can recall his treatment of the Nixon case,\(^96\) where it is said that if Nixon is human, then he is necessarily so. This indicates that there is a (metaphysical) necessity relating the two kinds of haecceities. Addressing the question of mutual relations between various natural kind haecceities, Kripke says that “of many such statements [like ‘Cats are animals’], especially those subsuming one species under another, we know a priori that, if they are true at all, they are necessarily true.”\(^97\) This claim relies on further assumptions related

---

\(^93\) At for example Kripke, 1980, 138
\(^94\) See the above quoted passage Kripke, 1980, 122.
\(^95\) In the case of samples this point is somewhat stretched but we can imagine a situation where we want to refer to a particular piece of gold.
\(^96\) Kripke, 1980, 46.
\(^97\) Kripke, 1980, 138.
3.2. Kripke’s Approach to the Semantics of Modal Statements

not only to the ontology of natural kinds but also to the problematic notion of a priori. We shall analyse the presuppositions of this sort of claims in the following section.

Hierarchies of Haecceities, Tony The Tiger, and Other Beasts

We can get a better idea of Kripke’s theory of natural kinds if we apply the aforesaid to a couple of examples. This should help us organise what we have learned so far. Let us therefore conduct an analysis of the epistemological and the modal status of the following sentences using a Kripkean perspective.

(1) Dogs bark.

(2) Tony is a tiger.

(3) Dogs are mammals.

If we read sentence (1) as a generic statement, it turns out to be contingent and a posteriori in virtue of the non-descriptionality of natural kind terms. A natural kind term refers directly and rigidly to the kind, and the kind refers directly to its instances. Strictly speaking, this needs to be qualified: we cannot say that a natural kind term rigidly refers to its instances because its extension can vary. In this generic statement, ‘barking’ is predicated of items belonging to the natural kind ‘dog’. On a naive reading, a single instance of a non-barking dog should falsify the sentence, and that is clearly undesirable. Kripke does not deal in his work with the problems raised by this kind of generic sentences.

We can analyse sentence (2) as necessary a posteriori by analogy with the Nixon example. As we pointed out, belonging to a particular kind is supposed to be a part of an individual’s haecceity.

\[^98\] In a non-actual possible world, the extension of a natural kind term may be different than in the actual world. If the definition of rigid designation demands that a rigid designator refers to the same thing in every possible world where that thing exists, we would have to amend the definition to account for the designation of natural kind terms.

\[^99\] For more on the problems of non-barking dogs, see Carlson, 1977, 56nn.
We can read (2) as claiming that *being a tiger* is a feature of Tony’s haecceity. Tony’s natural kind haecceity, *being this sort of thing*, puts him in the extension of the kind ‘tiger’, which presumably has its own haecceity (it is hard to tell what that would be except that clearly the kind ‘tiger’ is not a tiger).\(^{100}\)

Intuitively, a Kripkean analysis of the predicative use of natural kind terms – as seen here – is not very satisfactory. The non-descriptionality of natural kind terms together with the metaphysical necessity attendant to the relation between an individual and the kind to which it belongs, makes it somewhat unclear why sentences like (2) are used to convey information. There does not seem to be a connection between the semantics Kripke predicts for this sort of sentence and the function it has in a real discourse.

Sentence (3) should be analysed as necessary because subsumption of one species under another, if true, is necessary.\(^ {101}\) Moreover, we should know a priori that if such subsumption is true, it is necessarily true. Intuitively, this is not a very satisfactory position, because it ascribes metaphysical necessity to taxonomic principles. Yet we know that taxonomy changes all the time. A Kripkean analysis of sentence (3) makes ambitious realist claims about taxonomy. These are not trivial assumptions to make, and one would expect Kripke to provide some sort of argument for this position but, unfortunately, this he does not do.

Any reconstruction of Kripke’s analysis of the semantics of natural kind terms is tentative because there is little textual basis to work with. It is not hard to see why the analogy between natural kind terms in their referential function and singular terms is tempting. However, once we analyse natural kind terms as non-descriptive rigid designators, we run into problems with their predicative use. In his treatment of natural kinds, Kripke does not deal with the predicative use of natural kind terms. Using an extrapolation from the Nixon case is safe but the resulting position is not satisfactory.

\(^{100}\)See Frege, 1893, re-print in English 1952.

\(^ {101}\)See a quotation to this effect on p. 108.
3.2. Kripke’s Approach to the Semantics of Modal Statements

It is surprising that Kripke’s views on natural kinds, which have been very influential, are so incomplete with respect to even rather obvious issues, such as those we mentioned above.

### 3.2.8 Necessary Statements and Their Commitments

When investigating the interface between semantics, metaphysics, and the role of science in Kripke’s work, we quickly find that the borders between the three shift depending on what kind of statements and entities we consider. In this section, we shall look at different kinds of a posteriori necessary statements, and try to find out what has to be presupposed if we are to analyse them in a Kripkean fashion.

Let us start by considering statements that express identity between individuals, like ‘Hesperus is Phosphorus’, ‘Cicero is Tully’, and ‘I am Anna Pilatova’. According to Kripke, these statements – that is, statements composed of two distinct singular terms, each of which is either a proper name or an indexical – are a posteriori and necessary.\(^\text{102}\)

That these statements should be a posteriori seems immediately plausible: that Hesperus and Phosphorus refer to the same planet was at one time a discovery; that Cicero is Tully, or Woody Allen is Alexander Konigsberg is still news to some. And while I may know that I am Anna Pilatova a priori, to others it is a posteriori, otherwise there would be no need for me to ever introduce myself. And if ever I suffered amnesia, I would be happy to find out what my name is.\(^\text{103}\)

The a posteriori status of these identity statements being settled, let us turn to the claim that they are necessary. Where does that necessity come from?

\(^{102}\)See his argument at Kripke, 1980, 101-105.

\(^{103}\)This issue borders on the problems connected with describing what it takes to know who one is. If all I forget is my name (which is then supplied to me), the situation is different from one of forgetting everything except my name (then knowing that I am Anna Pilatova is not very helpful).
Once we treat proper names (and indexicals) as obstinately rigid designators, we assume that they designate the same thing in all possible worlds regardless of the existence of the referent in any given world, with the possible exception of the actual one.\textsuperscript{104} Coreferential singular terms designate the same thing in every possible world, which is by definition equivalent to saying that if two or more singular terms are coreferential, they are necessarily so. And, consequently, the identity statement that connects the two will be necessarily true as well. But even if singular terms were only persistent designators, and for example ‘Hesperus’ and ‘Phosphorus’ were coreferential names referring to Venus, an identity statement like ‘Hesperus is Phosphorus’ would still be true in every world where Venus exists. The sentence ‘If Hesperus exists, it is identical with Phosphorus’ would be never be false. Therefore, even if singular terms were only persistent designators, we would still get necessary identity statements that make an assumption of existence explicit, and that can be taken as close enough to necessity tout court.\textsuperscript{105}

In order to arrive at this conclusion, we had to assume that singular terms are rigid designators, and in order to make that claim, we had to presuppose at least an essentialist version of haecceitism. That is, we had to assume that in every world where it exists, the referent of a rigidly designating expression has the property of being the very entity it is.

Kripke, however, seems to argue that the necessity of certain kinds of identity statements must be derived by ‘philosophical analysis’:

Certain statements - and the identity statement is a paradigm of such a statement on my view - if true at all, must be necessarily true. One does know \textit{a priori}, by philosophical analysis, that if such an identity statement is true it is necessarily

\textsuperscript{104}If an obstinately rigid designator has no referent in the actual world, the situation becomes more complex. It could turn out that the entity is for example fictional.

\textsuperscript{105}There are well-known epistemic problems that arise in connection with identity statements. Kripke pointed these out in his puzzle about Pierre (Kripke, 1979). We deal with this topic at length in section 4.2.
It is unclear whether the ‘philosophical analysis’ here mentioned is supposed to be a semantic analysis, in which case this argument would be compatible with the reasoning we outlined above, or whether the analysis Kripke intends here is supposed to amount to something more.

I suspect the above-mentioned ‘philosophical analysis’ must mean more than an analysis of language, because Kripke seems to use this reasoning to support his analysis of identity statements involving natural kinds and substances. Kripke claims that statements such as ‘Water is H$_2$O’, ‘Gold is the element with the atomic number 79’, or ‘Cats are mammals’ are necessary, and, moreover, if they are true then they are a priori so, and can a priori be known to be necessary.

The kind of essentialism inherent in Kripke’s claim that these statements expressing ‘theoretical identifications’ are necessary goes beyond the adoption of the ontology of natural kinds and their haecceities. The modal status of these statements suggests concrete essentialist principles concerning natural kinds, such as that subatomic composition is a necessary feature of an atomic element, and that being a subkind of a taxonomically higher kind is an essential feature of a species. This is essentialism with much more metaphysical bite than the one we needed to make the rigid designation of proper names and natural kind terms work.

In addition, it seems now plausible to interpret the quotation above as an argument saying that for some sorts of theoretical identifications, we know a priori that if they are true, then they are necessarily so. We may be wrong about the particulars, it may just conceivably turn out that an atom of gold has 78 protons, but the principle that an identification of an element with a particular atomic number if true must necessarily be true is supposed to be known a priori.

---

107See Kripke, 1980, 116-117.
108As he calls them e.g., in Kripke, 1980, 116.
In parallel to the above-mentioned necessary statements about natural kinds, Kripke lists various a posteriori necessary statements about individuals. Statements like ‘Richard Nixon is not an inanimate object’, ‘Elisabeth II sprang from the very gametes she actually sprang from’, and ‘This table is necessarily made of the very hunk of wood it is actually made of’ suggest various essentialist principles concerning individuals.

Nathan Salmon convincingly argues that the essentialist principles implied by these kinds of statements (e.g., ‘Gold is an element with the atomic number 79’, and ‘Nixon is not an inanimate object’) cannot be derived from the semantics of the terms alone, or possibly from their semantics and some uncontroversial premises. The necessity of these statements derives from a previously assumed metaphysical theory of essentialism that is independent of a theory of reference.

It seems thus that Kripke makes two kinds of claims regarding essences. Firstly, he claims that certain general essentialist principles are known a priori. An example of this kind of principles is that having the parents one actually has is an essential property of persons. Secondly, Kripke claims that some particular a posteriori statements, e.g., ‘Gold is the element with atomic number 79’, if true at all, are necessarily true. Neither of these positions can be derived from an analysis of language. Let us now have a closer look at the assumptions that allow Kripke to make these claims.

3.2.9 Kripke’s Scientific Realism

The kind of essentialism involved in statements such as ‘Gold is the element with the atomic number 79’ involves claims about how things are in the world, which makes it a claim about metaphysical necessity, as Kripke freely admits. In this section, we shall look at Kripke’s motivation for accepting this view, and say something

109In all these cases, the property mentioned is supposed to be essential (necessary) of that particular individual or thing.

110Salmon, 1982.

111We shall investigate this claim in more detail in the following section.
about the assumptions that need to be made in order to derive it.

As we already indicated, Kripke claims that it is known a priori of some characteristic theoretical identifications (such as ‘Water is H₂O’) that if they are true, then they are necessarily so. This presupposes that we know certain essentialist principles in their general form (e.g., ‘chemical composition is a necessary property of a chemical compound’) in some sort of a priori manner. However, it is very hard to see what kind of a prioricity Kripke has in mind here.¹¹² It is quite implausible to assume that we could derive these essentialist principles by reflection on the meaning of natural kind terms alone. We must therefore assume that there is yet another kind of considerations at play here. Just what they are becomes clearer once we look at claims of an even more radical kind, which Kripke also seems to endorse, namely that some essentialist claims – and not just principles – concerning particular natural kinds are metaphysically necessary.

The motivation behind these claims emerges in passages such as:

Such statements [as ‘Gold is an element with the atomic number 79’] representing scientific discoveries about what this stuff is, are not contingent truths but necessary truths in the strictest possible sense.¹¹³

What this passage illustrates, is Kripke’s belief that science can discover not only truths about the actual world – which would be contingent – but also truths about all possible worlds, that is, necessary or essential truths.

Characteristic theoretical identifications such as ‘Heat is the motion of molecules’ might be interpreted as definitions, hence as analytic, a priori, and without any essentialist import. Yet Kripke insists that this is not the reading he intends, stating that

¹¹²The lack of clarity in Kripke’s notion of a priori, which we investigated in section 3.2.6, makes things only worse.
¹¹³Kripke, 1980, 125.
¹¹⁴And passages to the same effect, such as “...whether science can discover empirically that certain properties are properties are necessary of cows, or of tigers, is another question, which I answer affirmatively.” (Kripke, 1980, 128)
Chapter 3. Modal Statements, Individuals, Essences

...characteristic theoretical identifications like ‘Heat is the motion of molecules’, are not contingent truths but necessary truths, and here of course I don’t mean just physically necessary, but necessary in the highest degree - whatever that means.\textsuperscript{115}

In Kripke’s view, science is quite clearly assumed to be the tool of discovering how things are in a metaphysical sense, that is, independent of our perception of them, our language or knowledge. The meaning of the terms that various sciences use, like ‘heat’, ‘light’ or ‘tiger’, is taken to be independent of changes in our knowledge. Their meaning is supposed to be constant because they represent kinds, unchangeable abstract entities, and while the extension of a natural kind may change, its essence does not – and science is seen as the right tool for uncovering that essence.

Note that on the present view, scientific discoveries of species essence do not constitute a ‘change of meaning’; the possibility of such discoveries was part of the original enterprise.\textsuperscript{116}

These are some very strong assumptions, which should be backed by some powerful arguments. Kripke, however, does not seem to argue for his position at all. These basic presuppositions are just taken for granted. Most, if not all, of the a posteriori necessary claims Kripke makes are supported by Kripke’s view’s on science, and cannot not be upheld without making metaphysical assumptions. Kripke’s views on science are not, and cannot be motivated by the semantics of natural language – metaphysics is inherently independent of language. Our interest is in the semantic of natural language, and that is why we shall try to limit to a minimum the amount of assumptions that cannot be justified by an analysis of language. In this enterprise, metaphysics is something we should try to do without.\textsuperscript{117}

\textsuperscript{115}Kripke, 1980, 99.
\textsuperscript{116}Kripke, 1980, 138.
\textsuperscript{117}I think one should be cautious with metaphysics in general. It may sometimes be useful to introduce into one’s theory posits that are not further explained but they should be treated as such.
3.2.10 Kripke Without Metaphysical Assumptions

The question remains what Kripke’s position would look like without scientific realism, that is, without the view that science discovers metaphysically necessary truths. If he were not a scientific realist, could he still claim that proper names and natural kind terms are rigid designators?

It turns out that he could very well do without scientific realism. The essentialism necessitated by rigid designation of proper names or natural kind terms only requires the essential properties of the form being the very entity it is for individuals, and being this sort of thing for natural kind terms, that is, haecceities. In the case of natural kind terms, we adopted into our ontology natural kinds as abstract entities. If we are willing to make these assumptions, we can have rigid designation for both proper names and natural kind terms, a position that is essentialist in terms of further unanalysed haecceities. Furthermore, haecceities can be treated as posits simply introduced to make rigid designation work. And rigid designation is a claim about the semantics of singular and natural kinds terms. While it is open to refutation by counterexamples from spoken language, we have not yet encountered an example arguing directly against it.

We should note that the definition of rigid designation does not imply that once it is adopted, it has to be assumed for all the kinds of terms Kripke ascribes it to. We could hold that while proper names are rigid designators, natural kind terms are not, and analyse characteristic theoretical identifications not as referring to natural kinds but rather treat them as definitions. Depending on evidence from language, we can decide for what kinds of entities we want to assume haecceities. The rigid designation thesis implies some ontological commitments - adoption of haecceities - and we can decide for each kind of individuals whether to take that step or not.

\(^{118}\)The parallel between the haecceitism assumed in rigid designation of proper names and that of natural kind terms is not explored by Salmon, 1982 but it is, I believe, compatible with his views.
The story is different for the more particular essentialist principles because they go beyond the assumption of haecceitism. Kripke’s motivation of these principles is twofold: in their general form (e.g., ‘being made of a particular chunk of material as an essential property of an artefact’), these principles are supposed to be accessible to us a priori, by philosophical reflection.\footnote{See the above-quoted passage of Kripke, 1980, 109.} The problem with this sort motivation is that we can easily imagine a situation where someone arrives by her philosophical reflection at different conclusions. It is hard to imagine how one could settle differences of opinion in the realm of the a priori. Motivation from a priori reflection thus should be treated as idiosyncratic. Among other things, it is not supported by generally shared intuition about the meaning of the relevant words in language, and, as we pointed out in section 1.3, while we have to rely on our semantic intuitions to some degree, where this clearly gives rise to disagreement, the claim has to be supported by an independent argument.

So, Kripke’s claim that particular essentialist statements (e.g., ‘Gold is an element with the atomic number 79’) are metaphysically necessary should be seen as motivated by strictly and exclusively by scientific realism. As we already pointed out, Kripke does not argue for scientific realism, he simply assumes a particular form of it. We find it objectionable to employ controversial metaphysical assumptions in the context of a study of the semantics of natural language. We may be unable to do semantics of modal statements without employing our pre-philosophical modal intuitions, in case those modal intuitions vary among speakers and, on top of that, are not supported by generally shared semantic intuitions, one should seek some other kind of argument. Kripke’s reliance on scientific realism (without a convincing defence of that position) in the course of a semantic analysis is an example of an idiosyncratic use of modal intuitions, and as such it should be avoided.

When mentioning modal intuitions earlier (on p. 15), we said that we should not presuppose any connection between the semantic and the modal intuitions. Now we see that some connection should
be assumed. Where modal intuitions are not supported by semantic ones, or contradict them, they become problematic.

But what would happen with Kripke’s analysis of modality without scientific realism? As we have shown, what is possible or necessary is given by the modal properties of actual-world entities. If we do not know how to find out what the necessary properties of actual-world entities are, we are equally at loss when trying to figure out whether any possible world is indeed possible. Remember that whether something is or is not possible is independent of our ability to imagine such a state of affairs. Possible worlds that have to comply with metaphysical necessities – and Kripkean possible worlds are of this kind – are a liability if we cannot find out what those necessities are. If we are to give up metaphysics, we have to have a story about the kind of possible worlds we do want to work with, and about the necessity that holds such a framework together.

3.2.11 Conclusion

Dealing with Kripke’s conception of the semantics of modal statements, we started off by reconstructing his version of possible worlds, focusing on the claim that it is admissible to stipulate directly which individuals are involved in a possible world. According to Kripke, one may stipulate that an actual-world individual is involved in a non-actual possible world. This is equivalent to the adoption of transworld individuals. These, in turn, are needed if we want to make sense of the notion of rigid designation - we saw in our analysis of Lewis’s work that in a framework where individuals are world-bound the concept of rigid designation made no sense.

The main problem with transworld individuals is that one has to have a way of telling what counts as the same individual across various possible worlds, that is, under what conditions an individual retains its identity under change. In other words, one needs to be

\footnote{Our approach to the question of necessary properties may, however, differ depending on the kind of entity we consider. We may well take a piecemeal approach to ontology. We could believe that mathematical entities have necessary properties while rejecting non-trivial essentialism concerning persons.}
able to tell when something that may look quite unlike a particular actual-world individual is still our actual-world individual, and vice versa, when something that looks just like an actual-world individual is something else. This job cannot be done by anything less than sufficient and necessary identity criteria for individuals, and haecceities are the minimal presupposition one has to make. Particular essential properties (like essentiality of origin) can not do the job because they provide the necessary, but not the sufficient identity criteria. This is why the haecceities were brought in. A haecceity, as we saw, is a property that an individual has to have in every possible world in which it exists. It is the property of being the very entity it is. We can either treat it as a primitive (which, as Kripke says, is usually the practical thing to do) or analyse it further. The possibility of further analysing haecceities depends on there being bridging laws that would facilitate a reduction of identity criteria for one kind of entities to the identity criteria of other constitutive entities. The viability of ontological reduction of essential properties for particular kinds of entities is, Kripke says, to be decided by scientific inquiry.

The semantics Kripke proposes for natural kind terms is constructed to a degree in parallel with the semantics of proper names. The range of problems presented by the reference of natural kind terms, however, is significantly different from the problems of reference of singular terms, and in the present work we touched upon them only in a cursory way.

The main problem with Kripke’s possible world framework as we found it had to do with the role of metaphysical necessity, which was brought in to determine which worlds are possible and which only seem to be so. Metaphysical necessity is inherently epistemically opaque, which means that we – both as speakers or a linguistic community in general – are not in the position to know which worlds are possible and which are not, and this is built into the very core of Kripke’s possible world framework.

\[\text{footnote}{For an exposition of some of the problems see Carlson, 1977, Salmon, 1982, or Lowe, 1997.}\]
Moreover, Kripke investigates metaphysical necessity while neglecting other kinds of necessity. In natural language, modal terms are used in a variety of ways, and a possible world framework should have enough flexibility to capture them. In natural language, we seem to use successfully various kinds of modal statements, and we are, at least sometimes, fairly certain of their truth or falsity (which is hardly ever the case with metaphysical necessity). What we want is a framework that would enable us to better model how modal statements function in a natural discourse. And describing a possible world framework that takes its inspiration from real discourse shall be the task of the rest of this chapter.

3.3 Stalnaker’s Worlds

In our analysis of Lewis’s approach to possible worlds, we concluded that the realism that characterises it necessitates the adoption of worldbound individuals, which in turn leads to the counterpart theory. Both of these features detract from the theory’s potential usefulness in an analysis of natural language. Kripke’s approach does not suffer from these particular drawbacks but the elements of metaphysics, which could not be justified from the semantics of natural language, make his framework less than an ideal candidate for a theory we would like to work with. Stalnaker’s framework seems a more attractive option – while it works with transworld individuals, metaphysical assumptions seem to be absent from it. That is why in the following sections, we shall investigate Stalnaker’s approach to possible worlds in detail: we shall look at his motivation for using possible worlds, the requirements he sets for a possible world framework, and the way he goes about building one. An analysis of the problems he encounters, and of the way he deals with them, should help us to gather useful hints for our own analysis of modality.

Stalnaker has been influenced by the work of David Lewis. We analysed some parts of Lewis’s work at the beginning of this chapter.

\footnote{See for example Stalnaker’s contributions in Harper, Stalnaker, and Pearce, 1981.}
ter, and concluded that its use for an analysis of natural language is problematic. That is why here we shall focus on those parts of Stalnaker’s work where he departs, directly or indirectly, from Lewis’s line of thinking.

### 3.3.1 Stalnaker’s Motivation

In order to understand a possible world framework, one should first look at the problems it is supposed to help solve. In the present case, this is not too difficult. Although Stalnaker’s work covers a large number of topics, its main orientation remained constant: his primary interest has been in understanding representational mental states and their role in explaining human behaviour.

Representational mental states should be understood primarily in terms of the role they play in the characterization and explanation of action. What is essential to rational action is that the agent be confronted with a range of alternative possible outcomes of some alternative possible actions. The agent has attitudes, pro and con, toward the different possible outcomes, and beliefs about the contribution which the alternative actions would make to determining the outcome.\(^{123}\)

In this picture, people are seen primarily as agents, and language is treated as just one of the means by which an agent’s beliefs can be manipulated in a systematic fashion. Stalnaker’s main goal is to explain the connections between representational mental states and actions. He is interested in the attitudes, e.g., beliefs, which agents entertain with respect to alternative outcomes of their actions, i.e., with respect to possible states of the world. Which possible states of the world are relevant to an action, both as starting conditions and as results, is determined by the context. Propositions are then introduced as a way of distinguishing between relevant alternatives, and agents’ attitudes are treated in terms of their attitudes to propositions.\(^{124}\)

\(^{123}\)Stalnaker, 1984, 4.  
\(^{124}\)This characterisation is in line with Stalnaker’s own description of his aims as he presents it for example in Guttenplan, 1995, 561-568.
Stalnaker aims at balancing, in his account, two basic aspects of an explanation of behaviour. On the one hand, one can explain behaviour in terms of agents’ attitudes and their perceptions of what the different possible outcomes of their actions are. On the other hand, one must take into account that if agents’ actions are to produce, in general, the desired effect, they must be guided by a more or less accurate model of reality. To put it differently, we have on the one hand the perspective of the agent, and on the other hand the world she encounters. In order to describe the beliefs and desires that guide and motivate the agent in her actions, we have to take into account her epistemic situation. On the other hand, to provide a full description of an action, and of the success the agent aims at, we have to bring into the picture the way the world is.\textsuperscript{125} For example, if I want to buy some lettuce for a salad, I succeed if I buy not just something I believe to be lettuce, but only if that thing is, in fact, lettuce. In a description of intentional behaviour, it is crucial to make the right sort of connection between the epistemic possible worlds of the agent and the possible worlds that are independent of an agent’s wishes, beliefs or desires. Developing a framework that successfully balances out these two perspectives is, I think, what Stalnaker takes to be the core task of most of his work.

Our main aim here is to put flesh on the idea of this balance. It may seem that the two desiderata of explanation of behaviour sketched above – i.e., basically the internalist and the externalist perspective – lead to two distinct notions of content. On the one hand, in order to explain why an agent acts the way she does, we can ask about her perception of the situation she finds herself in. In characterising it, we feel compelled to employ the narrow, internalist conception of content. On the other hand, in order explain why the agent can act with a degree of certainty based on her previous experience regarding the outcome of her actions, we need a notion of content that takes into account the contribution of the environment, that is, the broad, externalist content. These two notions of content rest on different notions of possible worlds. We shall try to see how

\textsuperscript{125}See Stalnaker, 1984, 18-20.
Stalnaker balances out the internalist and the externalist perspective and the notions of possible worlds implied by them. To do this, we shall first look more closely at the main concepts Stalnaker uses, and explore their connections.

### 3.3.2 Possible Worlds

Unlike Kripke, and like Lewis, Stalnaker is explicit about the kind of possible worlds he wants to work with. It is possible, however, that his notion of possible worlds, even if seemingly explicit, hides tensions resulting from trying to account for both narrow and broad content. The conception of possible worlds that a theory uses, has extensive consequences for the notions of proposition and content, and that is why we shall focus on detecting possible tension between narrow and broad content throughout our inquiry.

Stalnaker deals explicitly with the concept of possible worlds in his oft-quoted article ‘Possible Worlds’, which later became a part of his *Inquiry*. He outlines his own conception of possible world by contrasting the differences between it and Lewis’s notion of possible worlds. We shall look now at how Stalnaker understands Lewis’s framework, at the points of difference between the two authors, and at the final picture arising from the comparison.

Stalnaker characterises Lewis’s theory of possible worlds as adherence to the four following theses:

1. Possible worlds exist.
2. Other possible worlds are things of the same kind as the actual world - ‘I and my surroundings.’

---

126 Stalnaker, 1979, 225-235.
127 Stalnaker, 1984, Chapter3.
128 I have already used Stalnaker’s characterisation of Lewis’s theory as Stalnaker presents it in Stalnaker, 1979, 227 in the sections of this chapter pertaining to Lewis. It is, however, in the interest of easier reading that they should be repeated here.
129 This quotation, as well as other quotations and direct paraphrases of Lewis’s work in this section come from and are based on Lewis, 1979a. The quotations in italics come from Stalnaker, 1979.
is only one world among others. We call it actual not because it differs in kind from all the rest but because it is the world we inhabit.

(3) The indexical analysis of ‘actual’ is the correct analysis. “‘Actual’ is an indexical. It depends for its reference on the circumstances of utterance, to wit the world where the utterance is located.”

(4) Possible worlds cannot be reduced to anything more basic. Lewis says, “When I profess realism about possible worlds, I mean to be taken literally. Possible worlds are what they are and not some other thing.”

The first thesis, Stalnaker says, is compatible with Lewis’s claim that we believe in possible worlds in virtue of believing that things might have been different from the way they are.\textsuperscript{130} Lewis claims that ‘ways things might have been’ exist but says so far nothing about the nature of those ‘ways’. It is the second thesis that contains the ontological commitment. Here Lewis says that possible worlds are concrete particulars. The actual world is ‘I and my surroundings’, and other possible worlds are more things like that. However, the motivation Lewis gives for thesis (1) – the claim that possible worlds are ‘ways things might have been’ – does not support thesis (2), which states that worlds are concrete particulars. If possible worlds are ‘ways things might have been’, then the actual world should be ‘the way things are’ rather than ‘I and my surroundings’. ‘The way things are’ is a property of a world, not a world itself. This is an important distinction because a property can exist unsubstantiated\textsuperscript{131}, and the way the world is could therefore exist even if the world that would be that way did not. Can we conclude that (2) is based on an equivocation between ‘the actual world’ in the sense given by ‘I and my surroundings’, and the sense given by ‘the way things are’? Yes, but (2) also has a deeper motivation that relates to thesis (3).

\textsuperscript{130}The reasoning in this paragraph closely follows Stalnaker’s reasoning in Stalnaker, 1979, 227-228.

\textsuperscript{131}This is true of most properties, and all the properties we are interested in at the moment. It does not hold for properties like being this person.
Chapter 3. Modal Statements, Individuals, Essences

Thesis (3) tells us that \textit{being actual} is a world-relative attribute. According to Lewis, the actual world is special because it alone is the concrete world. But this is a contingent fact: from the viewpoint of a counterfactual possible world, that world would be actual.\textsuperscript{132} At this point, Stalnaker starts disagreeing with Lewis. He asks

\ldots if there is no absolute property of actuality, does this not mean that, looking at things from an objective, absolute viewpoint, merely possible people and their surroundings are just as real as we and ours? Only if one identifies the objective or absolute viewpoint with a neutral standpoint outside of all possible worlds. But there is no such standpoint. The objective, absolute point of view is the view from the actual world, and it is part of our concept of reality that this should be so.\textsuperscript{133}

From our viewpoint, all other viewpoints are just possible.

We could, however, separate the semantic analysis of ‘actual’ from the metaphysical thesis that actuality is a relation between a world and things in it. According to Stalnaker, one can accept one thesis and reject the other, just as one can accept an indexical analysis of personal pronouns and be a solipsist, or accept an indexical analysis of tenses, and yet believe that the past exists only in memory and the future only as anticipation.\textsuperscript{134}

What is the picture of possible worlds we are left with? Stalnaker rejects Lewis’s full-blown realism concerning possible worlds in favour of moderate realism. He rejects the metaphysical interpretation of the indexical analysis of actuality but accepts the indexical analysis as semantically correct. To him, possible worlds do not exist in the same way in which the actual world does - rather, he seems to treat them as a useful instrument.

The concept of possible worlds that I am defending is not a metaphysical conception, although one application of the

\textsuperscript{132}See Stalnaker, 1979, 228-229.

\textsuperscript{133}At Stalnaker, 1984, 47. Even though the article ‘Possible Worlds’, Stalnaker, 1979, is basically the same as Chapter 3 of \textit{Inquiry}, Stalnaker, 1984, some formulations are more precise in the later version.

\textsuperscript{134}See Stalnaker, 1979, 229.
3.3.  *Stalnaker’s Worlds*

notion is to provide a framework for metaphysical theorizing. The concept is a formal or functional notion, like the notion of an individual presupposed by the semantics for extensional quantification theory.\(^{135}\)

Stalnaker thus defends a formal notion of possible world. Just like the concept of an individual, he is happy to keep it as a primitive. The main attraction of Lewis’s framework, as Stalnaker sees it, was its flexibility, and that flexibility he tries to preserve. By rejecting Lewis’s extreme realism, Stalnaker avoids some problems inherent in that position, e.g., the theory of counterparts and the question of the domain of all possible worlds. Different kinds of possibility and necessity as well as different contexts shall give rise to various domains of different kinds of possible worlds, but there does not need to be any definitive or maximal domain from which all kinds of possible worlds are drawn.\(^{136}\)

In his statement about the kind of possible worlds he wants to work with, Stalnaker manages to avoid at least some of the problems that plague both Lewis and Kripke. Indeed, if we accept both a ‘possible world’ and an ‘individual’ as primitive notions, it may seem that little more needs to be said on the subject. Stalnaker’s framework seems remarkably free of metaphysical and ontological assumptions. The difference between Lewis’s kind of ontological assumptions and those of Stalnaker is that Lewis asserts that existence of possible worlds without qualification. That makes his conception of possible worlds a metaphysical one. Stalnaker, on the other hand, acknowledges the usefulness of a possible-world approach to modality and avails himself of its principle advantages but avoids making an ontological commitment to possible worlds. By asserting that their existence is presupposed just for the sake of explanation, he puts it, so to say, in scare-quotes. This approach is not novel or limited to semantics. Hume\(^{137}\) attacks the notion of causality but

\(^{135}\)Stalnaker, 1984, 57.

\(^{136}\)I defend this interpretation later. See also Stalnaker, 1981, 135, quoted in the next section.

\(^{137}\)Hume, 1748.
admits its indispensability in everyday reasoning, and Adams proposes to treat modality by resorting to ‘world-stories’ without believing that world-stories really exist.\textsuperscript{138} Regarding possible worlds, the variety of approaches to their existence is considerable, and a black and white classification (as in ‘this author believes in possible world while that one does not’) is grossly insufficient. We are not going to investigate the issues of approaches to existence of possibles in detail,\textsuperscript{139} let us say only that the difference between Lewis’s extreme realism about possible worlds and Stalnaker’s approach is considerable, and that by treating possible worlds as tools within a theory of modality, Stalnaker does not seem to make a metaphysical commitment to their existence.

We shall now investigate some notions that are in Stalnaker’s work connected with the concept of a possible world, and look for problems that could be tucked away there.

### 3.3.3 Proposition

At the outset, we noted that Stalnaker’s main focus is on the explanation of intentional behaviour. It is important to add that for Stalnaker, a ‘respectable’ explanation is one given solely in non-intentional terms.\textsuperscript{140} This is something we should keep in mind while going through the following sections. Right now, we shall turn our attention to analysing in greater detail the kind of possible worlds Stalnaker works with and the use he puts them to. We have already noted that, by rejecting Lewis’s style of realism, Stalnaker escapes the problem of having to specify the domain of all possible worlds. In his framework, we have as many possible worlds as we need, no more, no less. Just how many worlds are needed, and of what kind, is determined by the context. We find an attractive statement of

\textsuperscript{138}Adams, 1979.

\textsuperscript{139}For a very nice overview of this discussion see Loux’ extensive introduction to Loux (1979).

\textsuperscript{140}See for example Stalnaker, 1984, 27: “I argued…that it is theoretically possible to solve the problem of intentionality - to give a naturalistic explanation of intentional mental states - without exploiting linguistic or semantic concepts.”
this approach in his ‘Indexical Belief’, where Stalnaker says

I doubt that it is plausible to believe that there is, independent of context, a well-defined domain of absolutely maximally specific possible states of the world... The alternative possibilities used to define propositions must be exclusive alternatives which are maximally specific, relative to the distinctions that might be made in the context at hand. But one can make sense of this requirement even if there is no ultimate set of possibilities relative to which any possible distinctions might be made. One might think of possible worlds as something like the elements of a partition of a space, rather than as the points of the space. The space might be partitioned differently in different contexts, and there might be no maximally fine partition. (This is only a rough analogy. And the space itself may also vary from context to context.)

To better understand the relation between possible worlds and context, we first have to introduce the notion of proposition. Stalnaker uses the notion of proposition to account for the content of representational mental states, e.g., beliefs, and also to express the impact of an utterance on the context and to capture the way an agent aims at changing something in the environment. He uses a basically Kripkean notion of proposition, according to which a proposition is a function from possible worlds into truth-values. This notion should be seen as set against the Russelian tradition, which holds that propositions are structured and reflect the structure of sentences that express them. Kripkean propositions are not structured in this way. Different sentences can express the same proposition, a particular sentence in different contexts can express different propositions, and, actually, a proposition need not even be expressed by linguistic means. Given a set of possible worlds, a proposition is determined by the subset of possible worlds in which

\[141\] Stalnaker, 1981, 135.

\[142\] See for example Stalnaker, 1998, 3: “...speech is action, and speech acts should be understood in terms of the way they are intended to affect the situation in which they are performed.”

\[143\] Stalnaker, 1987, 2.
it holds. And this is where Stalnaker, but not Kripke, connects the notion of proposition with the notion of context - the context determines which possible worlds are relevant to determining the proposition. He says that

If the alternative possibilities there are vary with the context, then so do the propositions which are, according to the conception of content I am sketching, just ways of distinguishing between the alternative possibilities. One can make sense of questions about identity and difference of the propositions expressed in different utterances or acts of thought only given a common context - a common set of possibilities that the propositions are understood to distinguish between. This yields a conception of proposition, which is less stable than, and very different from, the traditional conception, but it is, I think, more adequate to the phenomena of speech and thought.\footnote{Stalnaker, 1981, 135.}

By making the relevant set of possible worlds dependent on the context, Stalnaker’s propositions start behaving very differently from Kripke’s. For example, in Stalnaker’s framework, the notion of rigid designation becomes much weaker. Where in Kripke’s framework a rigid designator denotes the same individual in all (metaphysically possible) possible worlds, in Stalnaker’s framework a rigid designator designates the same individual in all the worlds of the context.

To sum up, according to Stalnaker, the proposition expressed by an utterance is a function over the set of possible worlds that are relevant to the context of an utterance – it divides those possible worlds into those where the proposition holds and those where it does not.

3.3.4 Context

We have now some idea of the role context plays in determining a proposition. It is time to turn our attention to the notion of context itself. Stalnaker offers a preliminary definition, saying:
I propose to identify a context (at a particular point in a discourse) with the body of information that is presumed, at that point, to be common to the participants in the discourse.\footnote{Stalnaker, 1998, 5.} 

In general, there are two kinds of things that participants in a conversation assume to be shared. Firstly, they normally take it for granted that both speakers and hearers are aware that a conversation is taking place, and that all of them are aware of their surroundings.\footnote{Stalnaker, 1998, 5 and Stalnaker, 1978, 323.} Secondly, participants assume they share knowledge of some less immediate features that ‘frame’ the situation. For example, if they talk about American politics, they may assume that all of them know who is the president. Both kinds of assumptions contribute to the information which participants assume is shared. These are the presuppositions of the participants.

We can represent the information that defines the context in which a speech act is taking place with a set of possible situations or possible worlds - the situations that are compatible with the information. This set, which I have called the context set will include all the situations among which the speakers intend to distinguish with their speech acts. The presumed common information - what is presupposed in the context - is what all these worlds have in common.\footnote{Stalnaker, 1998, 5.}

A speech act of assertion takes place in all the possible worlds of the context set. In effect, an assertion is a proposal to add information to what is presupposed, that is, it is a move to eliminate from the context set those worlds that are incompatible with what is asserted.

Should we see presuppositions as objective features of the agents’ interaction within a context or should we rather see them as necessarily tied to the perspective of one participant, usually the speaker? There is a degree of vacillation or perhaps just a lack of clarity in Stalnaker’s work regarding this point but the latter seems to result
in a more coherent reading. It is consistent with the repeated emphasis on information that is presumed to be shared (the presuming being done, one would suppose, by one agent at the time), as well as with the focus on the speaker, which we can witness in numerous quotations in this section. Even the passage where – as far as I could tell – Stalnaker defines the notion of presupposition for the first time supports the speaker-centred reading:

A proposition is presupposed if the speaker is disposed to act as if he assumes or believes that the proposition is true, and as if he assumes or believes that his audience assumes or believes that it is true as well. Presuppositions are what is taken by the speaker to be the COMMON GROUND of the participants in the conversation, what is treated as the COMMON KNOWLEDGE or MUTUAL KNOWLEDGE.\(^{148}\)

The structure of a speaker’s presupposition can be represented by a Kripke model, in which the accessibility relation is serial, transitive, Euclidean, but not necessarily reflexive.\(^{149}\) The requirement that it be transitive and Euclidian reflects the assumption that speaker presuppositions are transparent: speakers know what they are presupposing, so they that they are presupposing know they are presupposing \(P\) if they are, and that not if they do not. The requirement that the relation be serial reflects the assumption that the context-set is always non-empty, that there is always at least one possibility compatible with what is presupposed. The non-reflexivity is important here. It reflects the fact that some things the speaker presupposes may be false. This would happen either because an agent has a false belief or because she participates in some mutually recognised pretence. Moreover, it is not always, or perhaps not even usually, the case that participants in a conversation presuppose exactly the same things. An agent need not even believe that everything she presupposes is shared – it suffices if she pretends she does.\(^{150}\)

\(^{148}\)Stalnaker, 1978, 321, emphasis in the original.

\(^{149}\)Stalnaker, 1998, 6.

\(^{150}\)See the quotation immediately above as well as Stalnaker, 1978, 321, which is a direct continuation of the quotation above: “The propositions presupposed
Let us see now how Stalnaker puts the notions of context and proposition to work. His aim is to capture the effect the speaker’s utterance has on the context, that is, on the beliefs of the audience. This he models with the help of a two-dimensional matrix called a propositional concept. We can best illustrate how it works using Stalnaker’s own example:

I said You are a fool to O’Leary. O’Leary IS a fool, so what I said was true, although O’Leary does not think so. Now Daniels, who is no fool and who knows it, was standing nearby, and he thought I was talking to him. So both O’Leary and Daniels thought I said something false: O’Leary understood what I said, but disagrees with me about the facts; Daniels, on the other hand, agrees with me about the fact (he knows that O’Leary is a fool), but misunderstood what I said. Just to fill out the example, let me add that O’Leary believes falsely that Daniels is a fool. Now compare the possible worlds $i$, $j$, and $k$. Here, $i$ is the world as it is, the world we are in; $j$ is the world that O’Leary thinks we are in; and $k$ is the world Daniels thinks we are in.\footnote{Stalnaker, 1978, 317-318.}

The following propositional concept corresponds to this situation:

<table>
<thead>
<tr>
<th></th>
<th>$i$</th>
<th>$j$</th>
<th>$k$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$i$</td>
<td>T</td>
<td>F</td>
<td>T</td>
</tr>
<tr>
<td>$j$</td>
<td>T</td>
<td>F</td>
<td>T</td>
</tr>
<tr>
<td>$k$</td>
<td>F</td>
<td>T</td>
<td>F</td>
</tr>
</tbody>
</table>

This is a two-dimensional matrix where on the rows we find either the proposition the speaker expressed or the proposition as a hearer understood it (this, of course depends on who is the speaker). For example, in the second row, $j$, we see the proposition as O’Leary understood it, evaluated in the world as he thinks it is, $j$, and in the worlds as the speaker and Daniels think they are ($i$ and $k$). We construct a proposition as a function over the worlds that participants in the intended sense need not really be common or mutual knowledge; the speaker need not even believe them.”
think are relevant with respect to the situation at hand (this is of course as very simplified model). If we find the same proposition in two rows, we see that those two participants have understood an assertion as expressing the same proposition. In the columns, we find the worlds with respect to which we evaluate, that is, worlds in their role of a context (i.e., context worlds). Notice that in their role as context worlds, the world of the speaker and that of Daniels are the same (the verticals under $i$ and $k$). That means that the speaker and Daniels assign the same truth-values to the proposition expressed as they each understand it.

On the diagonal, that is in squares $ii$, $jj$ etc., we see whether the relevant participant himself thinks that the proposition as he or she understands it is true. In this particular propositional concept, we see in the square $jj$, that O’Leary, though disagreeing with the speaker about the facts, understands the utterance in the way the speaker intended. Daniels, on the other hand, does not understand the speaker’s utterance the way it was intended. We can view the diagonal as expressing a proposition, so-called ‘diagonal proposition’, which has some interesting properties. A diagonal proposition is relevant to describing situation in which the speaker or the addressee has only a partial knowledge of facts that are relevant to determining what is said. For example,\footnote{This example is adapted from Stalnaker, 1999b, 13.} if I get an undated postcard from San Francisco from my sister, saying, among other things, ‘it is warm and sunny here today’, I will not know exactly what it says (because I do not know when it was written), but I will know that it was written on a sunny day in San Francisco. The information I got from the postcard was the diagonal proposition of the propositional concept expressed by the writer, that is, my sister.

All this is well known. Yet, several more points need to be settled. Firstly, what determines which worlds occur in a propositional concept of a particular situation? This, in my view, is the place where the notion of context connects with Gricean maxims.\footnote{Which are introduced in the famous Grice (1975).} However simplified the propositional concept is, which possible worlds are
relevant to modelling a particular situation is determined by a set of Gricean principles, mainly the principle of relevance. In building a propositional concept, we implicitly rely on the Gricean maxims, which we assume the participants have internalised.\textsuperscript{154}

Secondly, we may find it surprising that in the quotation above, the world of the speaker is equated with ‘the world as it is, the world we are in’. This seems to contradict the explicit set up of the notion of context, which is said to be the

\[\ldots\text{set of possible situations that are compatible with what is presupposed, or taken to be common ground, by the participants in the discourse.}\]

Here, context is seen as consisting of information that is presumed to be shared by conversation participants. We then model that information using propositions, which are a function over the possible worlds which the participants think are relevant. The content of an assertion depends on the context in two ways: firstly, the context is the object on which an assertion acts, and secondly, it provides the source of information needed to interpret an utterance.

So both of the roles that contexts play require that they include a body of information: context-dependence means dependence on certain facts, but the facts must be available, or presumed to be available, to the participants in the conversation.\textsuperscript{156}

Context consists of possible worlds that are thought by the participants to be compatible with the information they share. Stalnaker gives us no reason to suppose that one particular participant’s context world is the actual one. In fact, the non-reflexivity of the model (which we mentioned on p. 132) rules that option out. Rather, if we look at the way Stalnaker describes context, it would seem that each participant thinks that his or her context set represents (a part of)

\textsuperscript{154}We shall return to the importance of Grice’s work to Stalnaker in section 4.4.5.
\textsuperscript{155}Stalnaker, 1998, 7.
\textsuperscript{156}Stalnaker, 1998, 5.
the actual world. This is not surprising once we go over the quotes above and asks ourselves whether the worlds Stalnaker works with are metaphysical or epistemic. It turns out that worlds consisting of information presumed by the speaker to be shared are quite clearly epistemic in nature. Nothing outside of epistemic alternatives can get into them. It follows then that the content they describe is a narrow, epistemic kind of content. But that does not seem to be the kind of content Stalnaker had in mind when setting out to explain intentional behaviour in non-intentional terms.

In order to get a fuller picture, we have to say more about content, focusing on what Stalnaker wants to say about the non-epistemic, broad component of it. And that is the topic of the following section.

### 3.3.5 The Possible and The Actual

We saw that in analysing particular conversational exchanges, Stalnaker uses epistemic possible worlds. Yet for any belief to be about something other than just itself, it has to depend in some way on the external world. Stalnaker addresses this problem by developing what he calls the ‘information theoretic account of intentional content’.\footnote{In the following, I discuss Stalnaker’s views as he presents them in Stalnaker, 1993. The problem he deals with is by no means unique to his theory. It is a version of the same problem that Block, 1986 is trying to address by making narrow versus broad content distinction, and the same is treated in the discussion about externalism versus internalism about belief.} Using a different idiom, we can say that after developing a story about narrow content, Stalnaker gives us a theory of broad content. To clarify his position, Stalnaker re-interprets the best-known argument for broad content there is: Putnam’s Twin Earth thought experiment.\footnote{It comes from Putnam, 1975b.} The Twin Earth thought experiment - and a good many other thought-experiments inspired by it – was designed to highlight the relation between the meaning of speaker’s utterances and the speaker’s environment.

Even though it is well known, let us briefly describe Putnam’s
example. Let us imagine a Twin Earth. The only difference between Earth and Twin Earth is that water on Earth is H\textsubscript{2}O, while on the Twin Earth there is something that looks and behaves exactly like the Earth water but has a different chemical composition, which we shall call XYZ. On Twin Earth, XYZ plays exactly the same part H\textsubscript{2}O plays on Earth. Let us also say that our scenario is happening before people knew about the chemical composition of water. Now, an actual, Earth speaker, O'Leary, has a twin on Twin Earth, a Twin O'Leary, who has all the same mental and physical properties as O'Leary does. Twin O'Leary should then have the same mental properties as O'Leary even when interacting with Twin water because there is nothing in his beliefs about that substance that could distinguish them from O'Leary's beliefs about Earth water. And yet, Putnam says, when O'Leary and his twin say something about water, they mean two different things. The meaning of their utterances differs depending on their environment. When O'Leary says 'There is water in the bathtub', he says something true on Earth, but would say something false on Twin Earth because there is no water there. Putnam argues that regardless of what Twin O'Leary thinks he means when using the word 'water', he cannot mean by it the same thing as O'Leary, who is an Earthling, because he, that is, Twin O'Leary, is not in position to refer to 'water'. ‘Water’ is a natural kind term, which in English inherently refers to H\textsubscript{2}O. Twin O'Leary lacks the requisite causal connection with H\textsubscript{2}O. That is why regardless of what Twin O'Leary thinks he means he cannot mean ‘water’ in the same sense in which O'Leary uses the term.

Putnam’s point is about the source of semantic values. Sometimes\textsuperscript{159} the Twin Earth experiment was understood as showing that what one means is not a matter of mental states but of social conventions and causal connections. According to this interpretation, words can do their semantic work without us knowing what goes on. Thus while speakers’ intentions and beliefs are inherently internal, their words depend for their meaning on external factors. This may conflict with our intuition that when a speaker is sincere

\textsuperscript{159}See e.g., Fodor (1987).
and means what he says, then the speaker says what he believes.

Tyler Burge pointed out\textsuperscript{160} that this is not the correct lesson to draw – according to him, Putnam shows that even belief and intention should be understood partly in terms external to the speaker. O’Leary’s twin then not only says something different from O’Leary when he says ‘There is water in the bathtub’ but also acts on a different belief and a different intention. Burge’s is a strongly externalist interpretation.

A theorist of an internalist inclination, on the other hand, tries to defend a position on which although what one sees and knows is partly a matter of external environment, what one thinks one sees and thinks one knows is a matter of the agent’s internal state. In order to account for the role of the external environment, the internalist distinguishes between the way things seem and the way they are. The externalist insists that this is a remnant of Cartesian dualism. This is a serious objection because the internalist motivation is usually epistemological in the first place. The problem is this: if the internalist position entails a need for distinguishing between the way things are and the way they seem, his endeavour is jeopardised right at the beginning.\textsuperscript{161}

Stalnaker is critical of the strongly externalist position. He feels that the claim that neither linguistic nor mental content is purely internal conflicts with certain common-sense intuitions. He says:

Perhaps the externalist need not deny that what I say is what I think, when I am sincere, but it is hard to avoid the suspicion that if an externalist theory of speech and thought is right then we don’t really know what we are either saying or thinking.\textsuperscript{162}

Stalnaker’s answer to the Twin Earth problems – the information theoretic account of intentional content – is an attempt to integrate the valid points of both the internalist and the externalist position. The idea is, basically, that

\textsuperscript{160}In Burge, 1979.

\textsuperscript{161}Stalnaker, 1993, 300-302. The sort of dualism he points to is exemplified for example in the work of Ned Block, Block, 1986.

\textsuperscript{162}Stalnaker, 1993, 300.
...states of mind can carry information, when there exists a pattern of counterfactual dependencies between those states and corresponding states of the environment... Representational states and systems carry misinformation as well as information in the strict sense, but according to the information theoretic picture, misrepresentation must be understood as a deviation from the norm. It is reasonable to assume that representational states are normally correct – that they are states that tend to represent things as they are.\(^{163}\)

The dependencies between internal states and states of the environment have to be systematic because only then can we understand a misrepresentation as a deviation from a norm. Normally, states of mind tend to represent things as they are. An internal state represents the world as being such that \(P\) if under normal conditions it would carry the information that \(P\). An information carrying state is a belief if it not only carries information but also monitors the systematic dependency between an agent’s environment and his actions.

It may be very difficult, if not impossible, Stalnaker admits, to give a non-circular account of normal conditions. For the time being, therefore, we have to make do with talking of a ‘tendency to carry such and such information’. Going back to the Twin Earth example, we can say that what makes it the case that my beliefs tend to depend on water is that water normally has certain observable properties, and that it normally is the only kind of thing with those properties around. These conditions do not obtain on Twin Earth, and that is why there the same internal states fail to carry information about Earth water.

The information theoretic story treats propositional content in terms of causal regularities and counterfactual dependencies that tend to hold under normal conditions. As to their particular form, such regularities are contingent, and at least partly external. Internal facts carry information in virtue of how the world tends to affect them. Therefore, if the world were different, the same internal states would carry different information.

\(^{163}\)Stalnaker, 1993, 302.
Using the information theoretic account, Stalnaker formulates his interpretation of the Twin Earth story: O’Leary’s internal state as he steps into a bathtub is about water because, approximately, if what is in the bathtub were not water, he would not be in the cognitive state he is in. The internalist position, according to which O’Leary’s beliefs are not about water but only about some stuff that looks like water (for example because he does not know that water is H\textsubscript{2}O), would lead to the rather undesirable consequence of sceptical doubt.

Stalnaker’s account predicts that O’Leary’s beliefs are, under normal conditions, about water (or whatever water-like substance that normally occupies a certain role in his environment), and that his reference would fail only if normal conditions suddenly failed to obtain. This contrasts with Putnam’s conclusions, according to which Twin O’Leary cannot have beliefs about water even though on the Twin Earth it is normal that the stuff that looks like water is XYZ.\textsuperscript{164} This conclusion, Stalnaker would say, is undesirable because we might be well be in O’Leary’s position, and on Putnam’s view\textsuperscript{165} our reference would then systematically fail. It would seem, therefore, that both the strongly internalist position and the position Putnam takes, lead to epistemologically unpalatable conclusions.

We have already noted that which conditions are normal is a contingent matter but the notion of normality itself is modal. In asking whether some condition is normal in a counterfactual situation, we ask either of two questions: ‘Would it be normal in the actual world?’, and ‘Is it normal relative to that counterfactual situation itself?’

Once we make this distinction, we can get at the information theoretic answer to Twin Earth: Normally, O’Leary’s internal state is about water (H\textsubscript{2}O), because that is the main source of his information. If water (H\textsubscript{2}O) suddenly changed to XYZ, O’Leary’s reference to water would fail. On the other hand, if it were normal in the

\textsuperscript{164}The same point could be made about the brain in the vat argument – it is normal for the envatted brains to be envatted, so why should they not be able to refer? Again, it is a crucial point of Putnam’s story that they cannot refer.

\textsuperscript{165}See p. 137.
actual world that water is XYZ, then XYZ would be what we would mean by ‘water’.

We have to recognise that attribution of content is context-sensitive, and that what normal conditions are is relative to context as well. Attribution of content is essentially contrastive. When we ask whether the stuff in the bathtub is water, we presuppose that we have a good enough concept of water based on our normal conditions. To us, water is the stuff that is normally in the lakes, has certain observational properties, etc. And, as Stalnaker says,

... attribution of content must be made relative to presumed facts about the background normal conditions. But any such can be called into question. The context of attribution can change, it changes what we say, but we don’t have to change our minds.\footnote{Stalnaker, 1993, 309.}

3.3.6 Content

We have seen that Stalnaker provides us with two accounts of content. Firstly, there is the content of assertions made in a particular context. That is what is described by propositional concepts. I have tried to show that this is, in fact, a narrow content.\footnote{The term and concept of narrow content is introduced in Block (1986).} Secondly, there is the notion of content derived from the information theoretic account. This notion of content is designed to capture the contribution of the environment, so it seems to be intended as an account of broad content. We should, however, keep in mind that the problem with which many theories of content struggle, and, indeed, the main task of a theory of content, is not just to describe the two kinds of content but, crucially, to explain the relation between them.\footnote{We can express this in various idioms, e.g., in terms of a relation between internalist and externalist intuitions, coherentist versus metaphysical theories of truth, internal versus external anchors, but the core of the problem remains the same.}

Stalnaker explicitly addresses the notion of narrow content in several articles, arguing, broadly speaking, against the usefulness of
the notion. The explicit aim of his project with respect to content is to develop a notion that bridges the gap between the narrow and the broad notion. If, however, what I argued for in previous sections is correct, then it would seem that what Stalnaker says he wants to do and what he actually does are two different things. That the actual work does not quite live up to the explicit project happens often enough and it need not trouble us per se. It would be troubling, however, if we found out that the two notions of content Stalnaker develops clash.

In order to clarify the relation between the notions of narrow and broad content as Stalnaker describes them, we shall have to explain his use of diagonalisation. This brings us to the locus classicus, the place where an analysis of ‘O’Leary believes that Hesperus is Mars’ is proposed.

In brief, the example runs as follows: O’Leary believes that Hesperus is Mars. He chooses to express his belief by saying ‘Hesperus is Mars’. Stalnaker analyses this assertion as follows: Let us suppose that Kripke convinced us that proper names are rigid designators, that is, that they refer to the same object in all possible worlds. A proposition expressing identity between two rigid designators, here the proper names ‘Hesperus’ and ‘Mars’, is then either necessarily true or necessarily false. Indeed, this example targets two points: one about the behaviour of proper names, the other about believing in necessary propositions.

In constructing a propositional concept of the assertion ‘Hesperus is Mars’, we take into account the world O’Leary believes to be in. Because O’Leary asserts what he believes, in his world the assertion is true. And because his assertion involves the identity of two rigid designators, it is necessary. In the actual world – as anyone who ever studied identity statements knows – this assertion is false, and necessarily so. The propositional concept then looks as follows:

\[ F \]

\[ \text{For example in Stalnaker, 1989, and Stalnaker, 1990.} \]

\[ \text{Stalnaker, 1987, 179.} \]
In \( j \), O’Leary’s world, ‘Hesperus’ designates Mars, while in \( i \), the actual world, it designates Venus.

There is a problem, Stalnaker notes, with ascribing a belief in a necessarily false proposition to an agent. Intuitively, one should think that speakers assert things they believe to be true. That is why in some cases, Stalnaker says, the content of an assertion is best described by a diagonal proposition (which we can read off squares \( ii \), \( jj \), etc.) derived from a propositional concept. The content of assertion is to be identified with the diagonal proposition in case the ‘standard’ interpretation would violate Gricean maxims, that is, when we would have the speaker assert something extremely implausible. In identifying the meaning of O’Leary’s utterance ‘Hesperus is Mars’ with the diagonal proposition of the propositional concept, we emphasise the fact that he made his utterance in ignorance of certain facts that were relevant to what he meant to say. In doing this, we avoid the need to saddle him with asserting a necessarily false proposition.

This example has a number of puzzling features. The first concerns the perspective from which it is constructed. In a previous example of a propositional concept, in the case of O’Leary being called a fool, the situation was described from the viewpoint of a speaker. A propositional concept is intended to capture the impact that a speaker’s assertion has on the belief alternatives of his audience. However, in this case, there is no audience. O’Leary is not imparting his views to anyone in particular. The situation involves just him and the actual world.

I think the preferred way of interpreting this propositional concept involves the assumption that it describes a situation from an outside point of view. In this case, the speaker is not O’Leary but someone describing the content of O’Leary’s beliefs. That person

\[\begin{array}{cc}
  i & j \\
  i & F & F \\
  j & T & T \\
\end{array}\]

\(^{171}\)Stalnaker, 1987, 179.
then also makes assumptions about what the actual world is. So even though this example was not intended as such, it is in fact a case of attitude ascription.

Further on, as I just mentioned, one of the possible worlds in the propositional concept above is identified as ‘the actual world’.\footnote{Stalnaker, 1987, 184.} Given the way the notion of propositional concept is set up, this seems rather surprising. Generally speaking, a propositional concept is supposed to represent context, that is a body of information or a

\[
\ldots \text{set of possible situations that are compatible with what is presupposed, or taken to be common ground, by the participants in the discourse.}\footnote{Stalnaker, 1998, 7.}
\]

If this is the case, then how can a particular viewpoint be identified with ‘the actual world’? What is the role of the actual world in a propositional concept?

It may be tempting to dismiss the use of ‘actual world’ in this particular example involving Hesperus and Mars as a slip of the pen. We could do that if this were the only place where Stalnaker invokes the ‘actual world’ in connection with a propositional concept. But we have already noted above, when first introducing the notion of propositional concept, that the viewpoint of the speaker was equated with the actual world there as well. In that example – where the speaker calls O’Leary a fool – the perspective of the speaker is identified with ‘the world as it is, the world we are in’.\footnote{Stalnaker, 1978, 317.}

In another example, when analysing a stipulation that ‘Julius’ is the name for the person who invented the zip,\footnote{Stalnaker, 1999b, 15.} Stalnaker equates the actual world with the world where the stipulation takes place. A similar thing happens in the analysis of the statement ‘Sherlock Holmes does not exist’.\footnote{Stalnaker, 1978, 330.} Given this evidence, we cannot dismiss the occurrence of ‘the actual world’ in propositional concepts as a slip of the pen. It figures in most examples of propositional concepts that Stalnaker gives in his work.
In looking for an answer to this puzzle, we should start by sum-
ming up some relevant characteristics of Stalnaker’s possible worlds.
The possible worlds of a propositional concept are said to represent
the context, that is the worlds that individual participants think
they could be in. As we pointed out earlier, the notion of context is
closely connected with the notion of presupposition and the notion
of belief.

In general, to understand the content of a person’s belief,
ask what the world would be like if the belief were correct.
What is the world like, according to the person’s conception
of the way the world is? If we can give a coherent account, in
our own terms, of a way, or a set of ways, that things might
have been which seems intuitively to represent correctly the
person’s conception of the world and his place in it, then we
will have explained his beliefs and attitudes to propositions
- objects which conform to the received doctrines.\textsuperscript{177}

We can thus assume that each participant thinks that her context
set includes a representation of part of the actual world in the sense
of a broad content. But it is important to keep in mind that, as
Stalnaker repeats in a number of places,

\ldots according to the conception of content I am presuppos-
ing\ldots a person thinking [a] thought will be an inhabitant in
each of the possible situations compatible with his knowl-
edge.\textsuperscript{178}

This captures the insight that the agents themselves cannot be quite
sure which of the possible worlds compatible with their knowledge
is, or is closest to, the actual world in an externalist sense. In order
to understand Stalnaker’s use of the notion of actual world within
the context of a propositional concept, we have to look again at his
reasons for accepting a possible world framework in the first place.

When analysing Lewis’s claims about possible worlds, Stalnaker
rejected the metaphysical interpretation of indexicality of ‘actual’.

\textsuperscript{177}Stalnaker, 1981, 135.
\textsuperscript{178}Stalnaker, 1981, 142.
He said that “the objective, absolute point of view is the view from
the actual world, and it is part of our concept of reality that this
should be so.”\footnote{Stalnaker, 1984, 47.} Therefore, we might add, it follows from our con-
cept of reality that we identify our particular point of view with
‘the actual world’. The speaker takes his or her point of view to
be the actual world, even though he or she may entertain doubts,
be uncertain, or lack information regarding various features of it.
Possible worlds are introduced to model our attitudes to the world
that surrounds us.

Given the strongly epistemic flavour of the motivation and use
of possible worlds in Stalnaker’s work, it is only fair to ask whether
his position can be described as realist at all. Stalnaker asks himself
the same question:

Is the form of realism about possible worlds that I want to
defend really realism? It is in the sense that it claims that
the concept of a possible world is a basic concept in a true ac-
count of the way we represent the world in our propositional
acts and attitudes.\footnote{Stalnaker, 1979, 234.}

We can now conclude that the ‘actual world’ that figures in Sta-
laker’s propositional concepts is an epistemic concept that captures
the perspective of the speaker.

It remains now to see how narrow content – for we have seen
now that the content described in propositional concepts is, indeed,
narrow – connects with broad content described in the information
theoretic account of intentionality.

### 3.3.7 Narrow and Broad Content

Let us recall that the problem with an account of narrow content is
that unless it is complemented with an account of broad content, it
is hard to explain why agents’ beliefs are about the external world
at all. We shall now look at how Stalnaker handles this problem.
In Stalnaker’s information theoretic account (as we described it in section 3.3.5), the content of the agents’ beliefs can be about the actual world (in the externalist, broad sense) because that is the environment agents are normally in. On the information theoretic account, the actual world is the main source and reference point of the information an agent possesses. In fact, whatever is the main source of information for an agent is that agent’s actual world.

If, for example, an object \( x \) normally causes an agent to think that \( x \) is present, then that agent’s beliefs are about \( x \), and \( x \) is the source of information about \( x \). The information theoretic story is about general norms: normally, beliefs about \( x \) are caused by \( x \).

In particular situations, such as those we find described by propositional concepts, the information theoretic account cannot help us decide whether a particular piece of information is true. When building a propositional concept, we assume that the worlds participants think they are in can vary in details, even in details that are relevant to the situation at hand. Usually, all conversation participants are in the same physical surroundings but their perspectives, their perceptions of what the ‘actual world’ is, may vary.

The information theoretic account tells us why people’s conversations, though happening on the level of epistemic possible worlds, are about the external world. It cannot help us decide who is right in a particular case. It does, however, explain why people’s conversations are not just about beliefs – it explains why beliefs are about the world that surrounds us. This is also the point where communication connects with action: if we see speech as a kind of action, the information theoretic story explains how it can have an impact not only on the beliefs of the audience but also produce a desired effect in the actual world.

As it stands, the information theoretic story works on a different level and is much more global than the analysis of assertions. The final picture we are left with, accounts for narrow content of speaker’s assertions in particular situations, and for broad content in a general framework. It has been proving very difficult to establish a link
between narrow and broad content.\footnote{See for example the analysis in Lepore and Fodor (1992).}

Stalnaker does not seem troubled by the narrow versus broad content problems. His solution to the Twin Earth thought experiment aims at bridging the gap between the content conceived of as internal to the speaker (narrow content) and content as an inherently public phenomenon (broad content). We may feel that regardless of Stalnaker’s own perception of what he suggests, the proposal concerning the Twin Earth is internalist from the word go. Stalnaker is not likely to see it that way. This is because he never explicitly acknowledged that the possible worlds in the propositional concepts are epistemic in nature, and that the content they describe is thereby necessarily narrow. Especially when discussing necessary propositions and semantics of rigid designators, Stalnaker seems to think he still works within a Kripkean framework, as if not fully accepting the ramifications of the differences between his and Kripke’s framework. That is somewhat regrettable.

My analysis of the connection between narrow and broad content, however well supported by a previous study of the key notions, is therefore necessarily largely speculative.

\subsection*{3.3.8 The Resulting Picture}

Over the course of this subchapter, we have gathered a number of somewhat puzzling observations. On its own, each of them is just a little crack in the surface of the standard interpretation of Stalnaker’s work. By the standard interpretation I mean the approach that sees Stalnaker as working with a notion of context as a set of participants’ presuppositions, and reads the possible worlds he works with as rather akin to Kripke’s possible worlds. On this interpretation, Stalnaker’s attitude to context is often likened to Kaplan’s. This ‘standard’ reading can be found, for example, in the works, of Ulrike Haas (Haas-Spohn, 1994, Chapter 2), Craig Roberts (Roberts, 1996), Massimo Poesio (Poesio and Traum, 1997), Alessandro Capone (Capone, 2003), and Ben Caplan (Caplan, 2004).
3.3. Stalnaker’s Worlds

A careful examination of the central notions of Stalnaker’s approach to semantics has led me to a different reading. Its presentation shall follow the cues of questions that were left unanswered in the previous sections.

When characterising Stalnaker’s motivation (section 3.3.1), we pointed out that it is likely that Stalnaker’s main problem will be to find and describe the right balance between internalist and externalist concerns. For Stalnaker, this is crucial because an investigation of representational mental states and their connections to the surrounding world is at the very core of his enterprise. Later, in section 3.3.5, we saw that as a general theory of content, he endorses a version of moderate externalism. We ran into problems when we tried to connect his overt agenda about content (which we found in Stalnaker, 1990 and Stalnaker, 1993) with some implications of his analysis of conversation. In analysing his notion of a possible world (section 3.3.2), we concluded that Stalnaker adopts a very weak version of realism. Actually, he uses possible worlds to model attitudes of conversation participants. Stalnaker’s possible worlds are basically epistemic in character. In a sense, for Stalnaker, the actual world is ‘I and my surroundings’, but it is a different sense than Lewis tried to argue for.

In the last two sections, I outlined a somewhat speculative interpretation of the connection between Stalnaker’s notions of narrow and broad content. Having shown that the content characterised in propositional concepts is narrow, and the content derived from the information theoretic theory is broad, I tried to see how they fit together. I conjecture that the broad, information theoretic account, is intended to underpin narrow content. The theory of how broad content is derived is too general to be of practical use in analysing particular situations. This is not just because it is only roughly sketched: it is inherent in its very set-up, and rightly so since it may well be the case that, in principle, external circumstances fail to fully determine the mental state of an agent. However, we can make assumptions about some things that agents normally believe in given external circumstances. This is why the epistemic content of agents’ beliefs (and, consequently, their utterances) can be seen
as linked, in an underdetermined and general way, to broad content and the external world.

Why did I not draw on other people’s work in presenting my interpretation of Stalnaker? Why was I able to find only one person whose views on Stalnaker are somewhat similar to mine?\textsuperscript{182} I suppose it is because Stalnaker himself does not make much explicit connections between the various parts of his work. What we have are articles and passages that present the general program (e.g., his views on narrow content), and then passages that deal with particular problems using a different framework (e.g., propositional concepts). Most people who use Stalnaker’s work focus on particular applications of his two-dimensional semantics to descriptive semantics, mainly in natural language semantics. My goal here, however, was to trace Stalnaker’s use of possible worlds and his notion of content throughout most, if not all, of his writings. In absence of explicit connections between various parts of Stalnaker’s writing, I had to speculate to a degree. But it has been my aim to try and figure out what Stalnaker would have said had he focused on the questions that I am dealing with.

3.4 Conclusion

In the course of this chapter, we analysed the possible world frameworks of Lewis, Kripke, and Stalnaker, focusing on the aspects of their proposals that have consequences for the treatment of proper names. We treated each proposal separately because that way we could best uncover the internal coherence of each, and that, in turn, allowed us to describe some less obvious features of these proposals. Our aim was to see how the particular details of setting up a possible world framework influence the way proper names and their content are treated.

We saw that a strongly realist approach to possible worlds – such as we encountered in Lewis’s work – necessitates the adoption

\textsuperscript{182}I have in mind a paper by Peter Alward (Alward, 2004), which draws on his Ph.D. thesis.
of worldbound individuals, which, in turn, has consequences for the applicability of rigid designation, making it trivial and of little if any use. Thus while we concluded that Lewis’s framework is very general, has amazingly much internal coherence, and is, at first sight, free of essentialist commitments, its trademark feature, strong realism about possible worlds, has pervasive consequences that make its applicability to the study of modal properties of proper names rather limited. In the next chapter, in sections 4.1.3 through to 4.1.5, we shall see that the strongly realist approach to possible worlds, coupled with absence of any in-built essentialist features, has pervasive and rather undesirable repercussions for a foundational semantics based on this approach.

In our study of Kripke’s proposal, we concluded that the main disadvantages of the Lewisian conception were absent. Kripke’s framework allows for transworld individuals, and makes – not surprisingly – good sense of rigid designation. We saw that in order to make the minimal rigid designation work, one has to presuppose an essentialist version of haecceitism. That, per se, we did not find objectionable. What gave us a cause for concern was the scope and scale of essentialist theses and principles regarding particular natural kinds (e.g., the kind ‘water’ or ‘gold’), the adoption of which was not necessitated by the rigid designation as such. We argued that such metaphysical theses cannot be derived from an analysis of natural language. Furthermore, metaphysical necessity, closely connected with Kripke’s style of essentialism, turned out to be a very important feature in determining the domain of Kripkean possible worlds. Altogether, we found the part of Kripke’s framework that relies on an adoption of scientific realism objectionable. Moreover, we shall see (in section 4.2) that the descriptive semantics Kripke proposes for treatment of individuals causes problems on the level of foundational semantics.

Our reconstruction of Stalnaker’s proposal was to some degree speculative. While we tried to stay within the spirit of his proposal, sometimes we had to hypothesise what his answer to a particular problem would be. We saw that Stalnaker works with epistemic possible worlds but also has a story about the connection between
Chapter 3. Modal Statements, Individuals, Essences

the epistemic content and broad content. All and all, of the three proposals, we found his to be the best suited to the study of the functioning of proper names in a natural language. This we shall be able to confirm later, when reconstructing his proposal for the foundational semantics of names in section 4.3. Our own proposal for the descriptive semantics of proper names (chapter 5), too, shall bear marks of Stalnaker’s influence.

We have now the opportunity to make some remarks on the status of the semantics of modal statements. We saw both in Lewis’s and in Kripke’s case that building a framework largely in order to capture modal intuitions is a risky enterprise. In Lewis’s case, his desire to say something about the relative likelihood of various counterfactual situations (and, perhaps, an independent belief in the existence of possible worlds) lead to a development of a framework which seems to have little to contribute to natural language semantics. In Kripke’s case, his modal intuitions were incorporated into a system that was designed to account for natural language phenomena. This, however, resulted in a potential clash between our linguistic and modal intuitions, which we regard as an undesirable consequence. Thus it seems that while a possible-world framework has to allow for the expression of some of our modal intuitions, it should be general enough to accommodate the differences between speakers. Idiosyncratic modal judgements should not be built into the framework if it is to play a more general role in the semantics of natural language. A possible-world framework should primarily be a suitable medium for capturing linguistic intuitions (for example, rigid designation of proper names), and the influence of modal judgements should be kept at a minimal level.

Throughout this chapter we kept our focus on the connection between the possible world set up, its essentialist preconditions, the notion of an individual, and the kind of necessity that ties these notions together. In the following chapter, we shall continue our investigation of necessity, bring into focus the notion of a proposition, and see how various frameworks deal with the questions raised by the extensional character of propositions, that is mostly the problems of logical omniscience and necessary identity statements.