Preface to the German Edition

This book is an only slightly altered version of my doctoral dissertation accepted by the Faculty of Modern Languages of Tübingen University in summer 1994.

I am greatly indebted to my supervisor Arnim von Stechow for the professional, personal, and unwavering support that he has given me over the years in which this work was written.

I would also like to thank Irene Heim and Ede Zimmermann for their comments, their encouragement, and their model. Ede Zimmermann’s papers on context theory and on proper names greatly helped me to clear up my own thoughts, and Irene Heim’s contributions to the theory of meaning and the conversations that we held had a decisive impact on my views.

My husband Wolfgang Spohn’s help has been of immeasurable value. In the long discussions with him during the early phase of my work, I could develop my ideas and put them in more concrete terms; and afterwards, too, he has been a critical companion to my work in each of its phases. For that and for all the help that I have received from him, I would like to thank him here with all my heart.

Bielefeld, May 1995

Ulrike Haas-Spohn
# Table of Contents

**Introduction** ......................................................... 9

Chapter 1: The Initial Situation ........................................ 15

1.1 *The Cognitive Value of Necessarily True Statements* ............ 16
1.2 *Extension and Intension, Index, Context and Character* ........... 21
1.3 *The Content of Belief* .................................................. 31
1.4 *Context Change and Propositional Concept* .......................... 40

Chapter 2: On the Formal Structure of Belief Contents ................... 53

2.1 *Taking Stock* .......................................................... 53
2.2 *The Subject Parameter* ................................................ 57
2.3 *The Time Parameter* ................................................... 70
2.4 *The Place Parameter* .................................................. 74

Chapter 3: Predicates ....................................................... 81

3.1 *The Character of „Water“* ............................................ 81
3.2 *In What Way is „Water“ a Hidden Indexical?* ......................... 88
3.3 *What is a Possible Context?* ......................................... 92
3.4 *The Meta-Language Objection* ....................................... 98
3.5 *Essentiality Conventions and Hidden Indexicality* .................... 106
3.6 *Other Examples* ........................................................ 111
3.7 *Hidden Indexicality and Rigid Designation* .......................... 116
3.8 *Predicates and the Problem of Internalism* .......................... 120
3.9 *Formal Character and Subjective Meaning* .......................... 124
Chapter 4: Proper Names ................................................................. 135
  4.1 The Character of Names and Their Hidden Indexicality .............. 136
  4.2 Empty Names, Descriptive Names ........................................... 144
  4.3 The Descriptive Content of Names and the Meta-Language Objection . 147
  4.4 Proper Names and the Problems of Informativity and Internalism ... 152
  4.5 Names as Ambiguous Expressions .......................................... 156
  4.6 Names as Overt Indexicals: First Variant ................................. 162
  4.7 Names as Overt Indexicals: Second Variant ............................... 182

Bibliography ................................................................................. 179
Names Index .................................................................................. 185
Subject Index .................................................................................. 187
Introduction

In this book I am going to explicate and defend an individualistic concept to meaning. Such an enterprise is met with a lot of scepticism these days; but why is that so?

Any serious theory of linguistic meaning must account for two intuitions relating to this notion: On the one hand, meaning is something objective, pertaining to linguistic expressions by their being part of a language, connecting them with the world and the objects in it and, in particular, supplying truth-conditions for sentences. Yet on the other hand, meaning is something subjective, something that the users of a language associate with linguistic expressions. In particular, it gives a sentence a content characterising the belief that someone expresses by uttering the sentence and that someone acquires if he hears the sentence and trusts the speaker.

The classical theories of meaning have always had this two-fold task in view – and at the same time strained their basic concepts with it. This has already been the case with Frege’s senses of linguistic expressions, and especially with his thoughts, the senses of declarative sentences. On the one hand, Frege explicitly defines the thought expressed by a sentence to be its objective truth condition. On the other hand, it is this thought which is grasped by someone who understands the sentence, and which thus delivers its subjective meaning. And in this respect Carnap’s intensions do not differ from Frege’s senses.

The difficulties with this approach condense in a question that I will refer to as the problem of informativity: How can two statements with the same truth-conditions differ in informational content, how can necessarily true statements be informative, and how is it possible for rational persons to believe necessarily false statements? This problem, especially as applied to identity statements, already played a crucial role for Frege; and Fregean senses were designed to solve it. But the fact that Frege has not been able to specify senses of proper names in a non-subjective way indicates that this solution is incomplete.

That the classical theories of meaning are overstrained by this double task became once and for all apparent with Kripke’s Naming and Necessity (1972). By an investigation of the senses or intensions of proper names, Kripke made it clear that the distinction between apriori and a posteriori and that between necessity and contingency are logically independent. If it is its objective meaning on the basis of which a sentence must be judged necessary or contingent and if it is the subjective meaning on the basis of
which it is apriori or aposteriori – informative or not informative for individual subjects – then, just like these distinctions, objective and subjective meaning fall apart.

This problematic situation was further aggravated by Putnam’s and Burge’s work. Up to then, it had been a tacit assumption that subjective meanings should internally characterise the psychological states and, in particular, the beliefs of persons. However, in his The Meaning of "Meaning" (1975), Putnam used the example of natural kind terms to show that objective meanings are not even determined by the internal states of competent speakers: objective meanings are not in the head. Under the above-mentioned assumption this again implies that the subjective meanings of expressions must not be identified with their objective meanings.

In his Individualism and the Mental (1979), Burge goes one step further, claiming that not even beliefs are in the head. If Burge were right, then even the subjective meanings of utterances, the beliefs persons express by them, could not be characterised internally or, in Burge’s terms, individualistically.

As I have already said, this particular conclusion I want to reject. However, Putnam’s and Burge’s arguments are basically correct and must therefore be taken seriously. They pose a second problem to the theory of subjective meaning, one which I will call the problem of internalism: How can the beliefs that subjects express by uttering sentences be specified in an internal or individualistic way, i.e., such that two materially absolutely identical persons can also be described as coinciding in their beliefs and subjective meanings?

The discussion about individualism is not confined to the philosophy of language, but has gained even more attention within the philosophy of mind where it is, however, not so much concerned with linguistic meaning; rather, the meaningfulness and possibility of individualistic and anti-individualistic positions are discussed in general and the debate employs notions like mental representations, or the language of mind. In the present work, I would like to abstain from these topics. This abstinence is motivated by a possibly naive conviction that off-hand all intuitive plausibility speaks for individualism, and that the debate between individualism and anti-individualism does, in that sense, not concern the pros and cons of two equal alternatives. This intuition is initially blocked by Putnam’s and Burge’s arguments. But if, by some construction, Putnam’s and Burge’s insights could be reconciled with an internal characterisation of subjective meaning, the intuitive order of things would be restored; and that is what I want to achieve.

The basic ideas for solving the problems of informativity and internalism on which my theory of subjective meaning rests go back to Kaplan and Stalnaker. In his Demonstratives (1977), Kaplan treats both of these problems in connection with indexical expressions like ‘I’ and ‘this’. Investigating the interaction between context-dependence and modality, he develops the concept of character: The character of an expression is a function which for each context yields the intension of the expression in that context – where, as usual, intensions are functions from possible worlds, or more complex indices, into suitable extensions. Using the concept of a character, Kaplan arrives at an explication of Kripke’s distinction between necessity and aprioricity, if only for sentences containing
indexical expressions. Kaplan then proposes to identify the subjective meanings of such sentences with their characters; objective meaning, on the other hand, only applies to utterances, i.e. sentences within a context and, as before, consists in the intension the sentence has in the context at hand.

In order to solve Frege’s problem of informativity, Stalnaker introduces, in his *Assertion* (1978), a similar two-dimensional concept of meaning, i.e., one that also allows for dually dependent extensions: the notion of a propositional concept. Like the characters of sentences, propositional concepts determine which propositions – understood as sets of possible worlds – utterances would express if they took place in different contexts. Among the possible contexts there are then some in which the utterances belong to a completely different language than the one under investigation; this is the point at which Stalnaker’s propositional concepts go distinctly beyond Kaplan’s characters. At the same time, Stalnaker does not identify subjective meanings with the propositional concepts themselves but with their so-called diagonals. This identification is important because it allows for a one-dimensional concept of meaning that can be used to characterise belief contents. Another key point of Stalnaker’s approach is his holistic account of subjective meaning: What an utterance means to a subject does not only depend on the semantics of the sentence uttered but also on the subject’s previous beliefs. Hence propositional concepts are only defined for those contexts that the subject takes to be possible according to his belief state. They are thus not understood as objective semantic properties of sentences but as subjective properties of specific sentence tokens.

The theory which I am going to develop here can now, rather roughly, be sketched: I am assuming something which is usually taken for granted in linguistics semantics, but sometimes neglected in philosophy of language, viz. that a semantic theory for a given language must come up with a recursive meaning assignment. In principle, then, a recursive definition of the characters of the syntactically well-formed expressions, as demonstrated by Kaplan in his Logic of Demonstratives, is what we are after. For the time being, I will adopt the hypothesis that the character function already satisfies the above-mentioned double function of the concept of meaning: On the one hand it is an objective property that expressions possess qua being part of a language and that determines their relation to the world depending on context and index; on the other hand, it is supposed to represent the semantic knowledge that competent speakers of a language have. However, I will not take subjective meaning to be this semantic knowledge as coded in the character function, but rather – following Stalnaker – as given by the diagonal of the character. To account for Stalnaker’s holistic approach, the diagonal must always be regarded on the background of the subject’s belief set.

A subject’s belief set is meant to represent the totality of his beliefs. The idea of construing belief sets – and belief contents in general – as sets of possibilities originates with epistemic logic: a subject’s belief set is the set of those possibilities that are not excluded by his beliefs. The elements of this belief sets are also called the subject’s doxastic alternatives. It will, of course, have to be explained how belief contents can be characterised internally in terms of doxastic alternatives.
The diagonal of a sentence is then defined as the set of contexts in which the sentence is true. The claim that the diagonal represents a subject’s belief and thus a super-set of his belief set, only makes sense if it can be shown that belief sets can, or even must, be construed as sets of contexts, i.e., that a person’s doxastic alternatives have the formal structure of contexts. The second chapter will be devoted to this demonstration; there I will rely on the analysis of de se beliefs presented by Lewis in his *Attitudes de dicto and de se* (1979b). The core of the chapter is a presentation and defence of Lewis’s conception.

Everything else then hinges on one crucial point: There is only hope of ever solving the problem of informativity *via* a concept of subjective meaning in terms of diagonals of characters if these characters are not constant functions, i.e., if the sentences in question contain any context-dependent expressions at all. For otherwise the sentences’ diagonals would come down to the intensions they have in all contexts – which would be fatal because the observation that the intensions of sentences do not capture their informational content was, after all, our starting point. The only thing that is clear, then, is that the approach just sketched will be successful as applied to sentences that contain obviously context-dependent expressions like “I” and “this”; that much had already been demonstrated by Kaplan. But whether this approach still gives the right results when applied to Kripke’s and Putnam’s problems with natural kind terms and proper names, is anything but clear; it will first have to be shown that these expressions too are context-dependent.

In the third chapter I will address predicates in general and natural kind terms in particular. There I will use Putnam’s location of *hidden indexicality* and show how this hidden indexicality can be construed as context-dependence in Kaplan’s sense. Natural kind terms accordingly do not have constant characters so that the strategy of solving the problem of informativity carries over to them.

So far, the concept of subjective meaning is not going to meet the condition of internalism; the starting hypothesis that Kaplan’s notion of character can account for both subjective and objective meaning because it covers hidden indexicality, can no longer be upheld. In order to deal with the problem of internalism, we need a still more abstract notion of character, one that I will dub ‘formal character’. It abstracts from the language under consideration, turning the language to which an expression belongs into a contextual parameter, thus capturing the idea behind Stalnaker’s propositional concept. On the other hand, formal character is defined for syntactically identified expressions in a systematic and recursive way. But like Stalnaker’s it is a purely subjective concept; objective meaning can only be obtained from it by specifying the language of context. The above explanation of subjective meaning remains valid; at the same time the holistic element just mentioned becomes indispensable; and nothing will, according to this view, correspond to the traditional category of semantic knowledge.

In the fourth chapter I will apply the strategy developed in detail for predicates to the analysis of proper names. The question of their hidden indexicality can be answered in a completely analogous manner. However, proper names add a new problem, namely, their
multiple use. I will therefore have to address the surprisingly complex question of whether this multiple use is to be understood as a form of ambiguity, or rather as a form of overt indexicality.

Two expansions will at that point suggest themselves but will nevertheless remain unexplored. First, the semantics of proper names naturally ought to be embedded in a comprehensive semantic analysis of noun phrases – which would, however, involve a thorough treatment of overt indexicality, anaphoric reference, quantifiers, and other phenomena. That this embedding is missing will become apparent at the end of the fourth chapter, by the emergence of some open questions.

Second, it seems as if a thorough treatment of subjective meaning, and hence of beliefs and their contents, should be of immediate value for a semantics of belief sentences. I will, indeed, arrive at an account of de dicto belief reports; but I will not dig any deeper into the problem of belief sentences.

As far as semantic recursion will be made explicit here, it will not go beyond its first step, predication. Further steps, especially in the directions indicated, are yet to be made; only then the theory of subjective meaning proposed here would be well-founded. But I hope it can already be seen to be on the right track.
Chapter 1

The Initial Situation

The crucial keywords concerning the problems this work is dealing with and how it will be dealing with them, have already been given in the Introduction. However, our initial situation needs a much more precise and thorough account; only with this background will it be possible to understand our further proceedings.

Section 1.1 starts with our first core problem, Frege’s problem of informativity. It shows why Frege and Russell were unable to give a proper treatment of that problem; and it introduces aggravations due to Kripke’s and Putnam’s analyzes of proper names and natural kind terms. One aspect of the concept of subjective meaning is to explain the potential informativity of statements, and I will present various cases where truth-conditions, i.e. objective meanings of statements, cannot account for this aspect.

After this first exposition section 1.2 explicitly introduces the semantic concepts needed for an exact formulation as well as for a solution of the problems. To that end, Carnap’s extensions and intensions will be defined and, more importantly, Kaplan’s theory of context-dependence will be briefly presented. The central concepts are that of a character and related ones – particularly the diagonal.

Section 1.3 delves deeper into the problems. Assuming that the concept of subjective meaning is suitable for accounting for belief contents, I will explain why this goal cannot be attained by way of objective truth-conditions. The cases of Quine’s Orctutt, Kripke’s Pierre, and Burge’s arthritis will prove to be relevant here, because they show that neither de re reports nor de dicto belief ascriptions are internal descriptions of belief contents. Hence the traditional conception cannot be maintained, according to which subjective meaning can be equated with the interpretation of that-clauses.

In section 1.4 I am going to present Stalnaker’s approach to the subjective meaning of statements in terms of his theory of context change. This mainly involves an account of his notion of a propositional concept, which is formally reminiscent of Kaplan’s notion of a character, but based on a different, though not very transparent conception. This is important because my further strategy is even more inspired by Stalnaker than by Kaplan.

The steps to follow thereafter can be sketched more easily at the beginning of the second chapter, once the situation at the end of this chapter has been summarised.
1.1 The Cognitive Value of Necessarily True Statements

Equating objective and subjective meaning involves an identification of the truth-conditions of a statement with its informational content or, using Frege’s term, its cognitive value. If a statement is true and a person knows its meaning – i.e., its truth-conditions – the statement informs the person that the world must be such as to satisfy those truth-conditions. It follows that two statements with the same truth-conditions also have the same cognitive value and that a statement which is necessarily true has no informational content at all, because it does not restrict the existing possibilities as to the nature of the world.

There are, however, many statements which, according to a likely semantic analysis, express necessary truths and which can nevertheless be informative; and there are many pairs of statements to which one would want to ascribe identical truth-conditions and that still do not always, for every person, have the same cognitive value. An outstanding example is given by identity statements as already discussed by Frege (1892), who states the problem in the following way1:

"a = a and a = b are obviously statements of differing cognitive value; a = a holds \textit{a priori} and, according to Kant, is to be labelled analytic, while statements of the form a = b often contain very valuable extensions of our knowledge and cannot always be established \textit{a priori}. … Now if we were to regard equality as a relation between that which the names ‘a’ and ‘b’ designate, it would seem that a = b could not differ from a = a (i.e. provided that a = b is true). A relation would thereby be expressed of a thing to itself, and indeed one in which each thing stands to itself but to no other thing."

An identity statement like "the evening star is the evening star” expresses something that is necessarily true; it is uninformative, devoid of any cognitive value: nobody could deny, nobody could \textit{fail} to believe, that the evening star – or whichever object – is identical to itself. This is not so with "the evening star is the morning star”: this statement is obviously informative, it is very well possible for someone not to know, or not to believe, that the evening star is the morning star. At the same time it appears as if by "the evening star is the morning star” one expresses the very same thing as by "the evening star is the evening star”, i.e., a necessary truth: "the morning star” refers to the same object as "the evening star”, viz. the planet Venus, which is thus again said to be self-identical.

Frege himself tried to solve this apparent contradiction by distinguishing between the reference – his \textit{Bedeutung} – and the sense (\textit{Sinn}) of linguistic expressions: even though both "the morning star” and "the evening star” designate Venus, this planet is presented by "the morning star” in a different way than it is by "the evening star”. That is, the two expressions have different senses even though their reference is the same. Accordingly,

---

1 p. 56. Here and in what follows all page references to Frege (1892) relate to the English translation.
the sense of the entire sentence "the evening star is the evening star" is different from the
sense of the sentence "the evening star is the morning star", because the sense of the
whole is a function of the senses of its parts. This way it can be explained why one
believes the first statement without necessarily believing the second one, why the second
one is informative while the first is not; for it is the sense of a statement which constitutes
its cognitive value and which is believed, known, and doubted.

Russell (1905) too dealt with Frege’s problem and came up with his own solution.
Russell’s example is about George IV, who wants to know whether Scott is the author of
‘Waverley’ but certainly does not wonder whether Scott is Scott. Underlying Russell’s
thoughts there is the assumption that expressions with the same denotation can be
substituted for each other without thereby changing the truth-value of the sentence in
which they occur. Sentences like "George IV wanted to know whether Scott was the
author of ‘Waverley’” or "everybody knows that the evening star is the evening star”
seem to contradict this principle of substitution. Russell’s solution goes like this: definite
descriptions like "the author of ‘Waverley’” are not individual designators but
quantifiers, and sentences like "Scott is the author of Waverley” are not of the form "a =
b” but must be analyzed as "there is exactly one person who is author of ‘Waverley’ and
that person is identical with Scott”. This analysis no longer has a constituent
corresponding to the expression "the author of ‘Waverley’”, so that the problem of
substitutivity disappears. And at the same time the sentence’s potential informativity
becomes obvious.

Russell and Frege solve the problem of identity statements by rejecting a naive
semantic analysis of proper names and definite descriptions, according to which their
meaning lies in their reference. Consequently, sentences of the form "a = b” are not
bound to be necessarily true or necessarily false; only if "a” stands for the same name, or
the same (or a logically equivalent) description as "b”, do we have a tautology; otherwise
the result will be a contingent statement. This is the strategy of designing truth-conditions
so as to meet the requirements imposed on subjective meanings.

Frege’s and Russell’s position and the analyses pertaining to it are plausible as long
as one confines oneself to definite descriptions. The treatment of proper names, however,
leads to difficulties. Russell was forced to regard colloquial proper names as definite
descriptions in disguise, because on his view too there was no difference between "a = a”
and "a = b” in the case of proper names a and b. Frege faced the problem of specifying
the sense of a proper name. This was problematic in that senses had to be something
objective in Frege’s theory – something which is grasped by anyone with a mastery of the
language in question. But the sense of a proper name may, according to Frege, differ from
one language user to the next.²

² Frege (1892), p. 58, footnote *. 

---

"In the case of an actual proper name such as ‘Aristotle’ opinions as to the sense may differ. It
might, for instance, be taken to be the following: the pupil of Plato and teacher of Alexander the
Great. Anybody who does this will attach another sense to the sentence ‘Aristotle was born in Stagira’ than will a man who takes as the sense of the name: the teacher of Alexander the Great who was born in Stagira.”

In his lectures on naming and necessity, Kripke (1972) made clear that Frege’s and Russell’s theories and, quite generally, any approach to the meanings of proper names in terms of definite descriptions or senses corresponding to them were wrong. A proper name’s semantic contribution to the truth-conditions of a sentence in which it occurs only consists in designating a particular individual of which then something is said, and not in supplying characterising descriptions for that individual. Or, in Frege’s terms: proper names only contribute their reference to the truth-condition, not their sense.

If, e.g., one follows Frege and Russell in analyzing the name ”Aristotle” as ”the pupil of Plato’s who taught Alexander the Great”, one is forced to claim that the sentence ”Aristotle liked dogs” has the same meaning, i.e., the same truth-conditions, as the sentence ”the pupil of Plato’s who taught Alexander the Great liked dogs”. But this, according to Kripke, goes against linguistic intuitions. Both sentences are de facto true; yet the truth-condition of a statement does not only determine the actual circumstances under which it is true; it also determines when a counterfactual situation, a counterfactual course of history, would have satisfied the statement. Now, the first sentence would only be true in counterfactual situations in which a certain person, viz., the one we call ‘Aristotle’, likes dogs. The second sentence, on the other hand, correctly describes such courses of history in which some person who had been both Plato’s pupil and Alexander the Great’s teacher liked dogs – including even world histories in which Aristotle himself had never been Plato’s pupil or Alexander’s teacher. This shows that the name cannot be synonymous with the definite description; for, unlike the definite description, it designates the same individual in all counterfactual situations. Kripke speaks of proper names as rigid designators.

Kripke substantiates his rigidity thesis mainly by reference to the use of names in modal and counterfactual speech. For sentences like ”Aristotle did not have to be born in Stagira” or ”Aristotle did not have to teach Alexander” or ”if Aristotle had not studied with Plato” are felt to be meaningful statements – even if all we know about Aristotle is that he was a philosopher born in Stagira, or Plato’s pupil and Alexander the Great’s teacher. This, however, means that, even if we associate certain descriptions with a proper name, we do not talk as if these descriptions were the meaning of that name. From an intuitive point of view, none of the known descriptions of Aristotle gives a necessary property of Aristotle; but if one followed Frege and Russell in regarding the description as the meaning of the name, one would be forced to imply such necessities.

If names are rigid designators, identity statements with two proper names either express necessary truths, or else necessary falsehoods. If Hesperus is really identical with Phosphorus, there is no possible situation in which this identity is not valid. So the truth conditions of a sentence like ”Hesperus is Phosphorus” could not explain its potential informativity and Frege’s problem remains unsolved. Kripke does, of course, not deny that such identity statements can be of cognitive value, that it was an empirical discovery
that Hesperus is the same as Phosphorus and that, in a sense, the opposite could have turned out to be the case. And it would, of course, be absurd to claim that anyone who learns something new from "Hesperus is Phosphorus" is supposed to not have understood that sentence before. The conclusion thus seems unavoidable that objective meaning and subjective understanding of statements cannot be identified – at least not in the case of identity statements containing proper names.

But then difficulties also arise with sentences that are not identity statements. If proper names are rigid designators, a sentence resulting from another one by replacing a proper name with a different co-referential name has the same truth-conditions as the original sentence – this is at least so for the majority of syntactic constructions. Nevertheless, they could have different cognitive values; one may, e.g., compare "Hesperus is brighter than any other star in the evening sky" and "Venus is brighter than any other star in the evening sky". This general problem of substitution shows that Kripke's semantics of names has far-reaching consequences and that the question of the correct analysis of subjective meaning does not only arise for identity statements.

According to Kripke, not only proper names are rigid designators, but so are natural kind terms, i.e., expressions for biological species, chemical substances, or natural phenomena, like "tiger", "cat", "water", "gold", "light", or "heat". The objective meaning of such predicates, too, cannot be analyzed in terms of the descriptions that are usually associated with them; "tiger" and "wild feline of yellow tint with black stripes", are no synonyms, and neither are "gold" and "yellowishly shining, soft noble metal" or "light" and "that which makes the environment visible". Descriptions like this usually do not give essential properties of the kind, the substance, or the phenomenon. But whether something falls under a natural kind term or would fall under it in a counterfactual situation, does not depend on its superficial properties but only on whether it has the necessary essential properties: natural kind terms have this characteristic precondition that the objects or phenomena they are applied to form a natural kind, differing from other objects by deeper structural properties.

If, e.g., somewhere on earth one were to find layers of rock containing a material which looked exactly like gold and could also be processed in the same way but which, upon further analysis, proved to be of a different chemical texture, we would not call this material "gold"; it would not be gold. Or one might imagine that in those regions where gold is mined special conditions of the atmosphere led to optical illusions causing us to see gold as being yellowish; and suddenly these atmospheric conditions would disappear and we would see that gold is actually blue. Faced with that situation we would certainly not say that there is no gold, but rather that we had fallen victim to an illusion and that gold is just not yellow.

Let us assume that our chemical theory is correct, that gold is the element with atomic number 79, and that this is essential for gold. Then there is no possible situation in which gold is not the element with atomic number 79. Accordingly, the identity statement "gold is the element with atomic number 79" is a necessarily true statement. And so are sentences like "water is H₂O", "streams of photons of a certain energy range are light"
– or they are necessarily false, if our present theories are. Again we have identity statements the objective meaning of which turn out to be necessarily true, though they must obviously be regarded as informative statements. So the Fregean problem and the substitution problem pertaining to it also crop up with sentences containing natural kind terms: their truth conditions too cannot be identified with their subjective meanings, their cognitive values.

Considering identity statements that contain so-called indexicals, Kaplan (1977) discusses a further variant of Frege’s problem. Indexicals are expressions like “I”, ”here”, ”now”, ”yesterday”, ”this”, ”there”, ”the chair over there” – expressions whose meaning can only be adequately described if the utterance context, the situation in which they are uttered, is taken into account. Thus ”I” always refers to the speaker, the person who utters ”I”, ”here” refers to the place of utterance, and ”now” to its time; and the reference of ”this” or ”there” depends on which object or place the speaker points at while making his utterance.

Taken by themselves, indexical sentences, i.e., sentences containing indexical expressions, are not true or false; one can only ascribe truth-conditions to them if one views them in a particular context, thereby fixing the reference of the indexicals. If, e.g., the sentence ”the sun is shining here now” is uttered in Schwabing3 at 10 a.m. on March 20, 1990, exactly those world histories make that utterance true in which the sun is shining in Schwabing at 10 a.m. on March 20, 1990; on the other hand, the sentence ”the sun is shining here now” does not make a specific statement and thus does not have any specific truth conditions.

Indexicals turn out to be rigid designators too. As in the case of proper names, this becomes apparent if one looks at modal and counterfactual speech: for utterances of sentences like ”if I did not speak now,…” or ”I could be away from here now” do make sense; they could therefore not be synonymous with the statements ”if the speaker did not speak at the time of utterance,…” and, respectively, ”the speaker could be away from the place of utterance at the time of utterance”, which (at least on one reading) obviously make contradictory claims. So the description connected with an indexical, its descriptive meaning, does not enter the content, the truth condition, of an indexical sentence; rather, the description determines a referent in each context and this referent is then part of the truth condition. But this just means that indexicals are rigid designators or, as Kaplan puts it, directly referential.

This direct referentiality in turn implies that identity statements with indexicals are necessarily true or necessarily false. One may consider an utterance of ”this one is that one”, where the speaker, while uttering ”this one”, points at a person having coffee two tables away, and then, while uttering ”that one”, to a picture of a famous actor shown in the newspaper that the speaker is browsing through. This utterance is true if, and only if,

3 [Translator’s note: Schwabing is one of the livelier parts of Munich, Germany.]
the depicted actor is the person pointed at by the speaker. But if this is the case and if “this one” and “that one” do indeed refer to the same person, then there is no possible situation in which this utterance is false, because no individual is in any possible situation different from itself; of course, this does not mean that the same sentence uttered in a different context might not be false. Conversely, the utterance is necessarily false if its false at all. The same is true of utterance of sentences like “I am Franz Xaver Kroetz”. Only if Franz Xaver Kroetz utters this sentence will its utterance be true, and thus necessarily true. If anybody else is the speaker, he says something necessarily false by that utterance. However, such sentences are no doubt informative; for otherwise Franz Xaver Kroetz would, e.g., never have to introduce himself. So, in these cases again, the truth conditions of the utterance do not explain its potential informativity.

The consideration of indexicals calls for the introduction of new semantic distinctions. Thus, e.g., the descriptive sense of an indexical was distinguished from the meaning of particular occurrences of the expression. We talked about the truth conditions of utterances; but one would also want to account for something like the meaning of indexical sentences. In order to formulate this distinction and others, I will, in the next section, introduce the formal concepts providing the frame for the considerations to follow. The problem that rigid designators present for the identification of subjective and objective meaning, can then be given a more precise formulation.

1.2 Extension and Intension, Index, Context and Character

---

4 [Translator’s note: Franz Xaver Kroetz (*1946) is a successful Bavarian playwright who in the 80s had some memorable appearances on German television as an actor in a serial located in Schwabing (see previous footnote).]
In his attempt to improve Frege’s theory of sense and reference, Carnap (1947) systematically introduced the distinction between extension and intension. It still constitutes the foundation of formal semantics.

The extension of an expression is what was called its reference above and what Frege called its "Bedeutung". Up to now we have mainly been concerned with reference, which is the extension of individual terms, i.e., names, definite descriptions and pronouns: their reference is simply the object they denote. As to the extension of predicates, their number of places must be taken into account: the extension of a binary predicate is the set of objects to which the predicate applies, the extension of an n-ary predicate is the set of n-tuples of objects to which the predicate applies. The extension of a sentence is its truth-value; the truth-values are often written as the numbers 0 and 1, where 0 represents the value False and 1 represents True.

If an expression occurs in a syntactic construction in which it can be replaced by a co-extensional expression without thereby changing the extension of the whole construction, the latter is called an extensional construction and the expression is said to occur in an extensional context. Hence extensional constructions are characterised by the Extensional Principle of Compositionality: the extension of the entire expression is determined by the extensions of its parts and their mode of combination.

The concept of intension is meant to explicate the intuitive notion of meaning, thereby replacing Frege’s concept of sense. Intensions are constructed as functions that assign extensions to possible worlds. The idea behind this has already been mentioned: the meaning of an expression determines what the expression refers to, or would refer to under different circumstances; in particular, the meaning of a sentence determines what has to be the case in order for the sentence to be true.

So the intension of a sentence is constructed as a function from possible worlds to truth-values, a function assigning the value 1 to those worlds in which the sentence is true, and the value 0 to the worlds in which it is false. From a mathematical point of view, sentence intensions are thus characteristic functions on the set of possible worlds; the intension of a sentence can therefore be identified with the set of worlds that get the value 1. This set is the truth condition of the sentence or, as one also says, the proposition expressed by it. The set of all worlds is called the necessary proposition, the empty set is the impossible proposition. A proposition that is neither necessary nor impossible is called contingent.

The intensions of n-place predicates are functions that assign to any possible world the set of n-tuples of objects that lie in the extension of the predicate in that world. The intensions of one-place predicates are also called properties, and in the case of n-ary predicates I will analogously speak of n-ary properties.

Intensions of individual terms are functions assigning to a possible world the individual that the term denotes in that world; such functions are also called individual

---

5 As to definite descriptions, this is only true as long as they are not analyzed as quantifiers (following Russell), but rather as individual terms (following Frege and Carnap).
concepts. Following Kripke, an individual term whose intension is a constant function – assigning the same individual to every possible world – is called a rigid designator. In most languages of modal logic, variables are standard examples of rigid designators. If Kripke is right, proper names are rigid designators in natural language, but the intensions of definite descriptions normally are no constant individual concepts. Later, in section 3.7, I will discuss at length how the concept of a rigid designator is to be explicated for natural kind terms.

There are several syntactic constructions in which not the extensions of the parts, but only their intensions suffice to determine the extension of the whole expression. In this case, then, only intensionally equivalent expressions may replace each other without changing the extension of the entire expression. Examples are: the embedding of sentences under modals like "can", "must", "it is necessarily so that" as well as the colloquial "if ..., then..." construction. Such constructions are called intensional constructions and they are said to create intensional contexts. Intensional constructions satisfy the Intensional Principle of Compositionality, according to which the intension of a complex expression is determined by the intensions of its parts and their mode of combination.

As I have said, the concept of intension codes the idea that the meaning of a linguistic expression determines how its extension depends on the way the world is. However, closer inspection reveals that relativising extension and truth to worlds is not always enough. For there are expressions and sentences whose extensions not only vary with the world, but also with time and place. Thus, e.g., the extension of the description "the Chancellor of the Federal Republic of Germany" depends on the possible world – its extension is Helmut Kohl in the actual world, but would have been Rudolf Scharping if the Social Democratic Party (SPD) had won the last federal elections; but at the same time

---

6 Strictly speaking, Kripke (1972, pp. 269ff.) distinguishes rigid and strongly rigid designators. A rigid designator denotes the same individual in all worlds in which this object exists; a designator is called strongly rigid if it rigidly denotes a necessarily existing object, i.e., one that exists in all possible worlds. Usually, the objects denoted by proper names do not necessarily exist; one could thus have doubts as to the necessity of identity statements with proper names. However, if from this alone one concludes that a sentence like "Hesperus is identical to Phosphorus" might not express a necessary proposition, one would still have to admit that the sentences "if Hesperus exists, then Hesperus is identical to Phosphorus" is necessarily true. I do not want to go into the question of whether a rigid designator also denotes something in those worlds in which the object denoted elsewhere does not exist; the problems lurking behind this question are complex, but not relevant here. In what follows I will thus assume, without further discussion, that the intension is undefined on such worlds, and in that sense only partial. Mutatis mutandis, the same assumption is supposed to hold for the characters to be defined in a moment.

7 Intensional constructions have actually been defined as having the extension of the whole construction – and not its intension – depend on the intensions of the embedded parts. But then in recursive semantics the extension of an intensional construction must be defined for every possible world, because the construction itself may be embedded in an intensional context. An intensional construction, then, is intensional in at least one world, i.e., in at least one world its extension depends on the intensions of its parts. This, however, means that its intension too is determined by the intensions of its parts – just as the Intensional Principle of Compositionality would have it.
the extension of that description varies with time – in 1970 it is Willy Brandt in the actual world, in 1990 it is Helmut Kohl, and in 1998 it will be Gerhard Schröder in some possible world in which the SPD is going to win the next elections to come. The extension of the sentence ”it is raining” depends on both time and place; this sentence is true in some places at some times and false at other places and times.

These observations suggest that one should not construct intensions as functions on possible worlds, but as functions on more complex indices – e.g., indices that are triples of a world $w$, a time $t$, and a place $p$. In analogy to the rules of modal embedding, tenses and local constructions too can be described by recursive semantic rules: Just as one might say that a sentence of the form ”it is necessarily the case that $\phi$” is true in a world $w$ in case the embedded sentence $\phi$ is true in all worlds $w'$, one can now say that a sentence of the form ”always $\phi$” or ”sometimes $\phi$” is true at an index $<w, t, p>$ just in case, for all or, respectively, some time $t'$, the embedded sentence $\phi$ is true at the index $<w, t', p>$, and that a sentence of the form ”everywhere $\phi$” is true at the index $<w, t, p>$ just in case for all places $p'$, $\phi$ is true at $<w, t, p'>$.

However, such an expansion of indices is not enough for the analysis of indexicals. Rather, in giving a semantic account of indexicality one must have the extensions of expressions also depend on the context of utterance. That this dependence is not the same as the index-dependence discussed above becomes apparent from a comparison of indexical and non-indexical sentences. Let us consider the sentences ”on March 20, 1990, the sun is shining in Schwabing” as uttered at some arbitrary time, and ”the sun is shining here now”, as uttered in Schwabing on March 20, 1990. In a sense, the two utterances say the same thing, they have the same content, as Kaplan (1977, 1979) pointed out: they are both true in exactly those worlds in which the sun is shining in Schwabing on March 20, 1990. This truth-condition is expressed by the index-dependence of the extension. That the first utterance has this truth-condition is obvious; however, that it also attaches to the second utterance only becomes clear once the reference of the indexicals it contains has been determined by the context.

It follows that, in general, the intension of a linguistic expression is only given by the context of its utterance and that its extension is thus subject to a double dependence, on context and index. In this connection Kaplan also speaks of a need to distinguish between utterance situations and circumstances of evaluation. Moreover, from the necessity of this distinction Kaplan infers that two senses of linguistic meaning must be kept apart: the first one is the content – or as we have put it: the intension – that a linguistic expression has in a context; the second one is the so-called character of the expression, which is a function from contexts to intensions, viz. that function that assigns to each context the intension the expression has in that context.

In addition to the intuitive reasoning given above, there are theoretical arguments that show that the truth-conditions of indexical sentences can only be formulated in a recursive fashion if context and index are kept apart:

At first glance it does appear as if the introduction of complex indices – at least with a further expansion with a speaker parameter $s$ – already allowed for a treatment of
indexicals, because these indices could at the same time act as utterance contexts. One might then say that a sentence like "I am singing now" is true at an index \(<w, s, t, p>\) if \(s\) is singing in \(w\) at \(t\), or that a sentence like "it is raining here now" is true at index \(<w, s, t, p>\) precisely if in \(w\) it is raining in \(p\) at \(t\). However, such semantic rules could not be appropriately generalised. One problem to be encountered is this:

Consider the sentence "I am here now". It is part of the meaning that any utterance of it expresses something true, and semantic theory should reflect this. But the semantic rules just sketched can only achieve this if only such quadruples \(<w, s, t, p>\) were admitted as possible indices that are themselves contexts of utterance, i.e., those where \(s\) is located at \(p\) in \(w\) at \(t\); only then would the sentence be true at every index \(<w, s, t, p>\). But then these rules would have the same effect on the sentence "I am always here": according to its truth-condition, it is precisely true at an index \(<w, s, t, p>\) if for all \(t'\), "I am here" is true at \(<w, s, t', p>\). But if only indices that are also contexts are ever taken into account, the embedded clause – and thus the whole sentence – becomes true at every index. Intuitively, however, we know that utterances of "I am always here" are usually false and certainly not always true.

Lewis (1980) sums up the problems arising from identifying context and index in the following way. On the one hand, there are indexicals whose reference is given by contextually specified parameters. On the other hand, there are operators that ‘shift’ certain parameters of the index – temporal operators shifting the time component, modal operators shifting the world component, local operators shifting the place component; such operators demand the truth of embedded clauses to be evaluated only at indices with shifted parameters. But if contexts were to take over the role of indices, the result would be an inadequate treatment of these operators. For it would then be contextual parameters that are shifted by such operators. But shifting a contextual parameter does not necessarily lead to a possible context again, i.e., to a situation in which the respective speaker makes an utterance at the respective place and time. Moreover, such shifting must not be restricted to possible contexts; for we have just seen that otherwise utterances of "I am always here" – and similarly of "I am necessarily here" or "I am everywhere now" – could not be evaluated as false. If, on the other hand, indices are attributed the role of contexts, one cannot account for indexical expressions. For indexicals would then have to be evaluated at the given index. But these expressions can be embedded under operators that have shifted the index parameters; and then, inadequately, they would have to be evaluated at these shifted parameters.

Given such considerations, it follows that context and index must be strictly distinguished. This distinction must be kept apart from the further question of double indexing, the question of whether certain parameters must occur twice, both in the context and in the index. Rather, the criterion for parameters to occur in the index is: which intensional operators are there, and which parameters do they shift? And the criterion for

---

8 The line of reasoning essentially follows Kaplan (1979), S. 82f.
parameters to occur in the context is: which indexicals does the language contain, and which aspects of context do they refer to?\textsuperscript{9} Thus, e.g., it is often assumed (as in Lewis (1980)) that a speaker parameter is only needed in the context, not in the index, because there are no operators that shift the speaker. So the need for any parameters both in the context and in the index can only be justified by specific semantic analyzes.

The first time such double indexing was seen to be necessary was in Kamp’s (1971) investigations of the semantics of "now". "Now" always refers to the time of utterance; this calls for a fixed, unshifted temporal parameter. At the same time, however, another, shiftable parameter is needed for complex sentences with additional temporal operators. A sentence like, e.g., "those who do good now will be rewarded once" is true in a context and at an index if, and only if, there is a future index time \( t' \) at which those who do good at the context time \( t \) are rewarded – and not if those who do good at \( t' \) are rewarded at \( t' \), or those who do good at some time (as one could construe the sentence without "now").

It is time to explicitly introduce a few concepts that are important for the semantics of context-dependent expressions, and to fix some notation. The following definitions are mainly based on Zimmermann (1991) – one of the most comprehensive accounts of the theory of context dependence.

From now on we will write the character of an expression \( \alpha \) as \( \ll \alpha \ll \); so \( \ll \ll \) is the function to be recursively defined in a semantic account of a language. \( C \) shall be the set of possible contexts; \( c, c', c'' \) etc. represent elements of \( C \). \( I \) shall be the set of possible indices, \( i, i', i'' \) etc. represent elements of \( I \). A pair consisting of an expression \( \alpha \) and a context \( c \) is called an utterance (of \( \alpha \)); if \( \alpha \) is a sentence, \( \langle \alpha, c \rangle \) is also called a statement. The individual parameters of a context \( c \) are written as \( w_c, s_c, p_c, t_c \) etc., and similarly those of the index as \( w_i, t_i, p_i \) etc.

Which parameters context and index are supposed to contain\textsuperscript{10} will be subject to detailed discussion. But already now we will make the common assumption that contexts of utterance are special circumstances of evaluation, i.e., that contexts are at least as specific as indices; this means that all parameters of the index are also parameters of the context. Given a context \( c \), \( i(c) \) shall denote the index that results from eliminating the purely contextual parameters from \( c \)– if there are any.

The intension of an expression \( \alpha \) in context \( c \), i.e., the intension of the utterance \( \langle \alpha, c \rangle \), is: \( \ll \alpha \ll (c) \). The above-defined notions of an individual concept, a property, and

\textsuperscript{9} This does not mean that contexts must be defined as tuples supplying a corresponding parameter for each indexical. It rather means that contextual parameters must be assumed specific enough that the information coded in them covers all other aspects that may turn out to be relevant to the interpretation of indexicals. It is, e.g., sufficient to define contexts as triples consisting of a speaker \( s \), a world \( w \) and a time \( t \), because the place of utterance is then given by that place at which \( s \) is located in \( w \) at \( t \).

\textsuperscript{10} Contexts and indices are identified with triples of parameters and thus, strictly speaking, as descriptions of concrete situations. In the case of contexts, we assume a one-one correspondence between utterance situations and the corresponding parameter-lists; as to indices, no such assumption is made here – basically for reasons given in Lewis (1980).
a proposition must accordingly be re-interpreted as functions defined on indices instead of worlds alone. In particular, then, from now on propositions are defined as sets of indices. In actual practice, however, utterances of sentences usually express propositions for which only the world parameter is relevant, and where time and place parameter are idle because they are either bound or occupied by the values of the corresponding contextual parameters. I will thus continue to speak of sets of possible worlds as propositions. Context will always make clear what precisely is meant and the difference is usually irrelevant anyway.¹¹

The extension of an expression α in a context c at the index i is: ∥ α ∥ c (i). A sentence φ is true in c at i if, and only if ∥ φ ∥ c (i) = 1. Furthermore we would like to say when a statement < φ, c >, i.e. the sentence φ in context c, is true. This shall be the case if the proposition expressed by < φ, c > holds at c itself, i.e. if ∥ φ ∥ c (i(c)) = 1.

An individual term α is called rigid or directly referential if for every c it holds that ∥ α ∥ c (i) = ∥ α ∥ c (i'), for all i and i' (on which ∥ α ∥ c (i) is defined). Rigid expressions have constant intensions in all contexts.¹²,¹³

An expression α is called indexical or context-dependent if there are two contexts c and c' such that ∥ α ∥ c ≠ ∥ α ∥ c'.

An expression α is called absolute if ∥ α ∥ c = ∥ α ∥ c' for all contexts c and c'. Absolute expressions are not context-dependent; they have the same intension in every context.

An individual term is called absolutely rigid if it is rigid and absolute. The extension of an absolutely rigid expression depends neither on the context nor on the index.

An individual term is called descriptive if it is absolute, but not rigid. Descriptive expressions are context-dependent, but not index-dependent.¹⁴

¹¹ See also Kaplan (1977, p. 546). All page references to Kaplan (1977) relate to the 1989 published version.

¹² This definition corresponds to Kripke’s concept of rigidity. Kaplan would not subscribe to the restriction to indices at which the object denoted by α exists. For Kaplan the reference or a rigid of directly referential expression is totally independent of the index, i.e., it is always one and the same object, even at indices at which that object does not even exist (cf. Kaplan 1977, p. 492-494 and 497). As I have already said (cf. footnote VVV5 on p. VVV 23), I will not further investigate the difference between Kaplan’s and Kripke’s views, and not take the partiality in the above definition seriously.

¹³ Strictly speaking, Kaplan would like to make a difference between rigidity and direct referentiality. However, this difference cannot be captured by the formal apparatus of possible world semantics, but only appears when one resorts to structured propositions. For then one is in a position to say that the contribution a directly referential term makes to the proposition expressed by a sentence containing that term is just the object denoted, whereas the contribution of a rigid, but not directly referential expression consists in a complex condition that determines the very same object in every world (cf. Kaplan 1977, p. 494-497). I have, though, not been able to find out what theoretical role over and above rigidity direct referentiality is supposed to play. I am therefore not distinguishing between the two concepts here.

¹⁴ The concept of rigidity as well as all those depending on it are only defined for individual terms here, not for all kinds of expressions. For the present it is not at all clear what rigidity could mean in the case of predicates; I will, however, discuss this point in detail in section 3.7.
Kaplan defined a new operator, the so-called *dthat-operator* – here symbolised as $\nabla$ – which relates any index-dependences to the utterance context, i.e., it replaces the parameters of the index by the corresponding parameters of the context; the operator is defined by: $\| \nabla \alpha \| (c)(i) = \| \alpha \| (c)(i(c))$. For instance, the dthat-operator turns descriptive terms into indexical ones; thus "$\nabla$ the place of utterance" has the same character as "here".

Using the dthat-operator, one can also explicate the difference between the so-called attributive and the so-called referential readings of descriptions: thus when read attributively, the description "the inventor of the lightning-conductor" gets the character $(\text{the inventor of the lightning-conductor})_{\text{attr}} \| (c)(i) = \text{that individual in } w_l \text{ that invented the lightning-conductor in } w_l$. When read referentially, it can be understood in the following way: $\| (\text{the inventor of the lightning-conductor})_{\text{ref}} \| (c)(i) = \| \nabla (\text{the inventor of the lightning-conductor})_{\text{attr}} \| (c)(i) = \text{that individual in } w_c \text{ that invented the lightning-conductor in } w_c$. So the referential reading of a description is obtained by applying the dthat-operator to its attributive reading; the referential reading is thereby interpreted as directly referential – in the sense defined above. Maybe this interpretation does not quite capture what its inventor Donnellan had in mind (according to Donnellan, referentially used descriptions can apply to individuals that do not even have the descriptive property, see Donnellan 1966). Still, this explication of the concept has become quite common in the linguistic literature (cf. Heim 1991a, section 1.3), so that I will stick to it in what follows.

The so-called diagonal operator, whose usefulness in the philosophy of language has especially been advocated by Stalnaker (1978), is complementary to the dthat-operator $\Delta$. It will play a crucial role for us, which is why notation and terminology pertaining to it shall be introduced immediately. Stalnaker’s usage wavers between a syntactic and a semantic diagonal operator; we will need both.

The proper counterpart of the dthat-operator which, after all, is an expression of the object language, is the *syntactic diagonal operator* $\Delta$. It drags contextual dependences into the index, thus turning contextual parameters into index parameters. Of course, this only works if the above-mentioned assumption that contexts are at least as specific as indices may be turned round, i.e., if one assumes that contexts and indices are indeed structurally identical. Under this assumption $\Delta$ can be semantically defined by: $\| \Delta \alpha \| (c)(i) = \| \alpha \| (i)(i)$. Applying $\Delta$ to an indexical expression results in its descriptive paraphrase; for instance, "$\Delta$ here" is synonymous with "the place of utterance".

Even under the assumption of structural identity of contexts and indices, the syntactic diagonal operator is only partially defined in this way. After all, that utterance contexts are special circumstances of evaluation means that $C$ is a subset of $I$, and a proper one, i.e., that $C \subset I$.\footnote{\(\subset\) always symbolizes proper, \(\subseteq\) proper or improper inclusion.} Hence, for an utterance $< \alpha, c >$, $\| \Delta \alpha \| (c)$ is not well-defined for all indices $i$ from $I$, but only for those in $C$.\footnote{\(\subset\) always symbolizes proper, \(\subseteq\) proper or improper inclusion.}
The semantic diagonal operator – symbolised by \( \delta \) – assigns to each character a usually only partial character. So if \( \chi \) is a function from \( C \times I \) into a set of extensions and could thus be the character of a linguistic expression, \( \delta(\chi) \) is defined to be that function for which \( \delta(\chi)(c)(i) = \chi(i)(i) \). Formally, this means that \( \delta(\chi) \) is defined on \( C \cap I \times C \cap I \). However, since we are always assuming that utterance contexts are special circumstances of evaluation, the definition of \( \delta(\chi) \) only makes sense for the case that \( C \subseteq I \). So if \( C \) is a proper subset of \( I \), \( \delta(\chi) \) is a partial character only defined on \( C \times C \).

There is, of course, a direct correspondence between the syntactic and the semantic diagonal operators; we always have \( \ll \Delta \alpha \ll = \delta(\ll \alpha \ll) \). Distinguishing the two operators is thus more a matter of formal explicitness than of substance; and I will sometimes allow myself to speak of the diagonal operator \textit{simpliciter}.

We finally need the concept of the diagonal: For any expression \( \alpha \), \textit{the diagonal} \( \partial(\alpha) \) of \( \alpha \) is defined to be that function which assigns to any \( c \in C \) the extension of \( \alpha \) in context \( c \). \( \ll \alpha \ll \) \textit{is} \( \partial(\alpha) \) \textit{on} \( (i \ll c \ll) \). Diagonals, then, are functions from contexts to extensions. If \( \phi \) is a sentence, \( \partial(\phi) \) obviously corresponds to the set of contexts in which \( \phi \) is true; in what follows we will continue to go back and forth between representing the diagonal as a function and as a set. We will refer to the set of all contexts as the \textit{tautological diagonal}, and to the empty set as the \textit{contradictory diagonal}; a diagonal that is neither tautological nor contradictory is called \textit{informative}. If the assumption underlying the definition of the diagonal operator is met, i.e., if \( C \subseteq I \), we obviously have \( \partial(\alpha) = \delta(\ll \alpha \ll)(c') \) for some context \( c' \) (whose precise identity does not matter). But then \( \partial(\alpha) \) is also defined if this assumption is not met.

On the level of characters there is again a Principle of Compositionality to the effect that the character of a complex expression is a function of the characters of its parts. The question arises whether there are any constructions that are \textit{characterial}, whose extension is determined neither by the extensions nor by the intensions of its parts, but only by their characters. Kaplan (1977) calls such constructions monsters and he claims that there are no monsters in natural language, syntactic constructions being at best intensional. The only way of creating characterial constructions in natural language is by use of quotation marks. Thus, e.g., if \( \phi \) is an indexical sentence, the truth of “‘\( \phi \)’ is true in all contexts” depends on whether all utterances of \( \phi \) are true, and thus on \( \phi \)’s character. Moreover, the diagonal operator just defined is a monster, too. Kaplan’s thesis therefore implies that, in natural language, the diagonal operator could at best be rendered by quotation marks.

Traditional modal logic has always been concerned with one concept of necessity that Carnap (1947), §39, identified with the concept of logical truth but which can also be interpreted differently. However, our present framework at least allows for a distinction between metaphysical and epistemic necessity or, as we will also put it, between necessity and apriority:

Necessity does not apply to sentences but to statements or the propositions expressed by them. For any sentence \( \phi \) and context \( c \), \( < \phi, c > \) is \textit{necessarily true} just in case \( \ll \phi \ll (c) = I \) (\( I \) was the set of all indices), it is \textit{necessarily false} just in case \( \ll \phi \ll (c) = \emptyset \).
i.e., the empty set. Accordingly, $I$ is again called the *necessary* and $\varnothing$ the *impossible* proposition, and all other propositions are *contingent*.

Aprioricity, on the other hand, applies to sentences. A sentence $\phi$ is called *a priori (true)* in case $\phi$ is true in all contexts, i.e., if $\operatorname{I} \phi \downarrow (c) (i(c)) = 1$ for all contexts $c$ from $C$; hence sentences that are a priori true have a tautologous diagonal. *A priori falsehood* can be defined analogously. We will, moreover, follow Kripke (1972, p. 264) and identify analyticity with aprioricity plus necessity: A sentence $\phi$ is *analytic(ally true)* if it is a priori true and all of its utterances are necessary, i.e., if $\operatorname{I} \phi \downarrow (c) (i) = 1$ for all $c$ and $i$.

Logical truth is then defined as usual. A *model* $M$ of a language consists of $C$ and $I$, the character function $\downarrow \downarrow$, and a designated context $c_0$. A sentence is then called *true in model* $M$ if it is true in $c_0$; and it is *logically true* if it is true in all models. Thus, if one follows Kaplan (1979) and treats "I", "here", and "now" as constants, the above example "I am here now" will, according to this definition, be a priori true – even logically true – although its utterances normally are not necessary.

How does, after all these conceptual clarifications, the problem of section 1.1 present itself? There we had found that the objective meanings of statements containing rigid designators cannot be identified with their cognitive values or their subjective meanings. Given the above comments, it is clear that the objective meaning of a statement, its truth-condition, is here to be understood as its intension, the proposition it expresses. But it is still open what its subjective meaning should be. However, at least for those cases in which the rigid designators are indexical, we seem to have a hint to a solution of the problem.

We have seen that the character of an indexical sentence is not constant and therefore does not coincide with the truth-condition of a given utterance of that sentence. However, according to Kaplan – whom we will follow, in a way – the character function gives a sensible explication of the concept of meaning. One roughly speaks of meaning in this sense when one says that a language user knows the meaning of an indexical expression or its utterances. For even if he has a full command of his language, the user does not have to know the intensions of indexical utterances; he can only know them if he also knows the utterance context sufficiently well. Given this, identifying the character of an expression with its subjective meaning would seem to be the thing to do.

It is this solution to Frege’s problem that Kaplan (1977) proposes. But Kaplan’s proposal can only be satisfactory if two additional conditions are met. First, it would have to be explained in more detail how the character of a sentence can be understood as its subjective meaning, its cognitive value, and how, in particular, information transfer can be systematically described in terms of characters. Second, this solution would have to be carried over to all other cases of rigid designators. This would have to include a proof that any rigid designator is indexical, but not absolutely rigid. For absolutely rigid expressions do not just have a constant intension but also a constant character; and so identity
1.3 The Content of Belief

It is generally agreed that the concept of subjective meaning is supposed to achieve various things. We have so far concentrated on the aspect of informativity and the hearer’s perspective: the subjective meaning of a statement is to determine the information the statement conveys to a hearer, it is to be that which the hearer believes if he understands the statement and assents to it. There is, of course, a counterpart to this on the speaker’s side: the subjective meaning of a statement is to represent the thought that the speaker wants to express, that which the speaker believes if he seriously and sincerely utters a sentence, the information he thereby wants to convey to his hearers. The third role ascribed to subjective meaning derives from the other two and has to do with the semantics of attitude ascriptions.

If a person \( a \) utters an English sentence \( \phi \) or assents to \( \phi \), one can usually rightly claim that \( a \) believes that \( \phi \) – at least if certain conditions obtain: Thus, e.g., \( a \) must utter \( \phi \) seriously and sincerely, or assent to it seriously and sincerely; this also includes that \( a \) masters the language in which \( \phi \) is formulated. Furthermore, \( \phi \) must not contain any indexical expressions; for otherwise the sentence \( \phi \) would have to be changed in the belief ascription in the way typical for the transition from direct to indirect speech. That \( \phi \) be an English sentence is simply necessitated by the fact that otherwise ”\( a \) believes that \( \phi \)” would be a symbolic sequence that is neither well-formed in English nor in any other language. Given these assumptions, a sentence of the form ”\( a \) utters ‘\( \phi \)’” or ”\( a \) assents to ‘\( \phi \)’” implies the corresponding sentence ”\( a \) believes that \( \phi \)”’; Kripke (1979) has dubbed this connection the disquotational principle. Strictly, one should speak of the disquotational principle for English. But then one could formulate a disquotational principle for other languages, too. And given adequate English translations of utterances

---

16 For this reason Kaplan himself meanwhile no longer believes that character and subjective meaning can always be identified: According to Kaplan, proper names are absolutely rigid and Frege’s problem about Hesperus und Phosphorus can thus not be solved in terms of the characters of the names. (Cf. Kaplan 1989, pp. 597–9.)
in a foreign language, English also satisfies an extended disquotational principle, which reads: if \( a \) utters \( \phi \) or assents to an utterance of \( \phi \) and \( \psi \) is an English translation of \( \phi \), then \( a \) believes that \( \psi \).\(^{17}\)

Given the disquotational principle, the obvious thing to do is to interpret belief ascriptions of the form “\( a \) believes that \( \phi \)” as expressing a relation between person \( a \) and the subjective meaning of \( \phi \) for \( a \); after all, according to its first two functions, the subjective meaning is supposed to be whatever a person believes when uttering, or assenting to, \( \phi \). Ascriptions of other attitudes, like “\( a \) wants that \( \phi \)”, “\( a \) knows that \( \phi \)”, “\( a \) hopes that \( \phi \)” etc., are then analyzed in analogy to belief, as relations between \( a \) and the subjective meaning of \( \phi \). The third role of subjective meanings thus consists in their serving as the interpretations of complement clauses to attitude verbs and thus as contents or objects of attitudes.

It is usually assumed that these three roles are but different aspects of the same thing and that one cannot explain one without deciding about the others. It is still important to keep the three aspects apart because from their different perspectives different problems are to the fore. In this section we will concentrate on the aspect of attitude ascriptions and, from that point of view, present further reasons for not identifying subjective and objective meanings of statements.

As is common, e.g. in epistemic logic, we are going to model a person’s belief content by a set of possible worlds, which we will call the belief set of that person. The underlying idea is that, like linguistic statements, people’s belief contents or thoughts can be true or false and thus have truth conditions. Moreover, this model assumes that belief contents can be individuated by their truth conditions, so that distinct belief contents always have distinct truth conditions. A person’s belief set then consists of exactly those possible worlds that make all his beliefs true, those that are possible according to, and not excluded by, what the person believes; in other words, it is precisely the belief set and no smaller set in which the person takes the real world to be. We will refer to the elements of a person \( a \)’s belief set as \( a \)’s doxastic alternatives. Accordingly, a person \( a \)’s beliefs are inconsistent just in case \( a \)’s belief set is empty, i.e., if \( a \) does not have any doxastic alternatives.

The belief set model – and this is going to be crucial for the following discussion – is supposed to be an internal (or individualistic) characterisation of a person’s beliefs. That is, whether a possible world \( w \) belongs to a given person’s belief set is solely determined by that person’s inner psychological state or, as Putnam (1975, p. 220) would have it, his narrow psychological state. The usually understood criterion of membership can be roughly characterised like this: Imagine the actual epistemic state of a person as fixed and then place him in a world \( w \) which he may investigate in each and every detail. If he then finds no clues that \( w \) is not the actual world – if, in other words, he can in no way

\(^{17}\) Kripke (1979) proceeds in a different way. He formulates a principle of translation to the effect that belief ascriptions from different languages can translate one another. This difference does not affect the following discussion.
distinguish \( w \) from the real world as he knows it – then, and only then, will \( w \) be an element of his belief set. As a result, a proposition \( p \) is believed by a person if \( p \) is a superset of the person’s belief set; for it is then that he takes the actual world to be in \( p \).\(^{18}\)

As has been said above, this is only a rough and vague picture which will however become sharper. For we will see directly, and not just by way of the negative arguments of section 1.1, that the objective meanings of statements a person would assent to do not offer an adequate characterisation of his beliefs, and that it is therefore not appropriate to say that statements of the form “\( a \) believes that \( \phi \)” are true if the proposition objectively expressed by \( \phi \) is a superset of \( a \)’s belief set.

That the objective meaning of a complement clause cannot in general, but at best in the case of so-called de dicto belief ascriptions, serve as a belief content has already been shown by Quoin (1956). We will call a belief ascription de dicto if it can be reduced to an application of the disquotational principle or its extension: a statement of the form “\( a \) believes that \( \phi \)” is a de dicto belief ascription if and only if it says that \( a \) would assent to \( \phi \) (or an utterance of a translation of \( \phi \) into \( a \)’s language). These ascriptions must be distinguished from so-called de re belief ascriptions that take subjects’ beliefs to be about specific objects and where what the speaker is after with his belief ascription is mainly to

\(^{18}\) Another point should be mentioned here, if only to exclude its treatment from the present work: If a person’s belief is represented by a set of worlds – and not, say, by a set of propositions – one is forced to regard a person’s beliefs as deductively closed. This assumption is usually taken to be problematic, if not downright false; to see the totality of logical consequences of what one believes seems superhuman.

However, this kind of assessment always assumes that belief contents can be characterized by the objective meanings of statements, i.e., that “\( a \) believes that \( \phi \)” is true just in case the proposition objectively expressed by \( \phi \) is a superset of \( a \)’s belief set. It is less clear whether the same problem – which, following Stalnaker (1984), we will call the problem of deduction – still arises if, instead of the objective meaning, one uses the subjective meaning of \( \phi \) to characterize the belief content; of course, here all depends on how the concept of subjective meaning is defined.

I must, however, admit that the problem of deduction presents a serious challenge even to the framework of the theory to be developed in the course of this work, and that I do not have a complete solution to it. On the other hand, it must be pointed out that our investigations into objective and subjective meaning are not made dispensable by the possibly most important strategy for treating the problem of deduction, the theory of so-called structured propositions. The problem of deduction is just a special case of the general problem of accounting for the fact that attitude verbs create so-called hyperintensional contexts, i.e., contexts in which not even intensionally equivalent expressions can be substituted for each other salva veritate; the problems presented in section 1.1 belong here, too. The theory of structured propositions tries to account for hyper-intensionality by identifying the objective meaning of a sentence (in a context) not with a truth condition, but with a structured complex made of the meanings of sentence parts. But if these meanings are to be objective meanings in the sense explained above, this strategy is of no help with the problems of section 1.1: The two words “Hesperus” and “Phosphorus” have the same meaning which cannot be structured any further; hence complex sentences that only differ in these words will be assigned the same structured propositions – which again cannot explain their different cognitive value. So even the theory of structured propositions must be supplemented by a theory of subjective meaning.

My theory of subjective meaning will be able to solve the problem about Hesperus and Phosphorus, as well as some further problems. But it fails on those aspects of the problem of deduction with which the theory of structured propositions deals more successfully; viz. beliefs of mathematical and logical truths.
get this objectual reference across; it is then less important how the objects have been presented to the subject. Since in de re ascriptions reference is more at stake than belief content, the speaker can even describe the objects in terms the subject would never use in describing his own beliefs; accordingly, in de re ascriptions, co-extensional individual terms can be substituted for each other salva veritate.

Most belief ascriptions, then, possess both de re and de dicto readings and attempts have been made to describe this ambiguity in syntactic terms, as a scope ambiguity: the noun phrases in the complement clause can have narrow or wide scope in the logical form of the whole sentence. Narrow scope represents the de dicto reading of the noun phrase, and as the belief predicate does not provide an extensional context, on this reading co-extensional descriptions must not be substituted for each other. However, on a de re reading the noun phrase would stand outside the scope of the belief predicate, which would make substitutions possible, due to the extensional substitution principle.

Quoin (1956), though, made clear that de re readings cannot be adequately captured this way. Quine’s well-known example is the story of Ralph. Ralph takes a certain man with a brown hat, whom he has seen various times under dubious circumstances, to be a spy. Furthermore, there is a man with grey hair, whom Ralph only knows superficially, but then rather as a pillar of society, and whom he as only once seen consciously, which was on the beach. Ralph does not know that the man with the brown hat and the man on the beach are one and the same person, Bernard J. Orcutt. In this situation the sentences "Ralph believes that the man with the brown hat is a spy” and "Ralph believes that the man on the beach is not a spy” intuitively have true de re readings; in semi-formal notation, these could be rendered as, respectively, "that \( x \) that is a man with a brown hat, satisfies: Ralph believes that \( x \) is a spy”, and "that \( x \) that is a man on the beach satisfies: Ralph believes that \( x \) is a spy”. However, since the two descriptions have the same extension, these two sentences imply the sentence "there is a person \( x \) that satisfies: Ralph believes that \( x \) is a spy and Ralph believes that \( x \) is not a spy”.

If one now wants to argue that this belief ascription is true just in case the intension of the complement clauses "\( x \) is a spy” and "\( x \) is not a spy” are supersets of Ralph’s belief set, this would mean that Ralph is characterised as someone who takes a contradiction to be true. For the two occurrences of the variable \( x \) are bound by the same quantifier, which means that the intension of "\( x \) is a spy” is the set of all those worlds in which a certain person is a spy and that of "\( x \) is not a spy” is the set of worlds in which the same person is not a spy. However, these two sets have an empty intersection and, consequently, can only be supersets of Ralph’s belief set if the latter were the impossible proposition. But, from an intuitive point of view, one would not want to ascribe Ralph an internally contradictory belief.

So without further ado de re ascriptions cannot be analyzed in terms of wide scope and “quantifying in”. Quoin concludes that de re and de dicto ascriptions differ in that they make use of different belief predicates. De dicto ascriptions must be interpreted in terms of a binary predicate expressing a relation between a person and the content of the complement clause. In the case of de re ascriptions, on the other hand, one has polyadic
belief predicates expressing relations between the subject, certain other objects, and certain properties applying to these objects. For instance, the sentence "Ralph believes that the man with the brown hat is a spy", read \textit{de re} with respect to the noun phrase "the man with the brown hat", has the logical form "Believe³ (Ralph, the man with the brown hat, being a spy)" and thus expresses a ternary relation between Ralph, Orcutt, and the property of being a spy; the sentence "Ralph believes that Tom knows the man at the beach", read \textit{de re} with respect to "Tom" and "the man at the beach", has the logical form "Believe⁴ (Ralph, Tom, the man at the beach, knowing)" and expresses a four-place relation between Ralph, Tom, Orcutt, and the binary relation of knowing.

Obviously, such an analysis is only a first step towards an interpretation of \textit{de re} belief ascriptions; the second step must then be to explain how these polyadic belief relations obtain and when they do not obtain. Following Quin, Kaplan (1969) made a detailed proposal to this effect, according to which polyadic \textit{de re} predicates are reduced to the binary \textit{de dicto} predicate – roughly by supplying the missing characterisations of the objects from the subject’s point of view.

At any rate, the following moral may be drawn from Quin’s considerations: \textit{de re} belief ascriptions are not intended to be internal characterisations of belief contents; it is thus clear that, on a \textit{de re} reading, the belief predicate cannot be interpreted as a relation between the subject and the intension of the "that"-clause – not even if the terms in questions have been "quantified out". Quite analogous considerations hold for \textit{de dicto} readings of belief ascriptions.

Kripke (1979) gave a detailed account of this conclusion. His best known example is about Pierre, a Frenchman who only speaks French and has never left France (Kripke 1979, pp. 254f.). However, he has heard many of his co-patriots talk about the English town of London (which, of course, they call "Londres") and, due to their stories, he is inclined to seriously and honestly utter the sentence "Londres est jolie". In this situation we may apply the extended disquotational principle and conclude: "Pierre believes that London is pretty"; the \textit{de dicto} reading of this sentence would thus come out as true. Later Pierre moves to England, London even, but to a very ugly part of the town, which he practically never leaves. Since his neighbours do not speak a word of French, he is forced to pick up English the direct way. Among other things he learns that the town in which he lives is called "London". Due to is own impressions, Pierre is inclined to asssent to the sentence "London is not pretty", and again the disquotational principle implies: "Pierre believes that London is not pretty". At this time Pierre would, however, still assent to the French sentence "Londres est jolie"; he simply fails to realize that the ugly town in which he ended up is the place about which he has heard so many good things back in France.

In this story, the \textit{de dicto} readings of the belief ascriptions "Pierre believes that London is pretty" and "Pierre believes that London is not pretty" are intuitively true, while at the same time one would not want to describe Pierre as a person with internally contradictory beliefs. But if one interprets belief ascriptions such that the objective meanings of the complement clauses are contents of Pierre’s beliefs, this precisely comes down to ascribing a contradictory belief.
Lewis (1981) points out that Kripke’s contradiction argument, i.e., the second part of the story about Pierre, is not needed in order to prove the inadequacy of such an interpretation. More generally: in can be shown directly that the objective meaning of a statement containing proper names cannot be its subjective meaning. The following consideration reveals that the proposition expressed by "Londres est jolie" did not belong to the content of Pierre’s belief even at the time when Pierre still lived in France. Imagine a possible world like ours, except that in it the English have put a lot of money into redecorating the town of Bristol and have finally renamed Bristol into "Londer", in honour of Sir Øgdred Londer. The French call this town "Londres" and often praise their beauty. Pierre has also heard of it and is therefore inclined to assent to the sentence "Londres est jolie". At the same time the town of London has more and more become impoverished and meanwhile only consists of absolutely ugly slums. The French hardly ever speak of London, and it has never been mentioned in Pierre’s presence. Such a world is a member of Pierre’s belief set: according to whatever Pierre believes that world could be the world in which he lives; if one would lead him around that world and tell him this were the real world, he would be in no way surprised. But this world is not a member of the proposition expressed by "Londres est jolie"; for in it London is not pretty but quite ugly.

More generally, the objective meaning of a sentence containing a proper name – or any other rigid designator – could only belong to a person’s belief set if that person knew the essential properties of the object denoted by the proper name, i.e., those properties that determine that it is the same object, and not another one, in the various worlds. However, normally nobody ever has such identifying knowledge about anything. And it would be wrong to say that only someone who has such knowledge knows the meaning of a name or of the sentences in which the name occurs.

Putnam (1975) drew quite similar conclusions from examples involving natural kind terms. Putnam reasons that whatever competent users of language ‘have in their heads’, i.e. their narrow psychological states, does not determine the meanings, the intensions, of the expressions. The argument goes, so to speak, in the opposite direction: it is supposed to show that subjective meaning cannot fulfil the tasks of objective meaning – among which is the determination of extension. In this connection Putnam presents the following *gedankenexperiment*:

Imagine that somewhere in the Galaxy there is a planet we are calling "Twin Earth". This planet resembles our Earth in almost every respect, the only difference being that the liquid called "water" on Twin Earth does not have the chemical structure H₂O but another one to which we will refer as "XYZ". The superficial properties of the two liquids are completely identical, they can only be distinguished by chemical analysis. Now consider an inhabitant of Earth – we call him "Oscar" – of whom one could say, beyond any doubt, that he has a good control of English and, in particular, that he knows what the word "water" means. Oscar has quite a few correct beliefs about water – among other things that it is colourless, odourless, thirst-quenching, and that it fills lakes and rivers. But Oscar knows nothing about chemistry and, in particular, does not know that water
1.3 The Content of Belief

consists of H₂O molecules. Now, on Twin Earth there lives Oscar’s doppelganger, Twin Oscar, who resembles Oscar in every respect and, among other things, has the same beliefs about the colourless, tasteless, and odourless liquid that fills the lakes and rivers of Twin Earth and is called ”water” in Twin English. However, although the internal states of Oscar and Twin Oscar are identical, Oscar’s word ”water” still is not the same word as Twin Oscar’s ”water” and ”water” has different meanings in Twin English and in English: both in English and in Twin English ”water” is a natural kind term; and whether something lies in the extension of the English word ”water” does not depend on its surface properties but only on whether it has the same structure as that which on Earth is called ”water” in English. It would therefore not be correct to call the liquid filling Twin Earthian lakes and rivers ”water” in English; it is not water. The analogous point could be made for the Twin Earthian word ”water”.

Putnam’s conclusion is: Meanings are not in the head. According to Putnam, meanings have several components and they are subject to a linguistic division of labour: whatever the normal language user associates with a word – at least if that word is a natural kind term – are merely certain stereotypes that do not identify the extension and in extreme cases need not even apply to the extension. Identifying the extension, as well as structural features determining the extension in other possible worlds and thereby the intension of natural kind terms, is a task reserved to a group of experts – in the case of ”water”, the chemists. The intension of a natural kind term does therefore not depend on psychological facts about language users but on the way the world is that they inhabit.

Putnam is concerned with truth and with progress in scientific theorizing and he is thus chiefly interested in the objective semantics of words and the objective truth-conditions of statements, less so in a systematic theory of informational content, i.e., the subjective meaning of linguistic expressions. Still, Putnam has illuminated two points that are relevant to the development of such a systematic theory.

For one thing, Putnam says that natural kind terms possess a hidden indexicality: whether something on another planet or in another world is water depends on whether it is the same liquid as our water here. This idea points at the possibility to explain the subjective meaning of natural kind terms via their character. We will investigate this possibility in more detail in sections 3.1 and 3.2.

Moreover, it becomes clear why it is practically unavoidable that objective truth conditions and subjective informational content of statements containing natural kind terms fall apart: single persons – and sometimes even the language community as a whole – only very rarely have the knowledge necessary for determining the intension of a natural kind term. The situation is similar to the one we have described in connection with proper names: to know the intension of a natural kind term, one would have to know the essential properties of the kind or substance to which the term is applied.

The observation that competent speakers do not know the intension of an expression is not restricted to proper names and natural kind terms. With other expressions, too, it may happen that speakers whose mastery of the language is beyond dispute only have incomplete knowledge of meaning; in such cases we do not have incomplete knowledge
about objects in the world and their essential properties but incomplete knowledge of the facts involving the language community. In these situations subjective and objective meaning again fall apart. This can be made clear by an example from Burge (1979).

Burge’s story is about a person – let us call him Fred – who has quite a few beliefs pertaining to the disease of arthritis. Thus, e.g., Fred believes that he has suffered from arthritis for many a year, that it is better to have arthritis than cancer of the liver, and a lot more. Moreover he believes to have arthritis in his thigh; at least, he seriously and honestly utters the sentence ”I have arthritis in my thigh.” What Fred believes is wrong and the proposition objectively expressed by his utterance is even necessarily false; for it is part of the meaning of ”arthritis” that one can only have arthritis in the joints and there is thus no possible world in which Fred had arthritis in his thigh.

Now imagine a counterfactual situation that only differs from the actual one in that the language community Fred lives in uses the term ”arthritis” to cover all kinds of rheumatic pains and not, as we have it, just those occurring in the joints. This difference is, however, not supposed to reach into Fred’s personal sphere; Fred himself, including his history of learning and suffering and his total environment (inasmuch he has experienced it), are completely identical. Again Fred says: ”I have arthritis in my thigh.” But if we now want to describe his belief, we must not say that he believes to have arthritis in his thigh; we would instead have to take a different word – ”tartthritis”, say – to serve as the translation of the word ”arthritis” in the counterfactual language, and then go on to say that Fred believes to have tartthritis in his thigh.

One might object that the linguistic competence of actual Fred is so incomplete that his utterances about arthritis do not entitle us to ascribe beliefs featuring arthritis to him. But Burge, using ever new examples, makes clear that our linguistic competence, though possibly rarely that faulty, is almost always more or less incomplete; if one were to follow that argument and only applied the disquotational principle to utterances supported by a full mastery of the language, our common practice of belief attribution would completely collapse.

A different consequence seems to be inevitable instead: even de dicto belief ascriptions do not provide internal, but only external characterisations of persons’ belief contents. For it is prima facie clear that the cases described are de dicto belief ascriptions; after all, they are immediate results of applying the disquotational principle. This can also be seen from the fact that the expressions ”arthritis” and ”tartthritis” must not be replaced by co-extensional terms. Furthermore, it is clear that different de dicto belief ascriptions with different ”that”-clauses apply to actual and to counterfactual Fred: among other things, one ”that”-clause differs from the other in that it contains a word whose extension and

19 The fact that this utterance contains the word ”I” and is therefore indexical can safely be ignored in what follows.

20 I am assuming that Bill knows that there are no joints in his thigh.
intension differ from the word occupying the same position in the other sentence. Finally, it is clear that the differences in these *de dicto* ascriptions are not the result of any internal differences between actual and counterfactual Fred, but only of the external differences between their respective linguistic communities. In short: Even though counterfactual Fred does not differ from actual Fred in any internal respect, we must not use the same *de dicto* ascription for him; thus *de dicto* belief ascriptions cannot be internal characterisations.

Burge, however, takes his stories not only to imply that attitude ascriptions are external; he even claims to have shown that the belief contents themselves cannot be individuated internally – or, as he puts it: individualistically. The assumption he makes is that the content of a complement clause of a *de dicto* belief ascription is also the content of the belief described; complement clauses with different contents must then also describe different beliefs, so that Burge’s anti-individualistic conclusion is obvious: beliefs, like meanings, are not in the head.

I will not quite follow this train of thoughts, though. On the contrary, I would like to stick to the view that, below the level of *de dicto* belief ascriptions, there is an even more fundamental level – the level of internally characterised epistemic states, as introduced at the beginning of this section – and hence a level of individualistic individuation of these states. This assumption is at least *prima facie* plausible. It also explains why Burge’s story – like Putnam’s – initially creates the impression of dealing with a paradoxical situation. On the one hand, there is a clear intuition that Fred and his counterfactual counterpart – or Oscar and his twin – have the same beliefs; but then one cannot ignore the arguments to the effect that their beliefs cannot be the same. However, the contradiction dissolves once one allows for different levels of description and of individuating beliefs.

From Burge’s reflections we thus obtain further confirmation, and even aggravation, of the problems described so far. The confirmation lies in the observation that objective and subjective meaning may fall apart not only in natural kind terms but in many other predicates; the proposition expressed by the sentence ”I have arthritis in my thigh” as uttered by a person from our language community is not a superset of the belief set of that person, it is not believed by him – in the internal sense used so far – and the objective content of the complement cause of a *de dicto* ascription is not the content of the belief described.

The aggravation lies in the insight that *de dicto* belief ascriptions are not individualistic. For this implies that the three roles that I have assigned to subjective meaning at the beginning of this section do not completely fit together. According to the first two roles, both beliefs and subjective meanings were supposed to be characterizable internally. But the above account of the third role cannot be entirely correct; a *de dicto* belief ascription cannot just express a relation between the subject and the internally characterised subjective meaning of the complement clause. Though the semantics of belief ascriptions is not the official subject of the present work, I will briefly return to the analysis of *de dicto* belief ascriptions in section 3.9.
1.4 Context Change and Propositional Concept

From several perspectives now reasons have appeared to not equate the objective meaning of a statement, its truth condition, with the subjective meaning that it has for the person who makes or hears it. The concept of subjective meaning has so far only been characterised in very general terms; summing up, we may say that the problems and examples discussed lead to the following desiderata for a theory of subjective meaning:

A theory of subjective meaning should solve the problem of informativity. That is, it should explain why two statements with the same truth conditions can have different cognitive values, why necessarily true statements can be informative, and how it is possible that necessarily false statements are believed by rational persons.

Furthermore, a theory of subjective meaning should solve the problem of internalism, as it has been posed by Kripke’s reflections on the one hand, and Putnam’s and Burge’s on the other. That is, it should describe belief contents and thus people’s subjective meanings in an internal way and such that Pierre from Kripke’s story is ascribed a consistent belief content – for Pierre is not subject to epistemic irrationality – and that the several doppelgangers in Putnam’s and Burge’s stories are, respectively, ascribed the same belief contents and their utterances the same subjective meanings – for their internal identity has after all been part of the stories.

Finally, a theory of subjective meaning should be systematic. It therefore must not treat these problems in ad hoc way, making special assumptions. On the contrary, it must generate the subjective meanings of all complex expressions in a recursive way; otherwise it could not be explained how ever new expressions can have subjective meanings for speakers and hearers.

Stalnaker’s writings are an important starting point for the theory of subjective meaning which, in view of these three desiderata, I am going to develop. An account of at least the essence of his approach is therefore in order.21

Within the theory of meaning, Stalnaker distinguishes a semantic and a pragmatic component. Semantics determines which proposition a sentence expresses in a context and how the proposition expressed depends on features of the context in which it is uttered. It does so by providing a recursive definition of the characters of all expressions of the language under study – where utterances are regarded as abstract constructions, viz. as pairs of a sentence and a context. Pragmatics, on the other hand, takes utterances as

21 The followong account of Stalnaker’s theory is mostly based on his ”Assertion” (1978) and his ”Semantics for Belief” (1987b). The basic ideas of his position can, however, already be found in several papers from the early seventies: ”Pragmatics” (1972a), ”Propositions” (1972b), ”Presuppositions” (1973), ”Pragmatic Presuppositions” (1974), ”Indicative Conditionals” (1976a), and ”Possible Worlds” (1976b).
concrete events or speech acts and describes their meaning in terms of a theory of action. Both the production and the interpretation of utterances are instances of rational behaviour and it is therefore part of the task of pragmatics to explain why and to what end persons make utterances and which strategies they apply in the interpretation of utterances.

In this pragmatic sense, utterances are always made and interpreted against a background of previously held beliefs. In rational and co-operative conversations, all participants try to adjust their background assumptions. Whoever speaks then attempts to design his contributions in such a way that their interpretation only requires assumptions he takes to be commonly known and accepted. Background assumptions in that sense are what Stalnaker calls speaker presuppositions. The totality of the presuppositions of a speaker is represented by a set of possible worlds, the set of worlds that are possible according to the speaker presuppositions. This set is also called the context set or, as we shall say, following Kratz (1978): the speaker’s conversational background (Redehintergrund). Ideally, the context sets of all interlocutors coincide, so that one may simply speak of the context set or the conversational background of an utterance. The conversational background of an utterance \( < \phi, c > \) is written as \( R_c \); so \( R_c \) is a further feature of any (ideal) utterance context \( c \).

Stalnaker mainly considers assertive utterances – a restriction to which the present work is also subject. An assertive utterance is characterised by its success which at least consists in the claim’s being accepted by the other participants in the conversation and its thereby becoming part of the common conversational background – which can be formally modelled easily by forming the set-theoretic intersection of the context set and the proposition the utterance expresses for the participants in the conversation. The point of an assertion is thus to reduce the context set: if the sentence uttered is accepted, only such worlds remain in the context set after its utterance which are compatible with the content of the utterance. This corresponds to the idea mentioned on p. VVV 12 VVV that informativity can be explained in terms of eliminating possibilities; and, of course, assertions should be informative.

We can now be somewhat more specific about what pertains to the pragmatic component of the theory of meaning, namely a general theory of change of context sets describing how knowledge of meaning as formulated in the semantic component is applied in forming a new context set when interpreting an utterance made on the background of a given one. Pragmatics should thus specify the context change potentials of sentences: for any context set \( R_c \) and any sentence \( \phi \) uttered in \( R_c \), it is to say what

---

22 Kaplan (1977, section XIII, 1979, p. 91, 1989, p. 584) also points out the difference between utterances as sentences-in-contexts and utterances in a pragmatic sense, emphasizing that his lexical semantics is only concerned with the former.
The concept of a context change potential I have borrowed from Heim (1982, 1983); in this form it does not appear in Stalnaker’s writings. Maybe it is even slightly at odds with his understanding of the relation between semantics and pragmatics, according to which pragmatics is just not concerned with the meaning of sentences as abstract types. But then a systematic pragmatic theory can hardly do without reference to abstract sentences; Heim’s concept therefore appears appropriate to me.

To begin with, each utterance of a sentence leads to quite a trivial change of the old conversational background: that this specific utterance has been made and, normally, that it has been made by a specific person at a specific time and place is added to the common assumptions. These new assumptions may affect the interpretation of the sentence uttered. If, e.g., Jack utters the sentence “I am hungry now” to his friends at noon and if everybody knows who is speaking and what time it is, the conversational background is supplemented by the information that Jack has uttered the sentence “I am hungry now” at noon; due to their semantic knowledge about “I” and “now”, the hearers can further conclude that on this conversational background that sentence means the same as the sentence "Jack is hungry at noon"; so in the end that Jack is hungry at noon is added to the conversational background.

This example also makes clear that it is not the actual utterance context, but the participants’ assumptions about this utterance context appearing in the conversational background, that determine the interpretation of indexical expressions and hence what the utterance means to them. For if all participants are wrong about the time, thinking it is not noon but only 11 a.m., they would interpret Jack’s utterance as "Jack is hungry at 11 a.m." and it would be the content of that sentence that gets added to the conversational background. Or, to quote an example from Daniels (1978): If Stalnaker says to O’Leary, who is standing next to Daniels: "You are a fool" and if O’Leary takes himself to be addressed by Stalnaker, while Daniels believes Stalnaker means him, then Stalnaker’s utterance has the same meaning for O’Leary – and for Stalnaker himself – as the sentence "O’Leary is a fool", whereas for Daniels it has the same content as "Daniels is a fool".

According to Stalnaker, it does not always have to be \( \| \phi \| (c) \), the proposition \( \phi \) objectively expresses in context \( c \), which is added to \( R_c \) – otherwise one could simply define \( R_c + \phi \) to be \( R_c \cap \| \phi \| (c) \), so that determining the context change potential and hence the pragmatics of assertions would be a trivial matter; rather, it is a proposition that essentially depends on \( R_c \). This explains – and it will still become clearer as we go along – why Stalnaker also has a subjective theory of meaning in mind when he talks about utterance meaning.

So far, the examples suggest a purely incremental model, according to which the new conversational background is obtained by simply adding the proposition the sentence uttered expresses as – subjectively in a sense – depending on the old conversational background. But Stalnaker’s views on the change of conversational backgrounds are more complex than that. An utterance can have an influence on the conversational

\[ R_c + \phi \] – the new context set resulting from interpreting and accepting the utterance of \( \phi \) in \( R_c \) – looks like.\(^{23}\)

\(^{23}\) The concept of a context change potential I have borrowed from Heim (1982, 1983); in this form it does not appear in Stalnaker’s writings. Maybe it is even slightly at odds with his understanding of the relation between semantics and pragmatics, according to which pragmatics is just not concerned with the meaning of sentences as abstract types. But then a systematic pragmatic theory can hardly do without reference to abstract sentences; Heim’s concept therefore appears appropriate to me.
background in yet other, more indirect ways; moreover, the conversational background can indirectly have an influence on which proposition an utterance expresses. In order to be able to give a more precise description of such kinds of influence, Stalnaker formulates three principles concerning the relation between context set and proposition expressed. Only if these principles are met can the assertion reach its goal of reducing the context set; this is what justifies the principles. Here they are:

1. The proposition expressed must always be true in some but not all worlds of the context set. This principle simply says that one shall not say anything that contradicts the current background assumptions, nor anything that has already been accepted as true; in the first case the context set would obviously not be reduced, but eliminated altogether, and in the second case one would not achieve any reduction whatsoever of the context set.

2. The utterance shall express a proposition with respect to every world in the context set, and this proposition shall have a truth-value in every world of the context set. The task of this condition is mainly to exclude utterances of sentences whose semantic presuppositions are not part of the general background assumptions, i.e. not true in each world of the conversational background. For – according to a common treatment of presuppositions – such utterances would be neither true nor false in some worlds of the conversational background and it would then be unclear whether or not these worlds are to survive in the new, reduced conversational background.

3. The same proposition shall be expressed with respect to each world of the context set. This principle concerns context-dependent expressions and demands their reference to be determined so far as to always denote the same object, no matter in which world of the context set they are uttered; only then the sentence in which they occur expresses the same proposition in each world of the context set. For if the propositions expressed varied with different utterance worlds, it would again be unclear how to reduce the context set, i.e., which proposition to add to the old conversational background.

To get a better grasp of these principles, one must realize that Stalnaker’s context set worlds play the role of the context as well as that of the index. That by uttering a sentence the assumption that this sentence has just been uttered gets into the conversational background just means that thereby the context set only contains worlds in which the utterance has taken place; so the context set consists of the utterance worlds that are possible for the hearer. At the same time, however, this set forms the possibility space that is to be further restricted by the utterance, it is "the material out of which propositions are constructed" (Stalnaker 1987a, p. 146). So every sentence \( \phi \) uttered on a conversational background

---

24 There is no contradiction here to our earlier considerations made in section 1.2 which showed that the set of indices must not be identical to the set of contexts. The point was that the interpretation of clauses embedded under intensional operators, and particularly under a necessity operator, must have access to indices that are not utterance contexts. However, outside intensional constructions, index and context collapse, as can be seen from the definition of the notion of a true utterance. Accordingly, in Stalnaker’s view any utterance of the sentence "I am here now" is true in all worlds of the context set; but this only means that the utterance is true, and not that the proposition expressed is necessary. Of course, necessary truth must not be defined as truth in all worlds of the context set; rather, intensional operators lead to the
$R_c$ can be assigned a function from $R_c$ to the power set of $R_c$, a function assigning to every (utterance) world $w$ in $R_c$ the set of those worlds $w'$ from $R_c$ in which $<\phi, w>$, i.e. $\phi$ uttered in $w$, is true. Stalnaker calls such a function the propositional concept determined by $\phi$ in $R_c$. The above principles can then be understood as conditions on propositional concepts: principle 2 excludes that the propositional concept of an utterance of $\phi$ in $R_c$ is only partially defined, principle 3 demands that it be a constant function, and principle 1 says that the value of that function is neither $R_c$ nor the empty set.

It is now tempting to simply regard the propositional concept of an utterance of $\phi$ in $R_c$ as the restriction of $\phi$ ’s character to $R_c$. Stalnaker’s early writings, and particularly Stalnaker (1978, pp. 318-320) do indeed suggest this interpretation; but in Stalnaker (1987b, pp. 182-183) such an identification is explicitly denied: character is supposed to be a semantic concept specifying the meaning of sentences and other expressions; propositional concept, on the other hand, is a notion from pragmatics that has nothing to do with the meaning of sentences but only with the content of utterances. The propositional concept of an utterance is not even a function of the meaning of the sentence uttered, because in order to determine it, one would not only have to know the meaning of the sentence in the actual world but also its meaning in the other worlds of the conversational background.

It appears that two reasons have motivated Stalnaker’s choice for not equating his propositional concepts with characters. The first has to do with his view that propositions as objects of attitudes must be constructed as sets of possible worlds and not as sets of more complex indices; this thesis is argued for at length in Stalnaker (1981). However, if one were to understand propositional concepts as partial characters, one would have to presuppose that the elements of the context set are entities that are at least as finely individuated as contexts – for character functions cannot just be defined for possible worlds, their arguments must be entire utterance contexts if they are to describe the semantics of indexical expressions like “I”, “here”, “now”, etc. In the next chapter we will, however, see that independent reasons speak against the view of attitude objects as sets of worlds. Thus Stalnaker’s first reason for rejecting the identification of propositional attitudes and characters loses its force.

His second reason presumably lies in the treatment of proper names. For if Stalnaker says that for determining the propositional concept of an utterance one needs to consider its meaning in all worlds of the conversational background, not just in the actual utterance world, he is particularly concerned with sentences containing proper names. Like Kripke (1972), Stalnaker interprets proper names as rigid designators; and he also seems to follow Kaplan (1977, section XXII, and 1989, pp. 598-599) in his view that proper names are not indexical and thus not only have a constant intension but also a constant character. Nevertheless Stalnaker wants to allow for a proper name to denote different objects in different (utterance) worlds and thus have different intensions; as we will see in the creation of so-called derived contexts that no longer need to have the property having hosted an utterance. See Stalnaker (1987a) for the notion of a derived context.
examples to come, he thereby wants to account for the fact that a person does not know what the name refers to.

The first example discussed by Stalnaker in (1978), p.330, is about the name "Sherlock Holmes". One has to imagine a person who is unclear about the status of the Sherlock Holmes stories; this person does not know whether they are fiction or factual reports. More precisely, the following worlds are possible according to the information that person has, they are in his context set:

- the actual world $w^0$, in which Sir Arthur Conan Doyle created a fictional character called "Sherlock Holmes" who does not correspond to any real person;
- a world $w^1$, in which there lived a famous detective named "Sherlock Holmes" in 19th century London, and where Sir Arthur Conan Doyle, not a detective himself, wrote a series of biographic stories about some of his cases;
- a world $w^2$, in which Sir Arthur Conan Doyle was himself a lousy detective named "Sherlock Holmes" and wrote self-glorifying accounts of his own cases using the pseudonym "Sir Arthur Conan Doyle".

Now let "$a$" designate Arthur Conan Doyle and "$b$" the famous detective from world $w^1$, so that $a$ exists in all three worlds $w^0,w^1$ and $w^2$, but $b$ only in $w^1$; we may then say that, depending on whether it takes place in $w^0$, $w^1$ or $w^2$, an utterance of "Sherlock Holmes" refers to nothing, to $b$ or to $a$, respectively.

Accordingly, sentences containing the name "Sherlock Holmes" express different propositions if uttered in worlds $w^0$, $w^1$, and $w^2$. For example, the sentence "Sherlock Holmes was a sharp detective", as uttered in $w^0$, expresses no proposition at all\(^{25}\); if uttered in $w^1$, it expresses the proposition that $b$ is a sharp detective, and on an utterance in $w^2$ it expresses that $a$ is. Hence, on a conversational background consisting only of $w^0$, $w^1$, and $w^2$, the utterance "Sherlock Holmes was a sharp detective" determines the following propositional concept $\Pi_1$:

<table>
<thead>
<tr>
<th>$\Pi_1$</th>
<th>$w^0_c$</th>
<th>$w^1_c$</th>
<th>$w^2_c$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$w^0$</td>
<td>$\varnothing$</td>
<td>$\varnothing$</td>
<td>$\varnothing$</td>
</tr>
<tr>
<td>$w^1$</td>
<td>$\varnothing$</td>
<td>$T$</td>
<td>$\varnothing$</td>
</tr>
<tr>
<td>$w^2$</td>
<td>$T$</td>
<td>$F$</td>
<td>$F$</td>
</tr>
</tbody>
</table>

The above representation shows which truth-value the sentence has as uttered in the line world and evaluated in the row world; $T$ stands for true, $F$ for false and $\varnothing$ for truth-valueless. The truth-value distribution presupposes, in line with Stalnaker, that a sentence containing a name that does not denote anything in its utterance context, does not get a truth-value in any (index) world and, moreover, that a sentence with a name denoting an

\(^{25}\) We only consider the literal meaning of the sentence we take to presuppose the existence of a referent for the name "Sherlock Holmes", and we ignore the reading that could be rendered by a paraphrase like "In the pertinent stories is it the case that Sherlock Holmes is a sharp detective".
object that does not exist in all words, only gets a truth-value in those worlds in which this object does exist.

In Stalnaker (1987b, pp. 183f.) we find a further example for propositional concepts at work, one that involves the name "Hesperus": There are certainly possible worlds in which the firmament appears to the layman as it does in the actual world, in which the very same planets revolve round the sun and carry the same names as they do here, in which, in particular, Venus is called "Venus" and Mars is called "Mars" – but in which, due to certain primeval perturbations, they are differently arranged around the sun; so in the evening planet Mars appears there at exactly the position at which Venus appears here and it even looks just like Venus looks here – at least to the lay eye. In these worlds, too, ancient astronomers have given names to the planet that is seen first in the evening sky; and from this name the name "Hesperus" derives, which the 20th century English speakers there use to refer to that planet.

Now let \( w^0 \) be the actual world again and \( w^1 \) one of the counterfactual worlds just described. According to Kripke's semantics of names, an utterance of "Hesperus is Mars" expresses the necessarily false proposition. But it is just the other way round if the utterances are not made in \( w^0 \), but in \( w^1 \); for there "Hesperus" is a rigid designator for Mars, not Venus. So on a conversational background consisting only of \( w^0 \) and \( w^1 \), the utterances "Hesperus is Venus" and "Hesperus is Mars" are assigned the following non-constant propositional concepts \( \Pi_2 \) and \( \Pi_3 \):

\[
\begin{array}{c|cc}
\Pi_2 & w^0_i & w^1_i \\
\hline
w^0_c & T & T \\
w^1_c & F & F \\
\end{array}
\quad
\begin{array}{c|cc}
\Pi_3 & w^0_i & w^1_i \\
\hline
w^0_c & F & F \\
w^1_c & T & T \\
\end{array}
\]

It is quite conceivable that a conversational background contains both \( w^0 \) and \( w^1 \). Many persons only have very amateurish knowledge of stars, not quite remembering what they have been told about Hesperus, whether it is Venus or Mars. According to these persons' beliefs, both \( w^0 \) and \( w^1 \) are possible; the corresponding context set contains both worlds in which Venus is called "Hesperus", and worlds in which, as in \( w^1 \), Mars is called "Hesperus".

The point of these two examples thus lies in the possibility that persons can be unclear about the referent of a name – and hence about the proposition an utterance with that name expresses. In these cases the proposition expressed varies with the worlds in the conversational background. If one were to account for this dependence by the character of sentences and the names they contain, one would obviously have to regard the character of a name \( \alpha \) as that function which assigns to a world \( w \) and an index \( i \) the object which is called \( \alpha \) in \( w \). Stalnaker apparently takes this to be wrong – and apparently for similar reasons as Kaplan does. For one would then not explain the meaning of a name, but only the various possibilities of using a string of sounds. But names are not mere strings of
sounds; they are symbols belonging to a specific language, thereby having a fixed meaning described by the character function of that language. Indexical expressions do have different referents in different contexts, due to the meaning they have in a language; one and the same string of sounds can have different referents when used as a name in different languages; but the semantics of a particular name does not allow for any variation. What causes the variability of the propositional concepts in these examples is not the semantics of "Sherlock Holmes" or "Hesperus" in English, but the insufficient knowledge the relevant persons have of that semantics; to that extent, they do not know English. This, then, is the second reason why Stalnaker considers propositional concepts and characters utterly different.

It is only in the fourth chapter when concerned with the character of names, that we will extensively discuss these observations and examples. The picture will become more involved there because we will find that, over and above the variability of propositional concepts due to semantic ignorance, and in contradiction to Kaplan and Stalnaker, names are still indexical, but in a Putnam-like, hidden way, and not like the paradigmatic indicators "I", "here", etc. For the moment we are happy with the insight that Stalnaker’s propositional concepts of utterances are indeed distinct from Kaplan’s characters of sentences.

After these comments we can return to Stalnaker’s theory of changing conversational backgrounds. It is only when its propositional concept satisfies the three above conditions, that the reduction of a context set by an utterance is a straightforward matter. The utterance then expresses one and the same proposition in each world of the context set, and this proposition is added to the old conversational background. Because principle 1 is satisfied, the result is always a reduction of the worlds in the context set.

The propositional concepts we have just seen in the examples, however, violate the second and third condition; otherwise we could not have used them to illustrate the variability of propositional concepts. But this does not mean that an utterance of "Sherlock Holmes was a sharp detective" or of "Hesperus is Mars" would be illicit or incomprehensible on the respective conversational background. It means that the hearers must resort to additional strategies of interpretation to attune the conversational background and the content of the utterance.

In general, there are two ways of reacting to violations of the three conditions. The first one consists in tacitly changing the conversational background so as to meet the conditions; this procedure is called accommodation.26 The second strategy is one of re-interpretation: the hearer gives up on the original interpretation of the utterance that violates the conditions and instead tries to find an interpretation that is in accord with the conditions.

Accommodation typically applies in connection with presuppositions. If an utterance is made and its presuppositions are not contained in the conversational background, a co-

---

26 The term is due to Lewis (1979a).
operative hearer does not react by breaking off the conversation; rather, he concludes that his context set does not match the speaker’s one, which obviously does meet the presuppositions, and thus tacitly adds the missing presuppositions to his conversational background. Thereby the second conditions on propositional concepts is satisfied again, and the utterance can now be interpreted without any problems. Due to such accommodation, utterances can even indirectly change the conversational background and thus implicitly convey information. This view proved to be particularly fruitful in presupposition theory (cf. Heim 1983). However, we will not go into this here, because presuppositions and pragmatic inferences in general are not our topic.

Accommodating the conversational background does, however, not always suffice; this can be gleaned from the above examples. Our first case, the utterance of ”Sherlock Holmes was a sharp detective” in the context set \( \{ w^0, w^1, w^2 \} \), produced the propositional concept \( \Pi_1 \), which does not only violate both parts of the second condition but also the third one. To begin with, accommodation is needed in this case: the sentence uttered has the semantic presupposition that there exists a referent of the name ”Sherlock Holmes”; so this assumption must be added to the conversational background, which means that the world \( w^0 \) where there is no Sherlock Holmes must be eliminated from the context set.

Given \( \Pi_1 \), this accommodation generates the propositional concept \( \Pi_1^* \):

\[
\begin{array}{c|cc}
\Pi_1^* & w^1_i & w^2_i \\
\hline
w^1_c & T & \varnothing \\
w^2_c & F & F \\
\end{array}
\]

However, \( \Pi_1^* \) still violates the second part of condition 2 plus – just like the propositional concepts \( \Pi_2 \) and \( \Pi_3 \) from our second example – condition 3; in all three cases, the propositional concept is not constant and it is therefore unclear which proposition is to be added to the old conversational background.

It is now that Stalnaker’s re-interpretation procedure comes to bear, which will turn out to be crucial for us. For in such a situation, says Stalnaker, co-operative hearers no longer assign to an utterance its regular propositional concept but rather the latter’s diagonalization; that is, they apply the diagonal operator \( \delta \) to that propositional concept.\(^{27}\) As already explained on VVVp.30VVV, the diagonal operator turns context dependences into index dependences – which is to say, in Stalnaker’s framework: context world dependences into index world dependences –, and it always generates constant propositional concepts. Also, diagonalization is always well-defined within a context set, because the index worlds coincide with the context worlds. Applying diagonalization to \( \Pi_1^\ast, \Pi_2 \) and \( \Pi_3 \), one thus obtains the propositional concepts \( \delta(\Pi_1^\ast), \delta(\Pi_2) \) and \( \delta(\Pi_3) \), which are finally in accord with all three conditions:

\(^{27}\) Since we are concerned with propositional concepts that are defined on \( R_c \times R_c \) and not with characters defined on \( C \times I \), the explanations from pp. VVV29f.VVV must be adapted accordingly.
After diagonalization a unique proposition remains to be added to the conversational background: the one which the utterance expresses in an arbitrary world of the context set – and this is the same proposition for any of these worlds. Clearly, as was already explained on p. 30, this proposition is just the (left to right) diagonal of the original, undiagonalized propositional concept (granting that in Stalnaker’s framework it is utterances, not sentences, that have diagonals).

Of course, diagonalization is not just a technical trick of Stalnaker’s, applied only to guarantee that his three conditions are satisfied somehow; rather, it is intuitively well motivated. If, for a given hearer, the utterance of a sentence does not determine a constant propositional content in a context set, this plainly means that for the hearer the reference of the expressions in depends on the context worlds in . As one may, moreover, assume that the hearer does not know which of the context worlds in is the actual one, he therefore does not know which proposition is expressed by . On the other hand, the propositional concept expresses the hearer’s conditional knowledge, which can be described as follows – if we briefly ignore the difference between propositional concepts and partial characters: The hearer knows that would express the proposition , if were the context of utterance, that it would express the proposition , if were the context of utterance, and so forth, for all contexts from . Moreover, of course, he still takes it that – whatever the actual context may be – the utterance of is true. Taking both into account, after the utterance of only those contexts remain possible for which , i.e., the contexts in which is true. The proposition that the utterance of expresses for him, then, is just the set of all these contexts, i.e. the diagonal of this utterance. This is what Stalnaker proposes with his method of diagonalization in all cases that are marked by such incomplete knowledge.

The intuitive adequacy of this proposal is best seen by considering examples involving indexicals. The following example is taken from Perry (1988, p. 9): Perry receives a postcard with a picture of a palm-rimmed beach. On the back there is the sentence "I am having a great time here these days", but the postmark, the signature and the sender’s address are blurred and illegible. So in this situation Perry neither knows what "I" nor what "here" or "these days" refer to. Maybe he has a certain conjecture about who
might have written the card, because not all of his acquaintances like seaside holidays and not all of those that do would send him a post-card. In that sense, his conversational background does not contain arbitrary situations in which this sentence is uttered but only ones that are compatible with his conjectures. But his conversational background still contains different worlds in which the indexical expressions "I", "here", and "these days" have different referents; so Perry does not know which proposition the sentence on the post-card expresses. But he does know that whichever of the possible acquaintances has sent him this post-card from whichever beach at whatever time, was having a good time there and then. So Perry adds to his conversational background just the proposition that the author of the post-card is enjoying himself at the time and place of writing, i.e., the set of all contexts in which "I am having a great time here these days" is true – and, surely, this set is nothing but the diagonal of the utterance. One can see clearly that the indexical expressions "I", "here", and "these days" are not taken literally, but in their descriptive senses; it has already been mentioned on VVV pp. 29f.VVV that this is just what the diagonal operator does.

Let us end our account of Stalnaker’s theory by returning to the identity statements "Hesperus is Venus" and "Hesperus is Mars". In the situation described above either of these statements would augment the old conversational background with its diagonal; and, in accordance with Stalnaker’s principle 1, this diagonal would be an informative proposition. This, then, is Stalnaker’s solution to Frege’s problem of informativity: Identity statements among rigid designators do express a necessary proposition in each context world; but they may possess a variable propositional concept on a given conversational background, in which case they always have an informative diagonal. And it is this diagonal that forms the cognitive value of the statement, its subjective meaning.

It must be admitted that Stalnaker’s writings on utterance meaning are of a rather programmatic nature. He does, of course, motivate his programme by several examples and problems, but there is no detailed and systematic body of rules which would help account for utterance meanings in an orderly fashion. In particular, the crucial notion of a propositional concept is only introduced by way of examples, without any indication as to how the propositional concept of an utterance could be determined from its conversational background and the meaning of the sentence uttered; as we have seen, Stalnaker even denies its dependence on the meaning of the sentence uttered.

Stalnaker’s ideas have been adopted at various places in the literature. Von Stechow (1982, 1984) gives a critical discussion of the method of diagonalization and its usefulness in an account of belief contents, and particularly of the contents of beliefs de se and de nunc. In the next chapter, I am going to treat these matters in detail too.

Gazdar (1979, ch. 6) presents an attempt at a precise definition of the notion of context change. However, Gazdar is exclusively concerned with the presuppositions28 and

---

28 The role of presuppositions in Gazdar’s theory is solely to contribute to the overall informative effect of the utterance, so that, in a sense, they are always accommodated and never act as pre-conditions on
implicatures of utterances; to this end, he develops a mechanism specifying the new conversational background $R_c + \phi$ for any given sentence $\phi$ and old background $R_c$. But the question of which proposition is expressed by a sentence $\phi$ on a conversational background $R_c$ is not considered; it is, rather, taken for granted that the answer to this question can be given in a semantics that is independent of the theory of context change. But then, according to Gazdar’s theory, the utterance meaning does depend on the conversational background in that it has an effect on which presuppositions and implicatures an utterance has.

Heim (1982, 1983, 1985) has systematically further developed Stalnaker’s ideas. In these works, too, presuppositions are a central topic which, however, is treated within the framework of a new theory of definite and indefinite noun phrases and anaphoric relations. Heim abolishes Stalnaker’s (and Gazdar’s) separation between a semantic and a pragmatic component. This way the context change potential becomes a basic concept in the theory of meaning; it is defined recursively for all atomic and complex sentences, the notion of truth being derived from it. I will not go into Gazdar’s and Heim’s theories here, though.

Rather, and as I have already announced, I am taking Stalnaker’s approach as the starting point of my efforts of arriving at a systematic theory of subjective meaning, and one that is able to solve the problem of informativity as well as the problem of internalism. At the beginning of the next chapter I will describe the strategy I am going to take and at the same time explain in which respects my approach differs from Stalnaker’s conception.
Chapter 2
On the Formal Structure of Belief Contents

2.1 Taking Stock

There are, above all, two ideas of Stalnaker’s that strike me as important and helpful in finding a solution to the problems described in sections 1.1 and 1.3 above: First, Stalnaker highlights the fact that utterances may not only convey information about the index but also about the context, and that therefore propositions, too, can be regarded as sets of contexts; this is the idea behind the method of diagonalization. Second, Stalnaker makes clear that subjective meanings must always be relativized to given background assumptions; this finally leads him to a holistic notion of utterance meaning.

According to Stalnaker, diagonalization is a strategy of interpretation which is only applied in special cases – viz. when knowledge about the meaning of an expression is incomplete, i.e. when the content of the expression is not constant across all worlds in the context set. However, the reflections in section 1.3 strongly suggest that this case is the rule rather than the exception. The examples discussed there showed that an incomplete knowledge of the intensions of expressions is common even among competent language users – and in the case of rigid designators or hidden indexicals it may well be inevitable. Of course, a more thorough reasoning is still needed that the different cases of incomplete knowledge can indeed be modeled via variable propositional concepts and thus construed as incomplete knowledge of the context of utterance. Such a reasoning would then establish that diagonalization is a ubiquitous strategy of interpretation. And this again would speak in favour of quite generally identifying the subjective meaning of an utterance with its diagonal, i.e., with the set of contexts in which the sentence uttered is true along with the given background assumptions. It is this idea that the following starts out from. Several differences with Stalnaker’s approach will, however, emerge.

The first basic point is that I will not adopt Stalnaker’s theoretical separation between propositional concepts as objects of pragmatics and characters as objects of semantics. I would rather like to show that it makes sense to develop a more general notion of character that is not restricted to semantic rules for indexical expressions but also covers the hidden context dependence of proper names and natural kind terms and, after yet another step of abstraction, even the cases of incomplete linguistic knowledge that Stalnaker discusses in connection with his propositional concepts. This generalized notion of a character does not square with Kaplan’s view on characters; I will, however, try to argue that it adequately reflects the specifically semantic knowledge of competent speakers and hearers.
Given this generalized character, Stalnaker’s propositional concept of an utterance \(<\phi, c >\) can be explained in terms of it: it is simply the restriction of the character of \(\phi\) to \(R_c \times R_c\). As a consequence, and in distinction to Stalnaker, the propositional concepts of utterances would be determined in a systematic way. With the generalized notion of a character, the diagonal \(\Phi(\phi)\)– the set of all contexts in which \(\phi\) is true – would be defined for any sentence \(\phi\). The diagonal of an utterance \(<\phi, c >\) in Stalnaker’s sense could then be obtained as the set-theoretic intersection of \(\phi\)’s diagonal with \(c\)’s conversational background.

However, the diagonal of an utterance in that sense still does not yield a suitable reconstruction of the notion of subjective meaning I am after. The reason is that Stalnaker’s objective does not entirely coincide with mine. Stalnaker wants to formulate a theory of the conversational contribution of utterances and thereby primarily takes the hearer’s perspective as his starting point. As we have already seen, his conversational contexts do not represent the actual beliefs of a particular participant in the conversation, but rather those assumptions which all conversational participants may rightly take to be common. In that sense, diagonals relativized to conversational contexts do not necessarily correspond to what speakers or hearers actually associate with the utterances, but only what they may rightly associate with them given their common assumptions. In distinction to this, I would like to develop a concept of subjective meaning that reflects the actual beliefs of speakers and hearers in producing and interpreting utterances.

A first consequence from this difference in objectives is that I cannot study utterances against the background of a common conversational context; I will rather have to relativize them to the belief sets of individual subjects. The belief set of a person \(a\) in a context \(c\) will be written as \(B_{a,c}\)^1; this notion is to be understood in the sense of my explanations in section 1.3 above.

If we now replace conversational contexts by belief sets, it is no longer possible to apply Stalnaker’s definition of utterance meaning, simply defining the meaning that the utterance of a sentence \(\phi\) has for a subject \(a\) in a context \(c\) as the intersection of \(\phi\)’s diagonal with \(G_{a,c}\). This would, as it were, be too holistic a notion of subjective meaning, because on it all the subject’s beliefs would enter the meaning of a single utterance. One may equally wonder why, on Stalnaker’s account, the entire conversational background is supposed to belong to each utterance meaning. But on my approach, this point leads to a particularly absurd result, at least as far as the speaker’s subjective meaning is concerned. After all, for an honest speaker, the subjective meaning of his utterance is always something that he already believes and hence a superset of his belief set. But intersecting the belief set with one of its supersets always results in the belief set itself; applying Stalnaker’s definition would therefore mean that all utterances a speaker may make (at a certain time) would have the same subjective meaning for him.

---

^1 This presupposes that an utterance context \(c\) always determines the belief set of any person \(a\) in \(c\). It does, of course, happen that an utterance \(<\phi, c >\) of a person \(a\) is interpreted although, intuitively, \(a\) is not part of the utterance situation \(c\) – as, e.g., in letters or taped messages. I will, however, ignore such cases and the complications arising from them.
For the hearer the result would not be quite so absurd because he normally does not already believe the content of an utterance; intersecting the diagonal of the sentence uttered with his belief set only produces a subset of that belief set, and a different one for different utterances. As was already observed, however, even in the hearer’s case does one get an inadequate result that way. For one would intuitively expect that not all of the hearer’s beliefs enter into the meaning an utterance has for him, but at best those that have something to do with that utterance. But it is clear that the very search for those of the speaker’s or hearer’s beliefs that are relevant to a given utterance is an intricate problem; and as far as I know it has never been solved in a satisfying way.

To sidestep this problem area, I will therefore only identify the subjective meaning of an utterance with the diagonal of the sentence uttered, and not with the intersection of this diagonal with the subject’s belief set. As a consequence, the subjective meaning of an utterance is the same for all subjects; for, as we have seen above, the diagonal of a sentence is supposed to reflect the semantic knowledge that is represented by the sentence’s character and that every competent language user has. In this sense the utterance meaning is not subject-dependent. It is nevertheless fair to call it subjective in that it meets individualistic criteria and is therefore internal to any subject; in section 3.9 I will discuss these matters in detail.

Now, by defining its subjective meaning as the diagonal of the sentence uttered, I do not mean to render the subject’s other beliefs irrelevant. In discussing the solution to the problems of the first chapter I will, of course, not just refer to the diagonal but to persons’ entire belief sets; and we will see that these problems can be solved completely by means these two concepts. This result can also be understood as a hint that there is no direct need for a concept of subjective meaning that is suitably relativized to beliefs or background assumptions.

The substitution of propositional concepts by characters leads to another difference with Stalnaker’s theory. Characters are, after all, functions from utterance contexts to intensions and, for the treatment of indexicals, utterance contexts must certainly be more finely individuated than possible worlds; so the diagonal of a sentence is not just a set of worlds, as Stalnaker would have it. Since subjective meanings are supposed to be belief contents, i.e. supersets of belief sets, it follows that belief sets, and belief contents in general, cannot be reconstructed by sets of worlds; their elements must rather have the same structure as utterance contexts.

This assumption does not only make sense; it is even cogent, as becomes apparent by considering beliefs de se and de nunc. In this connection, Lewis (1979b) has provided

---

2 I am assuming that utterances only expand the old context set and do not necessitate a revision of already existing assumptions. If one wants to avoid this restriction, one would have to resort to more general and more complicated theories of belief change like the one in Gärdenfors (1988). However, these complications are off my real topic. I thus concentrate on the simple case of expansion.

3 In the philosophy of mind (and particularly in Fodor 1987) there is, of course, a quest for just the notion of subjective meaning or narrow content that is neither as abstract as the diagonal not as holistic as its intersection with the belief set; such a notion is felt to be needed for an adequate formulation of psychological laws. I will, however, not enter this discussion here; the topic will briefly reappear in section 3.9 though.
convincing arguments that belief contents are generally properties of temporally localized individuals. Such properties can be constructed as sets of triples \( \langle w, s, t \rangle \), consisting of a world \( w \), a subject \( s \), and a time \( t \); the contents of beliefs would then indeed be sets of entities with a context-like structure. It is mainly this point which I want to pursue in the remainder of this chapter.

Finally, in my framework, the use of the diagonal operators \( \Delta \) and \( \delta \) is delicate – again as a consequence of replacing Stalnaker’s propositional concepts with character functions. When introducing them (cf. pp. \( \text{VVV}29f.\text{VVV} \) I had already pointed out that they are only sensibly defined if \( C \subseteq I \) – i.e. if contexts and indices are structurally identical – and that \( \| \Delta \phi \| \) and \( \delta(\| \phi \|) \) are only defined for indices that are at the same time contexts. Stalnaker’s semantic diagonal operator acts on propositional concepts that are functions defined on pairs of possible worlds, viz. on \( R_c \times R_c \); hence in his framework, the corresponding conditions are trivially satisfied. But in applying the diagonal operator to character functions, one cannot always take it that \( C \subseteq I \), i.e. that contexts and indices are to be parametrized in the same way:

Context should definitely contain a world parameter, a time parameter and a speaker (or, as I will also say: subject) parameter, and maybe also a place parameter. Indices must at least contain a world parameter and a time parameter; for these can be shifted by modal and temporal operators. This shiftability was the criterion that Lewis (1980) proposed to decide which parameters belong in the index (cf. \( \text{VVV} \) p. 26\( \text{VVV} \)). Now, this criterion is not a downright empirical one: whether a construction shifts an index parameter or not does, of course, depend on how the construction is semantically analyzed; and there is often more than one possibility as will be illustrated by the discussion of the place parameter in section 2.4. Nevertheless, most authors agree that indices do not contain a subject parameter and that there are no linguistic constructions that shift the speaker (see, e.g., Lewis 1980, p. 85, or Zimmermann, 1991, p. 169). I will also subscribe to this view for the rest of this book.

It is clear that in this case the formal prerequisite for applying the diagonal operator is no longer guaranteed. However, the guiding idea behind the diagonal operator can still be saved, simply by reference to the diagonal \( \partial(\phi) \) of a sentence \( \phi \). This is all that is ever needed in Stalnaker’s theory; and as we have seen on p. \( \text{VVV}30\text{VVV} \), it can be defined directly, i.e. without making a detour via the diagonal operator. I can thus adopt Stalnaker’s ideas without having to defend the structural identity of context and index.

As I said, I would like to discuss the first structural identity above in the next sections of this chapter. That is, I want to show that the contents of beliefs, and thus also the belief set of individual subjects, are aptly modelled in such a way that their elements are entities with context-like structures. At the same time I can then use simple examples of

---

4 Considerations in Heim (1991b) suggest that the first-person pronoun "I" has bound occurrences, just like a third-person pronoun. Bound pronouns are usually treated via a variable assignment; if one was to regard this variable assignment as a parameter of the index, one would then have an argument that, in a sense, even the speaker parameter may be shifted. See also Zimmermann (1991, section 4.1) on the problematic status of variable assignments and the analysis of pronouns in a context-theoretic framework.
indexicals, viz. "I", "now", and "here", to demonstrate how my theory of subjective meaning works.

Section 2.2 is about the subject parameter and the meaning of "I". I assume that contexts are at least pairs \(< w_c, s_c >\) of a world and a subject and I will argue that belief sets, too, must be constructed as sets of tuples \(< w, s >\) – or something equivalent – if the contents of de se beliefs, i.e., of beliefs that a person has about himself, are to be adequately represented. Section 2.3 contains an analogous discussion concerning the time parameter and beliefs de nunc. World, subject, and time thus turn out to be indispensable components of both contexts and the elements of belief contents. Section 2.4 is about the place parameter. It has not been used in the representation of beliefs; but it is usually regarded as a component of context and index. This is why I am taking it into consideration. However, at closer inspection it will turn out that, in all three areas, one can do without it.

Naturally, my discussion of these parameters will not lead to an exhaustive answer as to the formal structure of contexts and belief contents. For our present purposes, we may, however, ignore further parameters – like the variable assignment which would at least have to be considered in any serious semantic account of demonstrative noun phrases; for our later examination of predicates and names it will basically suffice to analyze contexts and belief contents via triples of the form \(< w, s, t >\).

The fact that I do not treat any pronouns other than "I" in this work results in the further restriction that in what follows only speaker meaning will be looked at, i.e., the subjective meaning that an utterance has for the person who produces it. Since the speaker’s "I" is a "you" or a "he" to the hearer, the subjective hearer meaning can only be seriously discussed once the treatment of these pronouns would also be settled. But then this change from the speaker’s to the hearer’s perspective does not seem to be such a big deal that my restriction should have to be considered as serious. Moreover, this exchange will not be relevant for the expressions to be examined in the third and fourth chapters; predicates and proper names do not diverge in speaker and hearer meaning, or only in that speaker and hearer are different subjects who interpret these expressions on the background of different belief sets.

### 2.2 The Subject Parameter

By way of a simple example, I will first show what the subjective meanings of utterances containing the first-person pronoun "I" look like; I am thereby taking the approach of identifying the subjective meaning of an utterance with the diagonal of the sentence uttered. Following Kaplan (1979, p. 91), we can define the character of "I" as:

\[ \| I \| (c) (i) = s_c \]
In every context of utterance, "I" refers to whoever utters it, the speaker \( s_c \) of context \( c \). According to our earlier definitions (cf. p. 4VV2VVV), "I" is indexical – its intension may vary with the context \( c \) – and it is rigid, i.e. its intension does not depend on the index \( i \).

For the discussions to come, I will make the simplifying assumption that contexts \( c \) are pairs \(< w_c, s_c >\) consisting of a world \( w_c \) and a subject \( s_c \) existing in \( w_c \). Similarly, for the time being I will assume that indices only contain a world parameter \( w_i \). As to the interpretation of predicates, I will also ignore all complications to be dealt with later and treat them as descriptive constants whose extension only depends on the index; so whenever \( P \) is a unary predicate:

\[
\models P \models (c) (i) = \{ x \mid x \text{ has property } P \text{ in } w_i \},
\]

and similarly for predicates with more than one place, though my first examples will be unary. So the only syntactic rule whose interpretation is of interest at the moment is the combination of an individual term \( \alpha \) with a unary predicate \( P \) into a sentence \( \alpha P \); it is interpreted by elementhood:

\[
\models \alpha P \models (c) (i) = 1 \text{ iff } \models \alpha \models (c) (i) \in \models P \models (c) (i).
\]

Now the truth-conditions and the diagonal of a simple example containing "I" can be given:

(1) I am sitting on a time bomb.

We will only analyze (1) as consisting of the individual term "I" and the unary predicate "be sitting on a time bomb"; we will not be concerned with the internal structure of the predicate, and especially not with the tense of the verb. We then have:

\[
\begin{align*}
\models \text{I am sitting on a time bomb} & \models (w_c, s_c) (w_i) = 1 \text{ iff } \\
\models \text{I} & \models (c) (w_i) \in \text{be sitting on a time bomb} \models (c) (w_i) \text{ iff } \\
& s_c \in \{ x \mid x \text{ is sitting on a time bomb in } w_i \}.
\end{align*}
\]

Given a context in which, e.g., David Kaplan is the speaker, (1) thus expresses the proposition consisting of all worlds in which David Kaplan is sitting on a time bomb; this is the objective meaning of (1). The diagonal of (1), however, is the set of all contexts in which the speaker of the context is sitting on a time bomb in the world of the context:

\[
\partial (\text{I am sitting on a time bomb}) =
\]

---

5 In Kaplan’s semantics, "I" is treated as a constant. As I have already mentioned (see p. VVV61, fn. 4VVV), this is not entirely unproblematic, since there seem to be cases in which "I" must be construed as a bound variable and others in which it could be; cf. Heim (1990, I.3.2, 1991b) and Chierchia (1989). We can, however, stay clear from such problems in the present work.
{<\ w_c, s_c > | l \ll I am sitting on a time bomb \ll (\ <\ w_c, s_c > ) (w_c) = 1 } = \\
{<\ w_c, s_c > | l \ll I \ll (\ <\ w_c, s_c > ) \in \ll be sitting on a time bomb \ll (w_c) = 1 } = \\
{<\ w_c, s_c > | l \in \{x \mid x \text{ is sitting on a time bomb in } w_c \} }.

Our theory of subjective meaning now says: if a person \( a \) seriously and honestly utters sentence (1) in a context \( c \), then the set \( \partial(1) \) of pairs \( < w, s > \) will be a superset of the belief set \( B_{a,c} \).

This presumes that belief sets in general are sets of pairs \( < w, s > \). The first question thus is how this is to be intuitively understood. The earlier explanation on pp. VVV 34f.VVV does not answer this question, because there belief sets only consisted of possible worlds. But it can be easily adapted to the present case where world-individual-pairs form doxastic alternatives:

We had said that a set \( B_\sigma \) of possible worlds is a model of the belief content of a person \( a \) in that \( G_\sigma \) just contains those worlds that are not excluded by \( a \)'s beliefs: if confronted with a world \( w \) from \( G_\sigma \), \( a \) would not find any reason to believe that \( w \) would not be the actual world. Analogously, we may now say that a pair \( < w, s > \) is an element of \( a \)'s belief set just in case \( a \) considers it possible that \( w \) is the actual world and he is the object \( s \) in \( w \). So the idea is no longer that person \( a \) is to be moved to each possible world \( w \) where he must just decide whether \( w \) could be the actual world – but rather that, in each world \( w \), \( a \) would have to take the role or perspective of an individual \( s \) in \( w \) and then decide whether this perspective is one that he himself might have, according to his actual beliefs: whether he himself could be the individual \( s \) in \( w \).

Given this explanation, \( \partial(1) \) seems to adequately characterize the belief expressed by (1). If Kaplan utters (1), he thereby expresses that, according to his beliefs, he is an object in the world that is sitting on a time bomb in the world, i.e., that all his doxastic alternatives are in \( \partial(1) \).

All this looks straightforward. However, I want to show that it is not only appropriate but even necessary to construe "I"-utterances in this or an equivalent way. So the matter must be discussed thoroughly, the more so since the topic is quite controversial in the literature. I will proceed in the following way:

First of all, the problems of informativity and of internalism provide a test for the adequacy of a theory of meaning; I will therefore demonstrate how these problems can be solved as far as they arise with the word "I". Next, I will show that any conservative attempt to describe the subjective meaning of "I"-utterances by propositions, i.e., sets of worlds, is bound to fail on principled grounds. This part will essentially be an exposition of the main arguments in Lewis (1979b); I will then briefly indicate why my proposal in substance coincides with Lewis’s. Finally, I will compare this proposal with the most important competing approaches to be found in the literature.

Let us first see how the diagonal satisfies the conditions on subjective meanings as formulated in section 1.4, i.e., how it solves the problems of informativity and internalism.

As to the problem of internalism , we have not explicitly formulated it for indexicals like "I" in the first chapter; there it was only introduced for proper names ("London"), natural kind terms ("water") and other predicates ("arthritis"), using Kripke’s,
Putnam’s and Burge’s examples. But, of course, analogous cases can be constructed with “I”. This may be seen from a (slightly adapted) example of Kaplan’s (cf. Kaplan, 1977, p. 533): Kaplan is sitting at home on his sofa watching a video. Sly murderers have placed a time bomb under the sofa and set the timer at five minutes; time is trickling off. Kaplan thinks that he is watching an exciting thriller on TV; the camera moves from a man on a sofa, hard to recognize from behind, to the bomb under the sofa with its clearly visible timer, to the murderers lurking in the background, and then back again. But what he sees is live; the sofa he sees is his own, and he himself is the man who is sitting on it. In his excitement he mumbles to himself: “This man is sitting on a time bomb” and immediately adds: “I’m fine; I am not sitting on a time bomb.” It is clear that the objective meanings of the two utterances: “This man is sitting on a time bomb” and “I am not sitting on a time bomb” contradict each other; for they are, respectively, the proposition that Kaplan is sitting on a time bomb and that Kaplan is not sitting on a time bomb. But intuitively – and this, to be sure, was just the problem of internalism – one would not want to ascribe contradictory beliefs to Kaplan. And this is what we get in the present case: We represent the subjective meaning of “I am not sitting on a time bomb” as the set of pairs \(<w, s>\), where \(s\) is sitting on a time bomb in \(w\), and this certainly does not contradict the subjective meaning of “This man is sitting on a time bomb”, which we may identify with the set of all pairs \(<w, s>\) such that in \(w\), \(s\) sees a man who is sitting on a time bomb.\(^6\)

The example reveals further conditions to be put on any explanation of the subjective meaning of utterances containing “I”. For an utterance with an “I” in it must always differ in its interpretation from the corresponding utterance that results by replacing ”I” by a definite description or a name of the subject. Thus, if someone says “I am sitting on a time bomb”, whatever he thereby thinks may affect his actions in quite a different way than what he thinks on uttering: “This man is sitting on a time bomb.” His other beliefs being equal, he will jump up and run away in the one case, stay and be thrilled in the other. If one wants to explain these different behavioural consequences, one must – ceteris paribus – obviously resort to different beliefs and thus different subjective meanings of the two sentences. Taken as subjective meaning, the diagonal satisfies this condition, too.

A further aspect of the problem of internalism that may arise with “I”, can be seen from another example due to Kaplan (1977, p. 531): Castor and Pollux are identical twins; they are psychologically completely alike – an assumption that should by now be familiar. Castor was born half an hour before Pollux. Now, if Castor says: “I am older than my brother” he thereby speaks truly; when Pollux says the same thing, he speaks falsely. One would still want to say that the two utterances have the same subjective meaning, that Castor’ and Pollux’s thoughts when making these utterances are the same. If the subjective meanings were the objective ones, then Castor’s utterance would have a different meaning from Pollux’s; the one says that Castor is older than Pollux, the other says the opposite. However, the diagonals yield an internal characterization, according to which the twins have the same belief, namely the set of all pairs \(<w, s>\), where \(s\) is older

\(^6\) This is more or less how Lewis would analyze the sentence (cf. Lewis 1979b, section XIII and XIV).
than \( s \)'s brother in \( w \). The difference in truth-value between the two utterances is still captured: Castor is right and Pollux is wrong because – if \( w_0 \) is the actual world \(<w_0, \text{Castor}>\) is in this set whereas \(<w_0, \text{Pollux}>\) is not.

The problem of informativity dissolves in the same way. Since we are only considering the speaker's meaning here, the term is slightly misleading, but the underlying problem is no less urgent. When Heimson says "I am Hume", the objective meaning of his utterance is the impossible proposition; when Hume says "I am Hume", it is the necessary one. But intuitively both Heimson and Hume believe something informative, namely to be someone who is identical with Hume – or plainly: to be Hume, i.e., the set of all pairs \(<w, s>\) such that \( s \) is Hume in \( w \).\(^7\) This set is precisely the diagonal expressed by "I am Hume"; and it is informative, as required. Thus, at least in the case of "I"-utterances, our theory of subjective meaning adequately meets the challenges presented in chapter 1.

The examples just discussed have already shown that the objective meanings of "I"-utterances are no good as subjective meanings. The same result can be reached by the following more fundamental considerations.

Let us first get straight how it can ever happen that a person has anything but himself in his doxastic alternatives. This may at first appear mysterious; for, in a sense, he who utters "I" knows what his utterance refers to – unlike the hearer who may not see the speaker and not recognize his voice. But this knowledge only reflects the fact that the doxastic alternatives of a person \( a \) must always correspond to his self-image: \( a \)'s belief set only contains pairs \(<w, s>\) such that in \( w \), \( s \) has all those properties that \( a \) believes to have in the real world. To be sure, this heavily restricts the set of \( a \)'s doxastic alternatives, but not enough that only \( a \) himself may appear in them; this would only be the case if \( a \)'s self-image covered all of \( a \)'s essential properties.

Hence, in a deeper sense, \( a \) need not know who he is; subsequent examples will illustrate this point. So a person can have distinct individuals among his doxastic alternatives, which then excludes that the objective meaning of "I"-utterances is also its subjective meaning. For it is always one and the same individual, the respective speaker, who enters the objective meaning of an "I"-utterance; any variations in the doxastic alternatives cannot be represented in that way.

But the crux of the matter here is the following, stronger claim: the subjective speaker's meaning of "I"-utterances cannot be represented as a proposition, i.e., a set of possible worlds; in other words: de se beliefs, beliefs a person has about himself, are not propositional attitudes. Only this proves that a finer individuation of doxastic alternatives is unavoidable; and only this justifies our theory which is, after all, based on such a finer individuation.

It is clear how all these claims are to be justified: One must find an intuitively convincing example of a person who is unclear as to which of two persons in a given world he is, so that his belief set contains at least two distinct pairs \(<w, s>\) and \(<w, s'>\) for one world \( w \). For, obviously, in such a case the two doxastic alternatives could not be

---

\(^7\) The analysis of proper names in the fourth chapter will bring out that this set is not identical to the set of all pairs \(<w, \text{Hume}>,\) where Hume exists in \( w \),
distinguished by worlds alone – the worlds are the same, only the perspectives or roles the person has in them differ. At the same time such examples further illustrate the earlier, weaker point, that a person may have different individuals as doxastic alternatives. Lewis (1979b, p. 535) adapted an example of Perry’s (1977, p. 492) in order to demonstrate this possibility:

Rudolf Lingens is lost in the Stanford library. He suffers from a strong form of amnesia; he does not remember his name, nor does he know where he is and how he got there. Lingens reads a lot of books, and in the library he finds a biography of Rudolf Lingens, which says, among other things that Lingens has lost his way in the Stanford library and is now on its sixth floor; in order to get out, he would just have to go down the stairs. Lingens also finds the biography of Adolf Dingens, a man who also suffers from amnesia and is also lost in a library, the Widener library. Dingens is on one of the lower floors, and to get out of the library he would have to go up the stairs. On the basis of his perception and his reading, Lingens has formed the hypothesis that he must be one of these two men. But he does not have enough information to know which of the two he is; neither does he know whether he must go upstairs or downstairs to find the exit. Only by choosing one of these alternatives will he find out whether he is Rudolf Lingens or Adolf Dingens.

If one now were to describe Lingens’s belief that he is either Lingens or Dingens by way of a proposition, i.e., if one assumed that his belief set only contains possible worlds, how could one then describe the additional knowledge Lingens acquires by going down the stairs? Which worlds would be eliminated from Lingens’s belief set when he finds out that he is Lingens in Stanford and not Dingens in Widener – given that there is one world in which both of these persons suffer from amnesia and they are both lost in libraries? One should note that it would not be correct to say: Before Lingens went down the stairs and found the exit, his belief set contained worlds like \( w \), in which Lingens is in Stanford and Dingens in Widener, and worlds like \( w' \), in which Dingens is in Stanford and Lingens in Widener; but after finding the exit, all worlds of the second kind are dropped from his belief set. For this is wrong in that Lingens already knew before that Lingens is in Stanford and Dingens in Widener– because this information he got from the two biographies he had read.

There is, however, no problem explaining Lingens’s gain of knowledge if we construe his doxastic alternatives as pairs of worlds and individuals: As long as Lingens does not yet know whether he is Lingens or Dingens, his belief set contains both pairs \( <w, \text{Lingens}> \) and pairs \( <w, \text{Dingens}> \) – for all \( w \) that are compatible with his beliefs; but once he has found the exit, all pairs of the form \( <w, \text{Dingens}> \) are eliminated from his belief set.

The fact that \( de \ se \) beliefs cannot be described by sets of worlds becomes even clearer by another example due to Lewis (1979b, p. 521f.):

A possible world \( w \) is inhabited by two gods, Zeus and Yahweh. Zeus lives on the coldest mountain in \( w \) and from there throws down thunderbolts, whereas Yahweh is sitting on the tallest mountain and throws down manna. The two gods are propositionally omniscient, they know exactly which of all possible worlds they inhabit; their common belief set, if modeled as a set of worlds, is the singleton of \( w \). Still, Zeus and Yahweh are
not really omniscient: neither of them knows whether he is on the tallest or on the coldest mountain, whether he flings down thunderbolts or throws down manna; neither of them knows who he is. (The gods not only have total propositional knowledge; if they acquire their knowledge by something like perception, this would also have to be total; otherwise they could use their respective perceptive perspective on the word to recognize their place in it.) So there is still something to know for Zeus and Yahweh, namely: who they are. But this information obviously cannot be propositional; for it cannot be captured by a reduction of their belief set if that only contains one element.

In this story, too, the problems disappear once one takes belief sets to be sets of doxastic alternatives: Zeus and Yahweh first only believe \( \{ < w, \text{Zeus} >, < w, \text{Yahweh} > \} \), but when they find out who they are, Zeus’s belief set reduces to \( \{ < w, \text{Zeus} > \} \), and Yahweh’s to \( \{ < w, \text{Yahweh} > \} \).

All these stories are about unusual persons in unusual epistemic states. This is, however, not essential. The stories could equally be about ourselves, if only we stop taking ourselves to be unique – i.e., once we admit possible worlds to our belief sets in which the whole history of our solar system happened twice in one universe. For in such possible worlds each earthling would have an indistinguishable doppelgänger and thus two doxastic alternatives in one world.

That \( \text{de se} \) attitudes cannot be understood as propositional attitudes was argued for at length in Lewis (1979b); I have just repeated his most important arguments. Lewis also proposed to use sets of possible individuals – instead of sets of worlds – in representing the contents of attitudes. Since Lewis presupposes an ontological framework, according to which each individual can only exist in one world (having so-called counterparts in other worlds), he can indeed represent attitude contents simply as sets of individuals; for in this framework each individual \( s \) exactly corresponds to one pair \( < w, s > \), where \( w \) is the world in which \( s \) exists. However, I do not wish to commit myself to Lewis’s anti-haecceitistic ontology here; nor do I want to go into all the intricate arguments in its favour, as presented in Lewis (1986), part 4. I would rather take the natural – if maybe naive – point of view, according to which a possible individual may well exist in several possible worlds. So if \( < w, s > \) is a subject’s doxastic alternative, then \( < w', s > \) can also be one of his doxastic alternatives even for distinct worlds \( w' \). Consequently, we must represent attitude contents as sets of world-individual pairs.

Sets of world-individual pairs can also be regarded as unary properties: usually (cf. p. VVV23VVV) a property is defined as an intension of a unary predicate, which is represented as a function assigning to each world \( w \) the set of those individuals that have the property in \( w \). To any set of world-individual pairs there corresponds exactly one such property, and \textit{vice versa}. For if \( Z \) is a set of world-individual pairs and the function \( f \) is defined by: \( f(w) = \{ s \mid < w, s > \in Z \} \), then obviously \( f \) is just this property; likewise, for any property \( f \), the corresponding set of world-individual pairs is: \( Z = \{ < w, s > \mid s \in f(w) \} \).

Lewis therefore also speaks of \( \text{de se} \) beliefs as self-ascriptions of properties: when a set of possible individuals or world-individual pairs represents a person \( a \)’s belief content, this means that \( a \) ascribes to himself the property determined by this set.
Propositional belief is then simply a special case of self-ascription. According to Lewis it is represented by a set of individuals containing, for any world \( w \), either every or no individual from \( w \); on our account, it would have to contain, for any world \( w \), either all pairs \( < w, s > \), where \( s \) is any individual existing in \( w \), or no such pair at all. Accordingly, and again abstracting away from any background assumptions, the subjective meaning of ”Snow is white” would be the set of all pairs \( < w, s > \), where snow is white in \( w \) and \( s \) exists in \( w \). It is clear that in such a case \( s \) is completely irrelevant for determining the set: the speaker’s perspective does not enter the subjective content of his utterance. These considerations show that there is no problem in giving a uniform treatment of propositional and non-propositional belief in terms of properties. In the course of this work, we will, however, see that practically all beliefs and, at any rate, all subjective meanings have an irreducible perspectival or de se component, making propositional belief practically irrelevant.

Let us briefly sum up: If we identify the subjective meaning of an utterance with the diagonal of the sentence uttered, we obtain a result that coincides in substance with Lewis’s proposal to analyze beliefs in general as self-ascriptions of properties.\(^8\) We have given two kinds of arguments. We have shown that, first, our proposal solves the problems one runs into with the subjective meanings of ”I”-utterances, and, second, that doxastic alternatives must at any rate be more finely individuated than possible worlds. We have not given any argument that this finer individuation must be done in terms of world-individual pairs. But then this is the second easiest way, once possible worlds alone turned out to be insufficient; and all our examples have proved this procedure natural and feasible. But it is not undisputed; there are, indeed, quite different approaches to these problems. To end this section, let me therefore discuss alternative semantic theories of utterances with ”I”, or de se attitudes; they come in three different kinds.

First, there are approaches that agree with Lewis that the subjective meaning of an ”I”-utterance is a property, but at the same time do not make the systematic connection between properties and utterances via diagonalization. Such theories have been advocated by von Stechow (1982, 1984) and Chierchia (1989), among others. In von Stechow’s case the basic idea is, roughly, to analyze belief in general as a de re attitude to structured propositions and then identify the subjective belief content with the de se attitude underlying the de re attitude. Chierchia, again roughly, treats ”I” as a variable bound by a \( \lambda \)-operator and thus semantically interprets a sentence containing ”I” as a property. However, both von Stechow and Chierchia concentrate on the interpretation of belief reports – and particularly of de se belief ascriptions – whereas in this work I restrict myself to simple, unembedded sentences. These approaches are therefore hard to compare to mine.

\(^8\) Lewis (1979b) only argues that properties, not propositions, are the objects of belief, without formulating an explicit mechanism assigning to each utterance the property it expresses. But Lewis (1983, p. 230) – the first postscript to “General Semantics” – contains a hint indicating that our method of diagonalization might be in his sense. For there he says: ”It will not escape the reader of ‘Attitudes De Dicto and De Se’... that these ‘contexts’ are the same as the ‘subjects’ that self-ascribe properties, and that sets of them are the self-ascribed properties. This agreement might well be exploited in the semantic analysis of de se attitude-sentences.”
A second family of theories agrees with Lewis that the content of a de se belief cannot be a proposition, but then the subjective meaning of the corresponding utterance is not identified with a property or the diagonal of the sentence uttered, but with its entire character; the most prominent advocates of this theory are Kaplan (1977) and Perry (1977, 1979). But what motivates this approach? Characters certainly encode more information than the diagonals they correspond to, but it is not clear whether this additional information ever becomes relevant in capturing subjective meaning. To show that one would have to find examples of two utterances with the same diagonals but different characters, which at the same time intuitively carry distinct subjective meanings. But such examples are never discussed by either Kaplan or Perry. On the contrary, as we have already seen above, all cases used by Perry (1977, 1979) or Kaplan (1977) to back up their theory can equally well be treated under the weaker assumption that it is the diagonal and not the character of an utterance that makes up its subjective meaning.

Apart from that, Perry and Kaplan meanwhile no longer support their original view. Perry (1988) now identifies the subjective meaning of an utterance with the proposition that the utterance is true; this is, of course, substantially the same as the diagonal of the utterance – even though Perry, in his situation-theoretic framework, takes a different approach to propositions. Kaplan (1989), on the other hand, doubts that the character theory can be extended to a general theory of subjective meaning, because he thinks it

9 Let us try to construct two sentences φ and φ’ that have the same diagonal but different characters. We may assume that there is a semantic distinction between the referential and the attributive use of definite descriptions and (as already explained on p. VVV29VVV) that this difference can be interpreted as a one between a context-dependent and an index-dependent reading of the description:

\[ \parallel (\text{the} \ P)_{\text{ref}} \parallel (c) (i) = \text{the object in } w_c \text{ which is } P \text{ in } w_c \]
\[ \parallel (\text{the} \ P)_{\text{attr}} \parallel (c) (i) = \text{the object in } w_i \text{ which is } P \text{ in } w_i. \]

(Here P is a unary predicate and it is taken for granted that each world contains exactly one object that is \( P \).) A sentence like (a) would then have the readings (b) and (c) with characters \( \parallel (b) \parallel \) and \( \parallel (c) \parallel \), as given in (d) and (e):

a. The first person to climb Mount Everest is a New Zealander.

b. (The first person to climb Mount Everest)ref is a New Zealander.

c. (The first person to climb Mount Everest)attr is a New Zealander.

d. \( \parallel (b) \parallel (c) (i) = 1 \) iff that individual from \( w_c \) who makes the first successful ascent to Mount Everest in \( w_c \) is a New Zealander in \( w_i \).

e. \( \parallel (c) \parallel (c) (i) = 1 \) iff that individual from \( w_i \) which who makes the first successful ascent to Mount Everest in \( w_i \) is a New Zealander in \( w_i \).

So \( \parallel (b) \parallel \) and \( \parallel (c) \parallel \) are distinct, though they have the same diagonal, since replacing \( w_i \) by \( w_c \) in either case amounts to the same thing. Hence our theory would predict here that the difference between the referential and the attributive reading of a definite description is only one in objective meaning, not in subjective meaning; Kaplan’s and Perry’s theory, though, would imply a difference in subjective meaning.

I do not want to decide her which of the two is the more adequate prediction, because this would necessitate a much more comprehensive discussion of the problems surrounding the interpretation of definite descriptions in general and the referential/attributive distinction in particular. The example was only meant to illustrate that my proposal also differs empirically from Kaplan’s and Perry’s character theory.
cannot solve the problem of informativity in the case of names: according to Kaplan, proper names are not context-dependent.

The third theory is that of Stalnaker (1981), which is his most detailed discussion of *de se* attitudes. The method of diagonalization is as important there as it is in his other writings, and in that we have followed him. But in another respect, he prefers to stay conservative, thereby most drastically differing from Lewis’s approach. For, unimpressed by Lewis’s, Perry’s and Kaplan’s arguments, he sticks to the view that all beliefs, including those about oneself, can be analyzed as attitudes to propositions. This view is the next and last to be considered here.¹⁰

Stalnaker also attends to Lingens’s serious amnesia. In his variant, Lingens gleans from the Lingens biography, among other things, that Lingens is the cousin of a notorious spy called Orcutt. Now he is standing somewhere in the Stanford library, wondering: ”Am I Lingens? Am I the cousin of a notorious spy?”, which – that much is clear – can be a serious question for him. According to Stalnaker, this is how the content of Lingens’s belief is to be described: In all worlds that are compatible with whatever Lingens believes, there is a person called ”Lingens” who has all the properties the biography Lingens read ascribes to Lingens; among these is the property to be the cousin of a spy. Moreover, in all these worlds there is exactly one person who is sitting in a library, suffers from amnesia, has just finished reading a biography of a person called ”Lingens” and now wonders: ”Am I the cousin of a spy?”. But only in some, not in all worlds are these two persons identical with each other. If Lingens receives a positive answer to his questions, then all those worlds in which he is called ”Lingens” and is distinct from the person asking the question to himself, will be eliminated from his belief set.

The upshot of Stalnaker’s proposal – for which, strangely, only circumlocutions can be found in Stalnaker’s writings – is thus: the subjective meaning of ”I” in utterances or thoughts can be expressed by definite descriptions like ”the one who makes this utterance” or ”the one who thinks this thought”. On the one hand, this sounds incontestable; after all, we have also described the character of ”I” as well as its diagonal as ”the speaker”. On the other hand, one may wonder how ”this utterance” and ”the one who makes this utterance” are to be understood and whether *de se* attitudes can really be represented by propositions in this way.

An assessment of this question is particularly difficult because Stalnaker’s frame of thought differs so much from ours. The basic difference lies in his view of the relation between semantics and pragmatics, which for Stalnaker are two quite distinct domains. Apart from the – already critically discussed – incompatibility of characters and propositional concepts, it implies an ontological difference. Our basic entities are sentences and contexts, and utterances are defined in terms of them; Stalnaker’s basic entities, however, are sentences, which semantics is concerned with, and utterances, which

¹⁰ Boër und Lycan (1980) seem to take a conservative view on *de se* attitudes, too; at least they argue that the *de se* problem is only a special case of the *de re* problem. But they say little about the analysis of *de re* attitudes and they do not explicitly address the question of whether it is possible to treat all beliefs as propositional attitudes.
pragmatics is about. These two ontological frameworks are certainly interchangeable – at least if Stalnaker adds a relation saying which sentences are uttered in which circumstances. But this difference explains that we cannot raise the objection against Stalnaker that an utterance is a sentence in a context $<w,s>$ and that thus ”this utterance” already makes reference to the subject of the utterance.

But maybe a variation of this argument is closer to the point? Stalnaker seems to merely shift the indexical reference from ”I” to ”this utterance”, thus not going beyond Reichenbach’s (1947, Kap. VII, § 50) old insight, according to which indexicals can be represented as ‘token reflexive’. But this argument fails too. For it must be kept in mind that Stalnaker relativizes all pragmatic discussion to context sets that only distinguish possibilities relevant for the respective concerns. This may well only involve the utterance being made, and no other one; then ”this utterance” means as much as ”the only utterance”.

Returning to our framework, it is now clear that this answer is not sufficient. So there only remains to also interpret ”this utterance” descriptively, e.g. as ”an utterance made by such-and-such a person in such-and-such a situation” where ”such-and-such” stands for the speaker’s beliefs about himself, his situation, and his utterance.

Can Stalnaker’s theory, construed in this way, be saved in our framework? It would at first seem so. Let us reconsider the above Dingens version of the Lingens example, which was supposed to demonstrate the irreducibility of de se attitudes. Stalnaker would presumably analyze the case like this: In each world of Lingens’s belief set there is a person $x$ who is called ”Lingens”, has such-and-such properties and is in Stanford, as well as a person $y$ who is called ”Dingens”, has such-and-such properties and is in Widener, and a person $z$ who has such-and-such properties, is in a library in such-and-such a perceptive situation and now asks himself: ”Am I Lingens?”; let us refer to the properties assumed for $x$, $y$ and $z$ as $P$, $Q$ and $R$, respectively. Then Lingens believes that there are $x$, $y$ and $z$ such that $Px$, $Qy$ and $Rz$ and $z = x$ or $z = y$. So at first sight four kinds of worlds can be found in his belief set: (a) those in which there are $x$, $y$ and $z$ such that $Px$, $Rz$, $z = x$, $Qy$ and also $Ry$, (b) those in which there are $x$, $y$ and $z$ such that $Px$, $Rz$, $z = x$, $Qy$ and not $Ry$, (c) those in which there are $x$, $y$ and $z$ such that $Qy$, $Rz$, $z = y$, $Px$ and also $Rx$, and (d) those in which there are $x$, $y$ and $z$ such that $Qy$, $Rz$, $z = y$, $Px$ and not $Rx$. It must be noted, though, that the worlds of kind (a) are the same as those of kind (c); for, surely, conditions (a) and (b) both just say that there is someone to whom $P$ and $R$ apply as well as someone to whom $Q$ and $R$ apply.

There now is a good answer to the question as to which worlds are dropped from Lingens’s belief set when he finds out that he is Lingens: namely, precisely those of kind (d) (though not the (c)-worlds – which, as we have seen, are just the (a)-worlds). Was our above conclusion premature, then? Not really. We had, indeed, neglected the fact that – simply by wondering ”Am I Lingens?” or thinking something else – Lingens does have a description of himself that he does not have of anybody else – namely as someone asking himself ”Am I Lingens?” or having such-and-such a thought. Moreover, the rest of the perceptive situation can now be hooked onto this; for, obviously, the one who has such-and-such a thought is just the one who has such-and-such perceptions. This detail creates an almost inevitable asymmetry in the persons existing in Lingens’s belief set; and
it is this detail that Stalnaker’s position is built on. Let us suppose Lingens’s belief set only consisted of the (a)- or (c)-worlds. This is pretty fantastic; Lingens would not only have to not exclude but even to assume that his amnesiac fellow-sufferer over in the other library is in the same perceptive situation and also wondering "Am I Lingens?". But then this is certainly not impossible; apart from amnesia, there are, after all, telepathy and all kinds of miraculous things. So finally there does not seem to be an answer to the question as to which proposition Lingens learns when he gets the information that he is Lingens.

But even here Stalnaker sees a last resort, which consists in saying that somehow the (a)-worlds are not quite the (c)-worlds; if so, this information would then exclude the (c)-worlds. Stalnaker explicitly takes this way out when confronted with Lewis’s example of the two gods, in which connection it may be easier to understand. It will be remembered that Zeus and Yahweh had complete propositional knowledge, only considering a single world possible; at the same time they are not supposed to know who they are. That is, Zeus does not know whether he is Zeus or Yahweh, whether he is sitting on the coldest mountain or on the highest one; and when he utters "I am Zeus" he does not know whether it is he who is uttering the sentence. How can Stalnaker represent his increase of knowledge on learning that he is Zeus. In his comment Stalnaker says that in this case Zeus’s original belief state should, after all, not be represented by a single world; rather, Zeus’s belief set contained two worlds w and w’ with no qualitative difference whatsoever, but which still differ in that it is Zeus who is sitting on the highest mountain in w, whereas in w’ it is Yahweh – though Yahweh in w’ has the same properties as Zeus in w. Stalnaker thus takes a stand for haecceitism here in accepting that there are distinct but indistinguishable objects and worlds and that questions of identity among objects across worlds are not reducible to questions about qualitative properties. This way Stalnaker can represent Zeus’s increase in knowledge as excluding world w’ and thus maintain his position that de se beliefs can be represented propositionally.

On the other hand, Stalnaker himself admits that it does not make a big difference whether one talks about distinct but indistinguishable worlds or of distinct centered worlds – where centered worlds are just our world-individual pairs <w, s>. For any centering s of world w in a Lewis-style belief set of a person a Stalnaker may assume a variant w’ of w that does not qualitatively differ at all from w, but in which a is s. Any substantive difference between Lewis’s and Stalnaker’s method seems to disappear this way.

Still, Stalnaker makes out an advantage in his method: Although objective and subjective meaning are well distinguished on Lewis’s approach, the informational content of an utterance is relativized to particular subjects, differing from one subject to the next. Stalnaker, on the other hand, can also distinguish between objective meaning and informational content; but since he regards the informational content of an utterance as a set of worlds, it is always the same even for different subjects – at least if these subjects make sufficiently similar background assumptions. Stalnaker holds that thereby his theory gives a simpler and more direct account of communication: If in a context c a person a utter a sentence φ with informational content A and b hears and understands this utterance, then b’s beliefs are just enriched by the set A – if A is a set of worlds. If, however, the subjective meaning of φ for a is the diagonal φ(φ) of φ, then it is, of course,
not this set that is added to hearer b’s belief set; if, e.g., a utters "I am Lingens", then b, who takes this utterance to be true, does not believe that he is Lingens, but that a is.

Is this last argument of Stalnaker’s waterproof? I believe that it is just the other way round: the alleged advantage turns into a disadvantage. Of course, how to account for communication within a Lewisian framework is still open. But then it does seem to me that this permanent switching between I and You in communication is a performance that theory should not just explain away. But the bigger problem is that, by giving up subjective relativization, Stalnaker’s approach gets into complementary troubles. Let us once again consider Zeus and Yahweh. According to Stalnaker, their common belief set is the set \{w, w'\}, where w is the real world, in which Zeus is sitting on the coldest and Yahweh on the highest mountain, and w' a qualitatively absolutely identical world in which, however, Yahweh is on the coldest mountain and Zeus on the highest one. If Zeus now learns that he is Zeus, then his belief set is just \{w\}; and if Yahweh learns that he is Yahweh, so is his belief set. But now let us suppose that not only Zeus, but Yahweh too believes to be Zeus. According to Stalnaker, \(B_{\text{Zeus}}\) then is \{w\}, but \(B_{\text{Yahweh}}\) is \{w'\}. We would certainly want to say in this case that Zeus and Yahweh believe the same thing. But in Stalnaker’s framework there seems to be no way of explicating this sense of "believing the same thing". To be sure, this result was to be expected; for a finer individuation of doxastic alternatives also involves more possibilities of describing sameness and distinctness of beliefs.

After all these turns of reasoning my final conclusion, then, is that, facing the problem of internalism, Stalnaker ultimately fails; at least some beliefs come out as distinct though they are identical from an internal psychological perspective. As a result, our proposal to identify the subjective meaning of "I"-utterances with their diagonal has stood the test of comparison. It will shortly turn out that it is also of use for other linguistic expressions.
2.3 The Time Parameter

The preceding section has shown that the elements of belief sets cannot just be possible worlds, but must rather be expanded by a subject parameter. In this section, we will present an analogous argument for a further parameter, the time parameter: belief sets must be constructed as sets of triples \(<w, s, t>\) consisting of a world \(w\), a subject \(s\), and a time interval \(t\) – to be understood in the sense that \(<w, s, t>\) is an element of a person \(a\)’s belief set \(B_{a,w,s,t}\) in a world \(w^o\) at a time \(t^o\) if, and only if, \(a\) considers it possible in \(w^o\) at \(t^o\) that \(w\) is the actual world, \(t\) is the present time and he is person \(s\) in \(w\) at \(t\).

Given our extensive discussion on the necessity of the subject parameter, we already know what counts in this kind an argument: we must find an example in which, intuitively, one would want to ascribe a person an informational increase, in a way unavailable to a theory treating belief contents just as set of worlds or world-individual pairs. The following story, due to Stalnaker (1981, p. 142/143) and slightly embellished by myself, provides a case in point:

Searching his car for his alarm clock, O’Leary climbs into the boot. It is nighttime and O’Leary knows that no one will hear him when he calls; he will have to wait until daytime. He decides to relax and sleep. As a matter of fact, O’Leary’s alarm-clock is in the boot, the alarm being set for both 4 a.m. and 5 a.m., which is known to O’Leary. When the alarm rings at four o’clock, O’Leary briefly wakes up. He wonders whether this was the first or the second time the alarm went off, i.e., whether it is four o’clock or already five, but immediately goes back to sleep. At five o’clock the alarm strikes again. O’Leary wakes up, but does not remember that he had already been awake that night. Again he wonders whether this may be the first or the second ringing, i.e., whether is only four o’clock or already five. And again, he is soon overwhelmed by his sleep.

Moreover, let us assume that, first, O’Leary takes it that he is going to wake up twice that night – after all, he set the alarm himself and he knows that he is waking up when, and only when, his alarm goes – and, second, that on his second awakening he does not remember having already woken up before – his situation is too dramatic to notice such trifles.

If one was going to represent the various contents of O’Leary’s beliefs by means of sets of worlds – ignoring, as we safely may, the subject parameter for the discussion of this example – one would then have to say that, both at four o’clock and at five o’clock, O’Leary’s belief set contains only such worlds where O’Leary wakes up both at 4 a.m. and at 5 a.m., wondering whether it is four or five and being unclear about whether he has already been awake before.

Let us now suppose that O’Leary, when waking up for the first time, gets an answer to his question, so that at 4 a.m. he now knows: it is now four o’clock, not five. How would we then describe O’Leary’s new belief? Which worlds are we to drop from his belief set? It should be clear that there cannot be a good answer to this question. Thus, e.g., it is not correct to eliminate all those worlds in which O’Leary wakes up at five, wondering whether
it is four or five; for he is still convinced that he is waking up both at four o’clock and at five o’clock.

The example about O’Leary can also be used to make another point. In analogy with Stalnaker’s proposal for interpreting “I”, one could try to analyze ”now” as subjectively synonymous with ”the time at which I utter ‘now’ in such-and-such a situation”, where ”such-and-such” abbreviates the assumptions the speaker makes about his present utterance situation; among other things, this includes assumptions on the relation of the time of utterance to other times. The story about O’Leary then shows that in general this is not sufficient; for it provides an example in which this analysis is not a definite description, because the predicate it contains equally applies to two times.

Now the parallel between the O’Leary case and the strengthened variant of the Lingens example (cf. pp. VV73f.) becomes clear, too: For O’Leary there are two times that are qualitatively identical and which he cannot distinguish – just like for Lingens there are two persons with identical properties and which he cannot tell apart. And just as, on the basis of purely qualitative properties, Lingens cannot decide which of these two persons he is identical with, O’Leary – again, on the basis of purely qualitative properties – cannot decide which of the two moments is the present one.

The strategy taken for “I” and attitudes de se now suggests itself for ”now” and attitudes de nunc: Adopt a finer individuation of belief sets and expand their elements by an additional time parameter t, thus obtaining triples < w, s, t >. One may then even regard belief contents as properties of individuals’ time slices or as temporally relativized properties of individuals, i.e., as properties that individuals have at given times. Again, that a set of triples < w, s, t > represents a person a’s belief content at a time t° in a world w° then means that, at t° in w°, a self-ascribes the property given by this set.

To return to the above example: Let w° be the world in which O’Leary lives; O’Leary’s belief set in w° contains the set of those triples < w, s, t > such that: w corresponds to O’Leary’s beliefs about w°, s is a person who exists in w at t and corresponds to O’Leary’s momentary self-image, and t is either four a.m. or 5 a.m. If at four o’clock O’Leary learns that it is four and not five, then all those triples are dropped from this set in which t is 5 a.m. There is no problem, then, in representing O’Leary’s increase in knowledge. And that in the above example O’Leary has the same thoughts at four o’clock as he has at five o’clock is captured by our method, too. But the method not only solves the problem of internalism, but also the one about informativity: the content of O’Leary’s belief when believing at four o’clock: ”It is now four o’clock.” – the set of triples < w, s, t > such that t is 4 a.m. – is again neither universal nor empty and thus informative; ignoring s, it is simply the property of being 4 a.m. that O’Leary ascribes to the present moment.

We may thus take it that there is ample motivation for modeling belief contents as sets of triples < w, s, t >. It must be kept in mind, though, that the additional temporal parameter

11 Stalnaker himself again appears to approach the O’Leary example with his method of haecceitistic multiplication (cf. Stalnaker 1981, p. 143). However, the arguments raised against this method of analyzing attitudes de se could be repeated mutatis mutandis. This is why I am again following Lewis (1979b) in my treatment of attitudes de nunc.
that it is also relevant for the subjective meaning of all expressions that directly or indirectly refer to the present time. Among these are expressions like "present", "yesterday", "next week" as well as the present, past, and future tenses; they, too, must be analyzed as deictic operators making reference to the time of utterance.

However, since the semantics of tenses and temporal adverbs involves additional complications, I will not go into it any further. Thus, when illustrating my theory of subjective meaning with utterances expressing de nunc attitudes, I will restrict myself to giving semantic rules for simple sentences with "now", pretending that they are tenseless. I will also neglect all uses of "now" in which it does not denote the time of utterance but is used anaphorically – as, e.g., in: "O'Leary climbed into the boot. Suddenly the lid closed. Now he was stuck."

To begin with, we must assume that the treatment of sentences containing temporal expressions requires both the index and the context to be extended by a time parameter. Consequently, contexts $c$ must be regarded as triples of a world $w_c$, a subject $s_c$, and a time interval $t_c$; $t_c$ is the time of utterance and is needed in the semantics of deictic temporal expressions. The notion of an utterance time must not be taken too literally, though; given a sentence $\phi$ in a context $<w_c, s_c, t_c>$, it would certainly be wrong to identify $t_c$ with just the time interval used for the production of $\phi$ in $w_c$ by $s_c$. Indeed, we do not even assume that $\phi$ is uttered by $s_c$ in $w_c$ at $t_c$ at all. As Kaplan (1977, p. 522) points out, as pairs of contexts and sentences, utterances are abstract constructions, not concrete events. The only condition we put on a context $c = <w_c, s_c, t_c>$ is that $s_c$ exist in $w_c$ at $t_c$. And when we consider a sentence $\phi$ in a context $c$, this means that we determine what the meaning of $\phi$ would be, if $\phi$ were uttered in $c$.

Indices $i$, then, are pairs $<w_i, t_i>$; as we already announced in section 1.2, the time parameter in the index will be needed for the semantics of temporal operators. Thus, e.g., "always" may roughly be analyzed by the rule:

$$\parallel\text{always }\phi\parallel(c)<w_{i}, t_{i}> = 1 \text{ iff } \parallel\phi\parallel(c)<w_{i}, t> = 1 \text{ for all times } t$$

In their interaction with the tenses, though, temporal operators face additional complications. This is why I do not want to go into them any further.

We will first have to rewrite the rule for interpreting predicates; it now reads:

$$\parallel P\parallel(c)(i) = \{x | x \text{ has property } P \text{ in } w_i \text{ at } t_i\}.12$$

---

12 Not all predicates are temporally dependent; thus, e.g., all abstract or mathematical predicates do not depend on time – or the world: $\parallel$ being an even number $\parallel(c)(i) = \{x | x \text{ an even number}\}$. It is not clear to me whether there are also predicates that are not temporally dependent but do depend on the world. These would have to be predicates whose extensions contain objects that do not exist in all worlds but do exist at all times in the worlds where they exist at all.
2.3 The Time Parameter

The rule for interpreting predication as well as that for the interpretation of "I" can be copied from above; we must only keep in mind that \(c\) and \(i\) now stand for \(<w_c, s_c, t_c>\) and \(<w_i, t_i>\), respectively:

\[
\| \alpha P || (c) (i) = 1 \iff \| \alpha || (c) (i) \in \| P || (c) (i),
\]

\[
\| I || (c) (i) = s_c.
\]

"Now" will be treated as a sentential adverb, i.e., if \(\phi\) is a sentence, then so is "now \(\phi\)" and we have:

\[
\| \text{now } \phi || (c) (i) = 1 \iff \| \phi || (c) (<w_i, t_c>) = 1.
\]

It is of importance for the interpretation of "now" that the time of utterance is not generally identified with the time it takes to utter "now" or "now \(\phi\)". For, depending on the context, "now" can refer to time intervals of varying length, all of which must, however, cover the actual or counterfactual time of production.

We can treat other temporal locutions in analogy to "now" (still neglecting any complications arising from tense). E.g., "tomorrow" can be interpreted as:

\[
\| \text{tomorrow } \phi || (c) (i) = 1 \iff \| \phi || (c) (<w_i, t_{tomorrow}>) = 1, \text{ where } t_{tomorrow} \text{ is the day following the day of } t_c.
\]

As a first example, let us consider (2):

(2) I am now sitting on a time bomb.

\[
\| \text{now } (I \text{ am sitting on a time bomb}) || (c) (<w_i, t_c>) = 1 \iff
\| I \| (c) (<w_i, t_c>) = 1 \iff
\| I \| (c) \in \| \text{be sitting on a time bomb} || (b) (<w_i, t_c>) = 1 \iff
s_c \in \{ x \mid x \text{ is sitting on a time bomb in } w_i \text{ at } t_c \}.
\]

Let us imagine that (2) is uttered in a context consisting of the real world \(w^o\), David Kaplan, and 12 o’clock noon on the 20th of July, 1978. Then the objective meaning of this utterance is the set of all worlds in which David Kaplan is sitting on a time bomb at noon of July 20, 1978.

The subjective meaning of the utterance, however, is the diagonal of (2), which is the set of all contexts \(<w_c, s_c, t_c>\) in which, at the time of the context, the speaker of the context is sitting on a time bomb in the world of the context:

\[
\partial (I \text{ am now sitting on a time bomb}) =
\{ c \| \text{now } (I \text{ am sitting on a time bomb}) || (c) (i(c)) = 1 \} =
\{ <w_c, s_c, t_c> \mid s_c \text{ is sitting on a time bomb in } w_c \text{ at } t_c \}.
\]
So our theory implies that David Kaplan – if he seriously and honestly uttered sentence (2) at the time indicated – ascribed himself the property of sitting on a time bomb at that time as the (then) present time. This is in accordance with the above considerations.

For a sentence like (3) we also get the desired result:

(3) It is now 4 a.m. on July 20, 1978.

We take this sentence to be composed of the sentential adverb ”now” and the sentence ”be 4 a.m. on July 20, 1978”, where:

\[ \ll \text{be 4 a.m. on July 20, 1978} \mathbin{\ll (c) \ (i) = 1} \text{iff} \]
\[ t_c = 4 \text{ a.m. on July 20, 1978}. \]

Then the diagonal of (3) is precisely what we want in view of the above discussion of the O'Leary case:

\[
\tilde{\partial}(\text{now} (\text{be 4 a.m. on July 20, 1978})) =
\{ c \mid \ll \text{now} (\text{be 4 a.m. on July 20, 1978}) \mathbin{\ll (c) \ (i(c)) = 1} \} =
\{ <w_c, s_c, t_c> \mid \ll \text{be 4 a.m. on July 20, 1978}(<w_c, t_c>) = 1 \}
= \{ <w_c, s_c, t_c> \mid t_c = 4 \text{ a.m. on July 20, 1978} \}. \]

That is, at the time in question O'Leary believes that his then present time is 4 a.m. on July 20, 1978.

2.4 The Place Parameter

In this section we are going to consider utterances containing the expression ”here” and contemplate a further parameter, the place parameter \( p \). For although it has never been discussed under the epistemological perspective taken here, it has always been a plausible candidate for a contextual parameter. So is it right to construct the elements of belief sets and hence contexts as quadruples \( <w, s, t, p> \), where \( p \) represents the place of utterance? And does it make sense to take indices to be triples \( <w, t, p> \) containing, apart from world and time, a place parameter?

It is clear that in utterances with ”here”, too, subjective and objective meaning diverge: While O'Leary is sleeping in the boot, his car is stolen. The thieves go to their boss’s house, open the boot, blindfold O'Leary and take him to the garden-house. The boss is Daniels, a man O’Leary knows and takes to be a good friend. Daniels had O'Leary kidnapped to hold him to ransom from his family. Sitting alone in the empty garden-house, O'Leary thinks: ”How uncomfortable it is here.” But since he would not have suspected Daniels to be his kidnapper, he certainly does not believe: ”It is uncomfortable in Daniels’s garden-house.” However, under the assumption that ”here” always refers to
the place of utterance, which in this case is the interior of Daniels’s garden-house, both utterances have the same objective meaning.

Of course, the fact that the subjective meanings of utterances with ”here” do not necessarily coincide with their subjective meanings is not reason enough to give the elements of belief sets a finer structure by adding a parameter \( p \) for the place of utterance. Indeed, it is easily seen that there is no such cogent argument as in the case of the subject and time parameters. For in the case of ”I” and ”now” we had seen that no definite descriptions can be found which generally reflect the subjective meaning of these expressions. But in the case of ”here” such a description is easily found; the subjective meaning of ”here” can simply be analyzed as ”the place at which I am now”, so that \( p \) can be defined to be ”the place at which \( s \) is at \( t \) in \( w \”).

Introducing the place of utterance into the context as a further parameter is therefore not absolutely necessary. There still are at least two reasons to do nevertheless so.

The first reason is the indeterminacy of the description specifying the place of utterance. For, strictly speaking, there are many places at which the speaker is at the time of utterance; so ”the place at which \( s \) is in \( w \) at \( t \)” is not a proper description. Let us reconsider the O’Leary case: What precisely would be the place in which O’Leary makes his relevant utterance of ”How uncomfortable it is here.”? Is it the region occupied by O’Leary at \( t \) in \( w \), or by the garden-house, or by Daniels’s entire premises? If one wants ”here” to always refer to the place of utterance, then – as in the case of ”here” – one must not insist that the place of utterance always coincides with the speakers spatial extension; rather, it may be a more or less inclusive environment around the speaker. This indeterminacy only disappears if it is clear which kind of spatial region is supposed to count as the place of utterance. This might be a small part of the body – ”here it hurts” – or a whole country – ”one can drive as fast as one wants here”. Which kind of region is at stake can only be gathered from further features of the context, among which the speaker’s intentions are not the least important. This context dependence, however, is presumably very hard to capture in terms of general rules. But if contexts contained a separate place parameter – if, in other words, they are taken to be quadruples \(< w_c, s_c, t_c, p_c >\) such that \( s_c \) exists in \( w_c \) at \( t_c \) and the region occupied by \( s_c \) in \( w_c \) at \( t_c \) is part of \( p_c \) – then this problem disappears, as in the case of the utterance time; for then the intended place must be explicitly noted as this place parameter.

This problem must not be confused with the additional problem of the vagueness of the place of utterance. The boundaries of the body part at which it hurts, like the borders of the German Federal Republic, are indeed vague – and so is practically every local designator. But the place parameter is not supposed to dissolve that vagueness; for then one could rightly object that vagueness is such a general problem that it would not make much sense to give these special cases a special treatment. The place parameter should rather solve the prior question of whether ”here” just denotes this part of the body, this room, this building, or even this country, etc.; and that is not a matter of vagueness.

A second reason may arise if indices contained a place parameter. For it seems to be a generally accepted principle that all index parameters are also parameters of the context (cf. Zimmermann 1991, p. 169). One motivation for this principle derives from the fact that the truth definition for utterances \(< \phi, c >\) makes reference to \( i(c) \), the index of context
and if all parameters of the index are also contextual parameters, \( i(c) \) can simply be
defined as the sequence of index parameters of \( c \) as occupied by the corresponding
contextual parameters of \( c \). It must, however, be admitted that there would be no
substantial problem of defining \( i(c) \) even in case only \( i \) but not \( c \), contained a place
parameter \( p \).

We are thus led to the question of whether indices should contain a place parameter.
In section 1.2 (p. VVV26VVV) we had said that all parameters that are shifted by
operators go into the index. One of our examples was "everywhere" which we took to be
analyzable as an operator that shifts the place parameter. We said that a sentence like "It
is raining everywhere" is to be analyzed like this:

\[
\| \text{everywhere} \| (c) (\langle w_i, t_i, p_i \rangle) = 1 \text{ iff } \\
\| \text{be raining} \| (c) (\langle w_i, t_i, p \rangle) = 1 \text{ for all places } p
\]

We thereby interpreted "everywhere" in analogy to the temporal operator "always".
If indices (and contexts) contained a place parameter, we could equally well interpret
"here" in analogy with "now":

\[
\| \text{here} \| (c) (i) = 1 \text{ iff } \| \phi \| (c) (\langle w_i, t_i, p_i \rangle) = 1.
\]

All other adverbials denoting a place could then be treated in the same way as "here";
they, too, occupy the parameter of evaluation \( p \) with the place they denote (in the respective
context).

However, these analyses presuppose that predicates, or at least those that can be
combined with locatives, are assigned their extensions relatively to places as well as
worlds and times. We would then have semantic rules of the following kind:

\[
\| \text{raining} \| (c) (i) = \{ \emptyset \mid \text{it is raining at } p_i \at t_i \}, \\
\| \text{sleeeping} \| (c) (i) = \{ x \mid \text{x is sleeping at } p_i \at t_i \}, \\
\| \text{seeing} \| (c) (i) = \{ \langle x, y \rangle \mid \text{in } p_i \text{ sees } y \at t_i \}
\]

But now problems arise. Let us take a closer look at the extension of "seeing". What
is it supposed to mean that \( x \) sees \( y \) in \( p_i \)? Do \( x \) and \( y \) both have to be in \( p_i \)? Intuitively,
a sentence like "Jack saw Peter in the market-place" may both mean that Peter was in the
market-place while Jack saw him – e.g., from the balcony of one of the surrounding
buildings – or that Jack was in the market-place when he saw Peter. This indicates that
locatives do not modify entire sentence meanings but rather localize arguments of the
verbs (see also Herweg/Wunderlich 1991 on this point). The only exceptions would be
sentences that are analyzed in terms of 0-ary predicates like "raining"; since such
predicates do not have any arguments to be localized, there is no other choice than treating
locatives in such sentences as sentential adverbs.

Another problem, pointing in the same direction, arises from the truth definition for
utterances which, as we have seen, reads: An utterance \( \langle \phi, c \rangle \) is true just in case \( \phi \) is true
at \( c \) and \( i(c) \) (cf. p. VVV28VVV). The upshot of this definition is that all unoccupied or
bound parameters of the index are assigned the values of the corresponding contextual parameters. But this means that, e.g., the sentence "it is now raining" has the same truth conditions in a context \( c \) as the sentence "it is now raining here", because we have:

\[
\| \text{now (be raining)} \| (c) (i(c)) = 1 \text{ iff } \\
\| \text{be raining} \| (c) (\langle w_c, t_c, p_c \rangle) = 1 \text{ iff } \\
\| \text{here (now (be raining))} \| (c) (i(c)) = 1.
\]

This result is certainly intuitively adequate. But if we follow the above rules and interpret all predicates as locally dependent, we get the same result even for sentences like "Jack is sleeping" of "Jack sees Peter". This, however, is intuitively inadequate: in a context \( c \), "Jack is sleeping" does not mean the same as "Jack is sleeping here", nor does "Jack sees Peter" mean the same as "Jack sees Peter here"; "Jack is sleeping" at best means as much as "Jack is sleeping somewhere (viz., wherever he is)".

Given this reasoning, we should take it that only some, nullary predicates like "raining", "snowing", etc. have locally dependent extensions. All other predicates should only be interpreted with respect to time and world. But that means that only in connection with such nullary predicates can locatives be analyzed as sentential adverbs occupying or quantifying the place parameter. In combination with verbs like "sleeping" or "seeing", on the other hand, they work more like predicates or, as the case may be, quantifiers. Thus the sentence "Jack is sleeping here" would have to be analyzed as "Jack is sleeping and Jack is here", and "Jack sees Peter in the market-place" as "Jack sees Peter and Jack is in the market-place" or as "Jack sees Peter and Peter is in the market-place"; and a sentence like "Jack knows a good restaurant everywhere" as: "For every place \( p \) it holds: there is an \( x \): Jack knows \( x \) and \( x \) is a good restaurant and \( x \) is in \( p \)."

Finally, it must be noted that locatives and local quantifiers can even function as arguments to verbs; thus, e.g., "live", "dwell", or "be" may express binary relations between objects and places. In sentences like "Jack lives here" or "The devil is everywhere", "here" and "everywhere" cannot be treated as sentential adverbs anyway.

Von Stechow (p.c.) mentioned that a heterogeneous treatment of locatives could be avoided if verbs like "raining" were analyzed in analogy to verbs like "sleeping", i.e., if "raining" would not be treated as a nullary predicate but as a unary one and thus as a personal construction; "it is raining" is then interpreted as "there is an \( x \): \( x \) is a rain". One would then no longer have to take "here" and "everywhere" as sentential adverbs but would rather interpret "it is raining here" or "it is raining everywhere" as "there is an \( x \): \( x \) is a rain and \( x \) is here" and as "for every place \( p \) it holds: there is an \( x \): \( x \) is a rain and \( x \) is in \( p \)"., respectively. "Raining" must then no longer be taken as a locally dependent predicate and the motivation for a place parameter would go away. However, there is a problem with this approach: "it is raining" would have to be understood as "it is raining somewhere", not as "it is raining here".

We may thus note that, although we have found reasons for having place parameters in context and index, we have not found any decisive arguments. Of course, we cannot exclude that a more comprehensive analysis of locatives revealed even better arguments in one or the other direction. But we do not want to elaborate the matter here. The aim of this
section was just to show that our theory of subjective meaning is consistent with either possibility. At the end, we want to look at some examples.

Let us first consider O’Leary’s utterance of (4):

(4) It is uncomfortable here.

I am assuming that in (4) ”here” is an argument of ”being uncomfortable”, i.e.: \( \parallel (c)(i) = \{ p | p \text{ is a place and in } w_i \text{ it is uncomfortable in } p \text{ at } t_i \} \). (4) thus receives the following truth-conditions:

\[
\parallel \text{being uncomfortable (here)} \parallel (c)(i) = 1 \iff \parallel \text{here} \parallel (c)(i) \in \parallel \text{being uncomfortable} \parallel (c)(i) = 1
\]

in \( w_i \) it is uncomfortable in \( p_c \) at \( t_i \)

The subjective meaning of (4) is the diagonal of (4). We have:

\[
\partial( \text{it is uncomfortable here} ) =
\{ c \parallel \text{being uncomfortable here} \parallel (c)(i(c)) = 1 \} =
\{ <w_c,s_c,t_c,p_c> \mid \text{in } w_k \text{ it is uncomfortable in } p_c \text{ at } t_c \}.
\]

Here we must take belief sets to be sets of quadruples \(<w,s,t,p>\) in the sense that a quadruple \(<w,s,t,p>\) is an element of a person \( a \)'s belief set in a world \( w^a \) at a time \( t^a \) and a place \( p^a \) if, and only if, in \( w^a \), \( a \) considers it possible at \( t^a \) that \( w \) is the real world, \( t \) the present time, he the person \( s \) at \( t \) in \( w \) and \( p \) the place in which he is at \( t \) in \( w \). If O’Leary believed nothing but \( \partial(4) \), this would then mean that he would take himself to be a person who is presently located at an uncomfortable place. This result is intuitively welcome.

But then practically the same result would have been arrived at if we continued to take the elements of belief contents to be just triples \(<w,s,t>\) and defined ”here” as ”the place at which \( s \) is at \( t \) in \( w \)”. For \( \partial(4) \) would then be:

\[
\{ <w_c,s_c,t_c,p_c> \mid \text{in } w_c \text{ it is uncomfortable in } p_c \text{ at } t_c \}.
\]

and to believe that again means to ascribe to oneself the property of presently being located at an uncomfortable place.

Our second example is sentence (5):

(5) It is raining here.

If we analyze ”raining” as a locally dependent, nullary predicate, \( \partial(5) \) is:

\[
\{ c \parallel \text{raining here} \parallel (c)(i(c)) = 1 \} =
\{ <w_c,s_c,t_c,p_c> \mid \text{in } w_c \text{ it is raining in } p_c \text{ at } t_c \}.
\]
if we treat "raining" as unary predicate – as sketched above – then $\partial(5)$ is:

$$\{ c \mid \exists x: \text{x is a rain and x is here} \mid (c) (i(c)) = 1 \} =$$

$$\{ <w_c, s_c, t_c, p_c> \mid \text{in } w_c \text{ there is a rain in } p_c \text{ at } t_c \}.$$  

The two results are identical and say that whoever believes (5) self-ascribes the property of presently being located at a place at which it is raining. We will spare ourselves the variants that arise from introducing the place of utterance by definition.

This is all I want to say about locatives; as I said, the point was to demonstrate that our theory of subjective meaning does not meet any adversities in this area. Since the remainder of this work will not even indirectly be concerned with locatives, we will from now on ignore the place parameter in both context and index.
Chapter 3

Predicates

According to the view developed so far, the subjective content of an utterance has two sources: first, the speaker’s specifically semantic knowledge which consists in his knowledge of the character function of the language in question and, second, further empirical assumptions made by the subject and collected in his belief set. This view has proved fruitful in the analysis of indexicals like "I", "now", and "here". I will now try to investigate how it carries over to other expressions.

First of all, I will be concerned with natural kind terms, concentrating on the by now classical "water" case. Following Putnam’s ideas, I will present the character of "water" in section 3.1 and then, in section 3.2, argue that this character must be understood as variable. It thus turns out that we can indeed construe Putnam’s notion of hidden indexicality as indexicality in the above sense. In doing so we are going to slightly stretch the classical concept of a character, which will put us in a position to apply the method of diagonalization to the problem of informativity for utterances containing natural kind terms.

In section 3.3 I will discuss the question of identifying the domain of the character of natural kind terms. This is a non-trivial question and only after answering it will my concept of a character be fully explained. In section 3.4 I will then discuss an obvious objection against the kind of semantic analysis proposed here, one that is usually taken to be cogent; I will, however, be able to show that this objection does not affect me.

The objective of section 3.5 is to generalize the results on the semantics of natural kind terms to arbitrary predicates. In section 3.6, the very abstract scheme resulting from these efforts will be further illustrated, mainly by functional and empty predicates.

In section 3.7 I am going to clear up how the concepts of hidden indexicality and rigid designation relate to each other, thereby correcting the wide-spread view that they are but two sides of a coin.

However, in section 3.8 it will turn out that the methods developed thus far still will not put us in a position to solve the problem of internalism as applied to predicates. To account for it yet another re-interpretation of the concept a of character will be necessary, and with it comes a new conception of the relation between semantic and empirical knowledge. These new views, as well as the concept of subjective meaning resulting from it, will be discussed in the final section 3.9.

3.1 The Character of "Water"

In "The Meaning of ‘Meaning’", Putnam was concerned with showing that meanings are not in the head: that the extensions of many linguistic expressions are not
determined by narrow psychological states and that meanings thus cannot be identified with the beliefs that competent language users hold. In section 1.3 we have already presented Putnam’s line of reasoning. It focussed on the case of Twin Earth, where there is no water but only different substance, which is also called "water" and, in spite of having the same surface qualities as water, differs in its chemical structure which is XYZ, not H$_2$O.

Let us now take a closer look at Putnam’s positive reflections on the meaning of "water". If the meaning of "water" cannot be equated with the properties that competent speakers associate with "water", how should it then be accounted for? Given our assumption that the basic concept of the theory of meaning is that of a character, we are thus aiming at a definition of the character of "water". Which conditions would such a definition have to satisfy, according to Putnam?

We must bear Putnam’s claims on the extension of "water" in mind: the world $w^*$, which we imagined to be actual, not only contained Earth and the speakers of English but also Twin Earth and speakers of Twin English; on Earth, the word "water" designated the set of all bodies of liquid with chemical structure H$_2$O, whereas on Twin Earth it denoted the set of all liquid portions with chemical structure XYZ. In the sense in which "water" is used in English on Earth – the sense of "water$_E$" – whatever the Twin English speakers call "water" is not water, and in the sense in which "water" is used on Twin Earth – the sense of "water$_T$" – whatever we call "water" on Earth is not water; "water$_E$" and "water$_T$" have different extensions. This also applies to the past when neither on Earth nor on Twin Earth chemistry was far enough developed to distinguish H$_2$O from XYZ; even in 1750 the extension of "water$_E$" already consisted of H$_2$O-bodies and that of "water$_T$" of XYZ-bodies (cf. Putnam 1975, pp. 223–5).

The fact that"water$_E$" and "water$_T$" have different extensions in $w^*$, of course, implies that the two words do not have the same intension; their intensions – regarded as functions determining extensions$^1$ – yielding two distinct values on at least one argument, viz. $w^*$. But Putnam makes even stronger claims about the intension of "water", namely that it is in some sense rigid: in other possible worlds "water" only denotes such objects which in all essential respects resemble whatever lies in the extension of "water" in the actual world. If it is actually the case that the bodies of liquid in the extension of "water$_E$" in $w^*$ are of molecular structure H$_2$O, then only such worlds are possible in which water$_E$ is H$_2$O, too. In other words: If being H$_2$O is an essential property of water$_E$, then the intension of "Water$_E$ is H$_2$O" is the set of all possible worlds, "Water$_E$ is H$_2$O" thus expressing a metaphysically necessary truth. Similarly for water$_T$ and XYZ; it is thus plainly impossible that water$_E$ be water$_T$ (cf. Putnam 1975, pp. 231–3).

What about the character of "water"? In "The Meaning of 'Meaning'", Putnam himself does not talk about the characters of linguistic expressions, and he makes no

---

$^1$ We always use "intension" in the sense defined above, according to which intensions are functions from possible worlds, or possible indices, to extensions. This is not the sense in which Putnam uses "intension"; he uses the term in a more vague, pre-theoretic sense that may be paraphrased as "those aspects of linguistic meaning that are beyond extension".
effort to describe anything like it. This is not too surprising, given that Putnam
developed his thoughts before Kaplan’s theory of character has been fully formulated.
Nevertheless we may construe some of Putnam’s claims on the semantic behaviour of
“water” as claims on the character of “water”.

Putnam explicitly speaks of the indexicality of “water” and other natural kind terms,
though he thereby does not use “indexicality” in precisely the technical sense defined
above:

“Our theory can be summarized as saying that words like ‘water’ have an unnoticed indexical
component: ‘water’ is stuff that bears a certain similarity relation to the water around here.
Water at another time or in another place or even in another possible world has to bear the
relation same to our ‘water’ in order to be water.” (Putnam 1975, S. 234; italics as in the
original.)

However, Putnam does want to make a difference between overtly indexical
expressions like ”I”, ”now”, ”this”, and the covertly indexical natural kind terms:

“The theory that natural-kind words like ‘water’ are indexical leaves it open, however, whether
to say that ‘water’ in the Twin Earth dialect of English has the same meaning as ‘water’ in the
Earth dialect and a different extension (which is what we normally say about ‘I’ in different
idiolects), thereby giving up the doctrine that ‘meaning’ (intension) determines ‘extension’; or to
say, as we have chosen to do, that difference in extension is ipso facto a difference in meaning
for natural-kind words, thereby giving up the doctrine that meanings are concepts, or, indeed,
mental entities of any kind.” (Putnam 1975, p. 234; all italics as in the original.

This passage – among other remarks, e.g. on pp. 233f. or p. 266 – makes clear that
Putnam would account for the context-dependence of overt indexicals like ”I” in terms
of an extended notion of index. However, as we have seen in section 1.2 (cf. pp.
VVV25ff.VVV), this procedure leads to undesired results. We should therefore think
about how Putnam’s remarks can be reconstructed on the background of the distinctions
between index and context and between character and intension. What Putnam was
after in the above passage would then presumably have to be reformulated like this:

“The theory that natural-kind words like ‘water’ are indexical leaves it open, however, whether
to say that ‘water’ in the Twin Earth dialect of English has the same character as ‘water’ in the
Earth dialect and a different extension and intension (which is what we normally say about ‘I’ in
different idiolects). (One could then advocate the further thesis that the meaning, i.e., the
character, of a natural-kind term determines its extension and that meanings are mental entities –
just like one says about overt indexicals like ‘I’ that their character, as depending on context and
index, determines extension and that the character describes whatever competent speakers know
about the meaning of ‘I’.) Or one could say that a difference in extension is ipso facto a
difference in character for natural-kind words, thereby giving up the doctrine that meanings
(characters) are mental entities of any kind.”

Let us take a closer look at these two possiilites of accounting for the indexicality of
”water”.
The gist of the first proposal seems to be to regard ”water_E” and ”water_T” as the
same word and English and Twin English the same language; the subscripts would then
only have the function of keeping different occurrences of ”water” apart – as one marks
the two occurrences of "this" in "I prefer this₁ to this₂" by subcripting. But what should the character of the English-Twin English word "water" exactly look like? Which feature of the utterance context would the intension of "water" depend on?

One may be tempted to say that it is the place of utterance: If "water" is uttered on Earth, then in every possible world it denotes those bodies of liquid whose chemical structure is H₂O; if it is uttered on Twin Earth, then in every possible world it denotes those bodies of liquid whose chemical structure is XYZ; and so forth. This idea could be read into the italicization of "around here" in the first quotation. But, clearly, it is no good. For according to Putnam, the earthling who goes to Twin Earth and says, pointing to the liquid he sees in a lake: "This is water." is mistaken. Making this utterance, he says something false because, as an earthling, by "water" he can only refer to things that resemble Earthian water in their essential qualities. Furthermore, the idea would have the absurd consequence that a sentence like "there is water in this vessel" could be truly said of on earth, though an utterance of the same sentence, with the same reference of "this vessel" would be false if the vessel were transported to Twin Earth. "Water" just does not work like "here".

The upshot of these reflections is that the contextual parameter with which the intension of "water" varied would have to be the language community: The extension of "water" at context c and index i is the set of all bodies of liquid whose chemical structure is H₂O, if the speaker in c belongs to the English-speaking community, it would be the set of all bodies of liquid whose chemical structure is XYZ, if the speaker in c belongs to the Twin English-speaking community, and so forth. This, then, would make the sentence "this is water" false if earthling Oscar utters it on Twin Earth, referring to the liquid in the lake; and it would render "there is water in this vessel" true if Oscar uttered it on Twin Earth about the vessel he has brought with him. However if, under the same circumstances, the same sentences are uttered by Twin Oscar, they would get the opposite truth values.

This view, which one might read into the italicisation of "our" in the first quotation, however, leads to absurd consequences too. Putnam makes this clear by way of a parallel and obviously counterintuitive example: Let us imagine that the word "beech" on Twin Earth referred to elms, whereas the word "elm" referred to beeches; and let us further assume that the properties I associate with "elm" or "beech" are the same – because all I know about elms and beeches is that they are broad-leaved trees – and that the same is true of my Twin Earth doppelganger. In that case one still would not want to say that my English word "elm" had the same meaning as his Twin English word "elm", viz. to designate elms when uttered by a speaker of English, and beeches when uttered by a speaker of Twin English. One would rather say that the Twin English word "elm" had the same meaning as the English word "beech" and that the Twin English word "beech" had the same meaning as the English word "elm".

What is obvious in this case also applies to our water-example. It is, as it were, pure coincidence that English and Twin English have the same word for these different liquids. This becomes particularly apparent when we translate Twin English sentences into English. Thus, if Twin Oscar says: "Water quenches thirst", he must, of course, not
translate this as "Water quenches thirst", but at best by something like "Twin water quenches thirst".

It should have been clear right from the outset that this – tentative – view is rather strange. For what is it supposed to mean that the language is kept fixed when at the same time the language community is allowed to vary as a contextual parameter? If, our intuition seems to say, a word has different meanings in two distinct linguistic communities but not in otherwise distinguishable contexts, then *ipso facto* these communities speak two different languages. Language and language community cannot be so easily separated; "water" does not work like "we" when referring to the language community.

One point is still in need of some clarification. What do we mean by one word as opposed to two distinct words? By phono-syntactic standards, "water_E" and "water_T" can certainly be regarded as one and the same word; after all, throughout our examples we have been assuming that English and Twin English are identical in these respects. However, if we were to use this identity criterion for words, our analysis would end up describing only possible meanings for purely phono-syntactically individuated words, and not the meanings of the words of a given language. And we would thereby have admitted that we failed to understand "water_E" and "water_T" in the same sense, i.e., assign it the same character, thus proving it to be the same semantic word. In that sense we must, at any rate, regard them as different words. But we will return to the difference between a purely phono-syntactic as opposed to a semantic characterisation of words and languages in section 3.9.

Let us now consider the second way of spelling out the indexicality of "water", by assuming that an extensional difference between natural kind terms already causes a difference in meaning. What precisely could the characters of "water_E" and "water_T" then look like? One thing we would have to take care of is that there be a context \( c \) such that \( \| \) water_E \( \| (c) \neq \| \) water_T \( \| (c) \). Given the above considerations we know, however, that the intensions of "water_E" and "water_T" must not depend on the context in a narrow sense, i.e., on the speaker or his language community, or the place of utterance; and that it does not depend on the time of utterance was plain obvious, needing no further discussion. So it could only vary with the world of context. For Earth and Twin Earth coexist in the same possible world \( w^{\#} \), but in any possible world "water_E" denotes \( H_2O \), whereas "water_T" denotes XYZ. How, then, shall we formulate the general rules determining the characters of "water_E" and "water_T"? Let us, for this purpose, look up Putnam again. There we find the following explanation of the meaning of "water" and its indexicality:

"[A]n entity \( x \), in an arbitrary possible world, is \textit{water} if and only if it bears the relation same_L [... to the stuff we call ‘water’ in the actual world." (Putnam 1975, p. 232; all italics as in the original.)

Neither could we say that "water" is ambiguous in English-Twin English. For this would imply that speakers of English-Twin English have the freedom of meaning this or that by "water". But they do not; they are bound by their respective linguistic communities.
Putnam does not intend this to be a definition of meaning, i.e., he would not say "water" is synonymous with "whatever is called ‘water’ in the actual world". For him this is just a way of defining meaning by an ostensive definition, just as one may determine the referent of a name by a definite description without thereby rendering the name synonymous with the description. In section 3.4 we will give a detailed account of Putnam’s reasons for this restrictive view. But let us first see how far we get if we still took the above quotation as our starting point for a definition of the character of "water". First of all, we may reformulate it in the following way:

\[ \ll \text{water} \rr (c) (i) = \{ x \in w_i \mid x \text{ bears same}_L \text{ to whatever we call "water" in } w_c \}. \]

We will essentially stick to this definition. But it is in need of various refinements to be discussed in this and the next few sections.³

Let us first be clear about which of the words "water" it is that we are going to describe, the English one or the Twin English one; and let us also make explicit who is supposed to be "we". After all, we do not want to understand the definition as having "we" simply refer to some group surrounding the speaker in c; for it has become clear that the intension must not vary with the speaker. Rather, we prefer "we" to refer to the English language community as long as we are concerned with the English word "water", and to the Twin English community when we are dealing with the Twin English word "water", etc.

Next, we should at least give a preliminary specification of the locution "calling something ‘water’". Of course, the definition is not intended to convey that \( x \) is water as soon as it bears same\(_L\) to something that some English speakers have once, possibly erroneously, called "water"; rather, we are relating to typical or paradigmatic cases. We are thus led to the following reformulation:

\[ \ll \text{water}_E \rr (c) (i) = \{ x \in w_i \mid x \text{ consists of the same liquid as that to which English ’} \text{”water” is typically applied in } w_c \}, \text{ or, respectively:} \]

\[ \ll \text{water}_T \rr (c) (i) = \{ x \in w_i \mid x \text{ consists of the same liquid as that to which Twin English ’} \text{”water” is typically applied in } w_c \}. \]

A more precise account of what it means to "typically apply" a natural kind predicate would more or less amount to a rehearsal of Putnam’s repeated utterances to the effect that the determination of extension should be investigated by socio-linguistic means, taking the linguistic division of labour into account. Hence the so-called causal theory of reference would have to be discussed in detail. But then historical tradition, a word’s causal pre-history of use, does not by itself determine its reference; causal

³ We will ignore the fact that Putnam has chosen a mass noun, "water", as his example; this would only distract from the main point. We are treating "water" exclusively as a predicate applying to single portions of water.
chains must always pass the filter of linguistic convention. Consequently, the theory of linguistic convention would have to be dealt with, too. Finally, since linguistic conventions can hardly be separated from the individual speakers, they too will play a role.

A lot has been, instructively, said about all this. Starting with Grice (1957), speakers’ intentions have been subject to extensive analysis (see, in particular, Bennett 1976). Starting with Austin (1962), linguistic conventions have been accounted for within speech act theory and theoretically analysed by Lewis (1969). And, starting with Kripke (1972) and Putnam (1975a, ch. 11-13), the causal theory of reference has been developed in great detail, most notably by Devitt (1981).

But, luckily, we need not enter into all this in any depth. We will just continue to speak of being-called-such-and-such-in-a-language or, e.g., of the English usage of "water" or, more extensively, of the causal-intentional complex of the English usage of "water"; we thereby always refer to the totality of empirical facts that are responsible for something being called such-and-such in a language – e.g., "water" in English – and that are further investigated in the literature just mentioned. Of course, this only makes sense if a language like English or Twin English is not conceived of as an abstract system of symbols but as a concrete historical structure comprising all these causal-intentional complexes pertaining to individual words.

To avoid any confusion, we have to keep two distinct linguistic tasks strictly apart. It is one thing to develop a recursive semantics, a recursive theory of subjective and objective meaning consisting in a specification of the characters of the expressions of a language – provided we have chosen the right approach here; this is the task we have decided to undertake. It is another thing to clarify what the relation of reference empirically consists in, which facts must hold for something to be called this or that in a language; the literature mentioned is concerned with this empirical theory of reference. For if the above definition of the character of "water" is on the right track, fulfilling the second task will be part of the first one, the definition of the character function; this is what is behind the locution "typically applied in English"). But that does not mean that we must carry out the second task here and now; for our theory, it is only crucial to eventually arrive at an adequate understanding of this connection between recursive semantics and the empirical theory of reference.

But let us first see what the above definition yields with regard to the extension and intension of "water". It is exactly what we want: If \( w_c = w^* \) and if the liquid that we typically call "water" in English is actually of molecular structure \( \text{H}_2\text{O} \), then in every possible world (including \( w^* \)), "water\(_E\)" as uttered in \( w^* \) denotes the set of all bodies of liquid \( x \) with molecular structure \( \text{H}_2\text{O} \). And if, in \( w^* \), the liquid that our twins typically call "water" has molecular structure \( \text{XYZ} \) in \( w^* \), then in every possible world (including \( w^* \)) "water\(_T\)" as uttered in \( w^* \) denotes the set of all bodies of liquid \( x \) with molecular structure \( \text{XYZ} \). So when Oscar goes to Twin Earth and says "This is water" about the liquid \( x \) in a lake, he speaks falsely, because:

\[
\ll \text{this is water}_E \ll (c)(i) = 1 \text{ iff } x \in \ll \text{water}_E \ll (c)(i) \text{ iff }
\]
3. Predicates

\[ x \in \{ y \mid \text{in} \ w_i, \ x \ \text{consists of the same liquid as that to which English } \text{"water" is typically applied in } w_c \} \ \text{iff} \]
\[ x \ \text{is of molecular structure } H_2O \ \text{in } w_i. \]

By the same token, we find that "Water is H\textsubscript{2}O" as uttered in \( w_c = w^* \) expresses a necessary truth; for we have:

\[ \ll \text{water}_E \ \text{is } H_2O \ \ll (c) (i) = 1 \ \text{iff} \]
\[ \{ x \mid \text{in } w_i, \ x \ \text{consists of the same liquid as that to which English } \text{"water" is typically applied in } w_c \} \subseteq \{ x \mid x \ \text{is of molecular structure } H_2O \ \text{in } w_i. \} \ \text{iff} \]
\[ \{ x \mid \text{in } w_i, \ x \ \text{consists of a liquid of molecular structure } H_2O \} \subseteq \{ x \mid x \ \text{is of molecular structure } H_2O \ \text{in } w_i. \}, \]

and this relation holds in all worlds \( w_i \).

Finally, the definition expresses Putnam’s intuition that "water" is rigid. We have not yet discussed what precisely rigidity as applied to predicates could be and whether it does play a crucial role here; we will only do so in section 3.7, when the conclusions drawn from the "water" case are to be generalised. But we can glean from the above definition that, in a possible index world \( w_i \), "water\textsubscript{E}" only denotes such objects which in their essential properties resemble whatever is water in the context world \( w_c \); and this is in accordance with Putnam’s explanation of the rigidity of predicates.

3.2 In What Way is "Water" a Hidden Indexical?

The crucial question for what follows is whether the above characters of "water\textsubscript{E}" and "water\textsubscript{T}" can help solving the problem of informativity in our way, i.e., employing the method of diagonalization. This question has so far not been answered, and it is not entirely straightforward to do so. For we know that in general the method of diagonalization fails on absolute characters; if the intension of "water\textsubscript{E}"”, like that of "H\textsubscript{2}O”, does not vary with context, then "Water\textsubscript{E} is H\textsubscript{2}O” not only expresses the necessary proposition in every context; its diagonal would be tautologous too, and hence not informative. Hence an informative diagonal only results if "water\textsubscript{E}” turns out to be indexical. But the above definition itself does not tell us whether "water\textsubscript{E}" is absolute or indexical; it can be construed either way. So we must specify it further.

As has just been explained, by the typical application of English "water", or the English usage of "water", we mean the causal-intentional complex underlying the reference of "water" in English. So the German usage of "water" only exists in possible worlds in which this causal-intentional complex exists. The question, then, is how this complex – which is only one of a countless number that make up the English language as a whole – can be isolated.
Asking this question wholesale, one gets entangled in irreconcilable vagueness. As a case in point, all actual, past and present speakers of English partake in this complex; through all these persons the usage of "water_E" has been established and reproduced. But no single English speaker's existence is essential for the existence of the English usage of "water". Even if, at the time of the Great Plague, the surviving half of the population had deceased and the deceased half had survived, the existence of the English usage of "water" would not have been affected. If, on the other hand, Europe had been completely depopulated in those days, historical continuity would have been sufficiently interrupted to block the existence of the English usage of "water". There remains, however, a large scope of vagueness between these two clear cases.

A different kind of vagueness can be found in the typical applications of "water_E". For instance, the Tyne is certainly an important English river. And "water_E" has typically been applied to the liquid in the Tyne; we may, indeed, assume that it is one part of the causal origin of the wide-spread English usage of "water_E". But then it is not an essential part. All of English history, right from its very beginnings, could have taken place somewhere further south, so that English native speakers would at best have seen the Tyne only occasionally. Or maybe, some time in the 18th century, all inhabitants of the Tyne and Wear county, without exception, could have left their homes, taken a boat and invaded Esbjerg in Denmark. Then maybe the Varde would have become an English river; at any rate, some of the important typical applications of "water_E" would have to be replaced by others. But, even after the invasion, the causal-intentional complex of the English usage of "water" would have remained the same.

We cannot – and do not want to – get rid of all this vagueness in the identity conditions of the English usage of "water". But then, for the problem of information only one aspect is important, which we will have to clear up now; it is part of the second kind of vagueness that we have just addressed. For the question is whether the causal origin with which the English usage of "water" commences, i.e., all these typical portions of water, shall be taken as parts of the causal-intentional complex.

If so, the character of "water_E" as specified above turns out to be absolute. For let us assume that the typical portions of water in the actual context world \( w_c \) consisted of \( \text{H}_2\text{O} \) and imagine another possible context world \( w_{c'} \) in which there is only XYZ but no \( \text{H}_2\text{O} \). Since the typical portions of water are \( \text{H}_2\text{O} \), they do not exist in \( w_{c'} \). However, neither is there, then, an English usage of "water" in \( w_{c'} \) that they could be part of. So, not only does "water_E" denote \( \text{H}_2\text{O} \) in all index worlds \( w_i \), if it denotes \( \text{H}_2\text{O} \) in the given context world \( w_c \); in the other context worlds, too, it can only denote \( \text{H}_2\text{O} \), or English plain does not exist there. This precisely means the "water_E" has an absolute character.\(^4\)

Moreover, its character is only partial, defined on such context worlds only in which the English usage of "water" and the English language exist at all. But then, according to the above rule, the character
The other possibility is that the typical portions of water are not themselves part of the English usage of "water" and that the complex this usage consists in rests, as it were, only on its causal base. One could then change this causal base from one context world to the next, e.g. from H₂O to XYZ, without thereby giving up the identity of the usage of "water". In this way we gain the freedom needed for a variable character of "water_E"; "water_E" would then be indexical.

Which alternative shall we opt for? I am not sure what Putnam would say here. He does not think in terms of characters and has thus not explicitly asked this question; and his remarks can be taken either way. As regards the analogous case of proper names, Kaplan (1977) takes the firm view that they are absolutely rigid; that is, as soon as the reference of a name – the causal origin of the usage of that name – changes, one gets a new word in a semantic sense.

One reason behind this decision certainly is that both proper names and natural kind terms like "water" are not indexical on the received view; they do not at all resemble the prime cases of indexicality. Burge (1982: pp. 104ff.), too, is impressed by this intuition. But then this intuition does not in the least contradict the indexicality of "water_E" as defined on VVV p. 28 VVV. It is the distinction between overt and hidden indexicality that must be noted here. In our theoretical framework it can be expressed like this: If the intension of an expression varies with the context world \( w_c \), we have a hidden indexical; but if it varies with the other contextual parameters \( s_c, t_c, \) and \( p_c \), it is an overt indexical. The usual sense of indexicality that Burge relates to only covers overt indexicality. If, however, we take "water_E" to be an indexical, it would have to be a hidden indexical, and not an overt one. We thus share this intuition and can even make it explicit; but it does not help us with our decision.

However, we have already presented a decisive argument above: one of the alternatives allows for a successful solution to the problem of informativity along by now familiar lines, whereas the other option not only excludes our solution but also does not even give a hint as to what an alternative might look like.

There is a related argument, having to do with the notion of analyticity. We have said in section 1.2 (cf. VVV p. 31 VVV) that a sentence of a given language is analytic just in case its character yields "true" for any context and any index for which it is defined; we were thus able to account for Kripke’s (1972, p. 264) intuition that a sentence is analytic if it is a priori and necessary. If we now assign absolute characters to both "water_E" and "H₂O", we get the counter-intuitive consequence that the sentence "Water is H₂O" is analytic in German. After all, intuition says that it (or an utterance of it in our world) is only necessary – which we have already accounted for – and, apart from that, its status is unlike that of, say, "Bachelors are unmarried".

In the absence of any other arguments, these reasons strike me as convincing. However, treating "water_E" as a hidden indexical tends to obscure the nature of the English usage of "water". For how are we to understand that this usage remains the

____________________________________________________________________________________

would have to be partial whatever the English usage of "water" might be. Hence this fact is of no concern to us at this point. We will return to it in the next section.
same across different context worlds, even though its reference, its causal origin, changes?

The English usage of "water" remains the same, I think, if all English native speakers’ patterns of reaction towards their environment as well as all their communicative patterns remain the same. This is precisely the case if, in the possible worlds of context, the causal basis of the English usage of "water" changes in a way that is unnoticeable for the entire English language community at the time in question. For if such a change cannot be noticed by any speakers of English, it could obviously not affect their patterns of reaction and communication. If, however, a change from one causal basis to another could in principle be detected by speakers of English, but when they go on talking about the new basis, the new typical cases, in the same way as they used to talk about the old ones, then the English speakers’ patterns of reaction have indeed changed; in the new context world they apply "water" to something noticeably distinct, something that they would not have called "water" in the old context world.

The point might become clearer with the aid of some examples. Let us consider the present English language community, our actual context world \( w_c \) where water is \( H_2O \) plus another world \( w_{c'} \). If \( w_{c'} \) results from \( w_c \) by replacing all \( H_2O \) in \( w_c \) with beer and if in \( w_{c'} \) English people talk about all this beer as "water", then their usage of "water" would quite obviously have changed; for everybody, children and fools included, can distinguish \( H_2O \) from beer, thereby revealing different patterns of reaction in \( w_{c'} \) and \( w_c \). If \( w_{c'} \) results from \( w_c \) by replacing all \( H_2O \) in \( w_c \) with a liquid that tastes, feels, etc. like \( H_2O \), which is, in other words, indistinguishable from \( H_2O \) to the layman but which still behaves differently, say, under electrolysis, then the English usage of "water" in \( w_{c'} \) must still be different from that in \( w_c \); for either the chemists show different reactions by calling the new stuff "water" in \( w_{c'} \), which they would not have done in \( w_c \); or the communicative patterns are different because the chemists still show the same reactions in \( w_{c'} \) but are no longer heard among the English speakers, so that the new stuff, indistinguishable as it is from \( H_2O \) for the vast majority, counts as "water" in \( w_{c'} \) – no matter what the chemists say. Only if the stuff that replaces water in \( w_{c'} \) is indistinguishable for every layman, every chemist, and even every high-energy physicist, can the English usage of "water" in \( w_{c'} \) be the same as that in \( w_c \).

The latter case being possible, it could in that sense still turn out to be the case that water is not \( H_2O \) but this other stuff – just as, 250 years ago, it could have turned out to be the case that water is not \( H_2O \), but Putnam’s XYZ.\(^5\) It is these cases that guarantee that the diagonal of "Water is \( H_2O \)" was, and still is, informative; for some context

\(^5\) Putnam graphically placed Twin English on a Twin Earth in a far-away galaxy of our world, which has led to a certain degree of confusion. This way he manages to give a non-example of the hidden indexicality of "water"; without any intergalactic traveling of English speakers, no language spoken in a faraway galaxy could be English. The intension of "water" in this context world simply consists of \( H_2O \). If, however, we place Putnam’s XYZ-story on our planet Earth in a counterfactual world, then the people there would be the English speakers with the English usage of "water" and the world would be one that proved the indexicality of "water". So the two cases have to be distinguished; the difference will become relevant in another connection (cf. section 3.8).
worlds, like our $w_c$, it will be true, whereas for others, like the last-mentioned $w_c^*$, it will be false. This, then, is our solution to the problem of informativity.

3.3 What is a Possible Context?

Given the above considerations, the following account of the character of "water" in English should be the most transparent one:

$$\ll \text{water}_E \| (c) (i) = \{ x \mid x \in w_i, x \text{ consists of the same liquid as those things that form the causal origin of the English usage of "water" in } w_c \}$$

This formulation brings out clearly that the causal origin of the usage of "water" can vary with the context world and that we only regard the superstructure to this basis as the English usage of "water"; only this superstructure must be kept fixed throughout all context worlds. Though we will continue to say that something is called this or that in a given language, this is now to be understood as short for a more complex locution of the above kind.

In the next section we will discuss conceivable worries against this analysis of the meaning of "water": in order to be able to do this in a sensible way, we must, however, first get clear about what shall be the functional domain of the character of "water" or, in other words, what is meant by a possible context in a meaning rule of the above kind.

Up to now we have been following Kaplan in assuming that any triple $<w_c, s_c, t_c>$ such that $s_c$ exists in $w_c$ at $t_c$ counts as a possible context and that, the character of any $\alpha$ is defined for all possible contexts. We have already pointed out (cf. section 2.3, VVV p. 78 VVV) that we are using quite an abstract notion of context here and that, strictly speaking, our contexts cannot be equated with possible situations of utterance. Let us first explain this point in some more detail.

Kaplan emphasizes that, in order to determine the content of an expression $\alpha$ in a context $c$, one need not take $\alpha$ to be actually uttered in $c$ (cf., e.g., Kaplan 1979, p. 91, or 1989, p. 584). Such an assumption would turn characters into partial functions: one would have to speak of $c$ as a possible context for an expression $\alpha$—viz. iff $\alpha$ is uttered in $c$—and could then only define $\ll \alpha \| (c)$ for those contexts $c$ that are possible for $\alpha$. If one, like Kaplan (1977, p. 91), also assumes that two distinct sentences $\phi$ and $\psi$ can never be uttered in the same context—because $\phi$ and $\psi$ cannot be uttered simultaneously and a change of utterance time means a change of the utterance situation—we find that the functional domains of the characters of distinct sentences are always disjoint. However, his assumption is only plausible if contexts are really taken as utterance situations. If one allows for a more abstract notion of context, according to which context time does not have to coincide with the time the utterance takes (cf.
section 2.3, (VVV p. 78 VVV), one gets the weaker consequence that the functional domains of the characters of two sentences are never identical, because no two distinct sentences $\phi$ and $\psi$ are always uttered in the same possible contexts.

However, in both cases would it be impossible to formulate a sensible notion of logical consequence or even synonymy. Logical truth had been defined as truth in all contexts of every model and, accordingly, a sentence $\phi$ logically implies a sentence $\psi$ if, and only if, in every model $\psi$ is true in all contexts in which $\phi$ is true. And two expressions $\alpha$ and $\beta$ are synonymous in a model if, and only if, they have the same intension in every context of that model. If one now were to define the character of an expression only for those contexts in which it is uttered, then according to the definition just given there could not be any two distinct but synonymous expressions $\alpha$ and $\beta$. Something similar is true of logical consequence: It could never be the case that a sentence $\psi$ is logically implied by a sentence $\phi$ in the sense of the above definition, because there are always contexts in which $\phi$, but not $\psi$ is true – simply because $\psi$ is not uttered in them.

One could, of course, define logical consequence differently, e.g. by saying that $\phi$ logically implies $\psi$ if, in every model, $\psi$ is true in those contexts in which $\phi$ is true and which are in the common domain of the characters of $\phi$ and $\psi$. (Of course, one must then not assume, as Kaplan does, that $\phi$ and $\psi$ can never be uttered in the same context.) But this would certainly not be a useful and intuitively adequate definition – it would be just as inadequate, indeed, as a definition of synonymy among two expressions as intensional equivalence in all contexts in which both expressions are uttered. From an intuitive point of view, the questions of whether a relation of consequence holds between two sentences or whether two expressions have the same meaning, have nothing to do with whether these sentences or expressions are uttered, or where they are uttered.

So there are good reasons for Kaplan’s decision to explain the character of $\alpha$ even for contexts in which $\alpha$ is not uttered. However, this view has its weak point, too. In particular, it has the consequence that, although the sentence ”I exist now” is true in every context and hence a priori, the sentence ”I am speaking now” is not. For $\forall \text{ I now exist } \exists (c) \ (\text{i}(c))$ is true just in case $s_c$ exists in $w_c$ at $t_c$, which is the case for every $c$ – because this is how possible contexts were defined; on the other hand, $\forall \text{ I am speaking now } \exists (c) \ (\text{i}(c))$ is true if, in $w_c$, $s_c$ is speaking at $t_c$ – which is not the case for all $c$ – because we allow for contexts in which no utterance is made at all. But then the intuition that ”I exist now” is a priori true is based on the knowledge that this sentence cannot be uttered falsely. It thus appears somewhat arbitrary to construe the notion of context in a way that only the first, but not the second, sentence counts as an a priori truth.

---

6 One might object that sentences can also be uttered without being uttered aloud, i.e. spoken; even the mere thinking of a sentence should count as an utterance. (We have actually been using this wide concept of utterance up to now.) Then, intuitively, not every utterance of ”I am speaking now”, but every utterance of ”I am thinking now” should be true – at least under the assumption that speaking always presupposes thinking.
We will still follow Kaplan in this point and only demand of possible contexts $c$ that $s_c$ exist in $w_c$ at $t_c$; it seems to me that the argument about logical consequence and synonymy clearly prevails over the objection just discussed. So if we evaluate the character of an expression $\alpha$ at a context $c$ this does not, strictly speaking, mean that we are considering the intension of an utterance of $\alpha$ in $c$ – but rather that we are asking which content $\alpha$ would have if $\alpha$ had been uttered by $s_c$ in $w_c$ at $t_c$. We are thus considering, as it were, not context $c$ but a counterfactual context which is like $c$ – with the possible exception that in it $\alpha$ is really uttered (plus any resulting differences like, e.g., that in $w_c$ at $t_c$, $s_c$ has enough command of the language to which $\alpha$ belongs that one truly can say he made an utterance in that language).

There is, however, a different point in which our conception of a possible utterance diverges from Kaplan’s notion of context: we take the character function of a language to be only defined for such contexts in which the language itself exists; we need this assumption in order to make sense of certain definitions – like that of the character of "water".

For what should the intension of water be, according to the semantic rule above, in a context in whose world English or the English usage of "water" did not exist? The definite description "the causal origin of the English usage of ‘water’ in $w_c$" would not designate anything in that case, which would in turn imply that the intension of "water" would be ‘empty’ in such a context $c$, i.e., a function assigning the empty set to each index $i$: in $w_c$, there would not be anything to stand at the causal origin of the English usage of "water" – because there is no English usage of water in $w_c$ – and thus nothing in $w_i$ could be the same liquid. I find this result intuitively inadequate. In contexts in which the English usage of ”water” does not exist, ”water” does not have an empty intension but no intension at all; the character of ”water” should simply not be defined for such contexts. We should rather reserve the above result for another case, namely for contexts in which the English usage of ”water” does exist but has no real origin because the English speakers who use the word ”water” somehow fell victim to a collective hallucination; only in such contexts I find it adequate to say that ”water” has an empty intension and, so to speak, designates a mythical notion.

We could, in this connection, not take the above way out and interpret the definition of the character of ”water” counterfactually, i.e., as saying, about contexts in which English does not exist, what the intension of ”water” would be if English did exist there. For this counterfactual mode of speaking could not be given any sufficiently precise sense. Let us, e.g., consider a context in a world in which there is $H_2O$ on one planet and $XYZ$ on another one, but in which neither English nor Twin English exist. If English had existed in that world, what would then be the causal origin of the English usage of ”water”? Would English be spoken on the one planet or on the other one? Such questions have no reasonable answers. Of course there do exist contexts in which it

---

7 A collective hallucination about water is hard to imagine. But history is full of all kinds of mythical entities that have been – and still are – taken for real.

8 This is reminiscent of Quine’s question concerning Bizet’s and Verdi’s nationalities if they had been compatriots (Quine 1952, p. 14f.).
would be clear what the English usage of "water" would be if only it existed. But once
one gets involved in such a counterfactual interpretation of the character function, one
ends up on a slippery slope with no grip between the harmless and the hopeless. A clear
policy can only be achieved by keeping away from the counterfactual way out and only
admitting possible contexts for "water" in which English exists.

Kaplan, however, explicitly rejects any such restriction:

"Given an interpreted language, a sentence is valid, if it expresses a truth in every context, includ-
ing those contexts in which the language doesn't or couldn't exist, or doesn't or couldn't have
that interpretation." (Kaplan 1989, p. 613)

For the remainder of the present section we will be dealing with this statement. Kaplan does not discuss this question in connection with the semantics of natural kind
terms but in a passage about Russell’s logically proper names (cf. Kaplan 1989, pp.
610ff.). However, this does not make any difference in the present context.

Logically proper names are, first, names that must denote something; but that which
they denote need not exist in all worlds. The assumption that names must denote
something could be reconstructed like this: If "a" is a logically proper name, then for all
c and all i , \( \parallel a \parallel (c) (i) \) is from \( w_c \) so that "a exists" is true in all contexts, i.e., a priori
true: \( \parallel a \exists \parallel (c) (i(c)) = 1 \) iff \( \parallel a \parallel (c) (i(c)) \) exists in \( w_c \). That a logically proper
name "a" does not denote a necessarily existing object, would mean that \( \parallel a \parallel (c) (i) \) did
not exist in all index worlds \( w_i \), i.e., that an utterance of "a exists" did not express a
necessary proposition: \( \parallel a \exists \parallel (c) (i) = 1 \) iff \( \parallel a \parallel (c) (i) \) exists in \( w_i \). In that sense
logically proper names are exactly like the indexical "I".

Second, however, logically proper names are not supposed to be indexicals. For
with a logically proper name I refer to an object with which I am directly acquainted
and about which I cannot be mistaken; then "a" can only denote this object and its
referent cannot change from context to context. Since a logically proper name, like any
proper name, is rigid, this means that it is absolutely rigid; we thus have: \( \parallel a \parallel (c) (i) = a \)
for all c and i in the domain of \( \parallel a \parallel \).

There is, however, a problem with this reconstruction. For if one assumes, as
Kaplan does, that all index worlds – or at any rate those in which there is at least one
speaker – are also possible context worlds, the question arises what "a" denotes if it is
uttered in a context with a world where a does not exist. Or, putting it slightly
differently: How can every utterance of a logically proper name denote something if
that which the name denotes does not necessarily exist?

As an illustration, Kaplan presents the following example: Imagine I call a certain
pain, to which I stand in an acquaintance relation, "Nixon". That I have this pain is not
a necessary truth; there are worlds in which Nixon does not exist. If I uttered "Nixon"
in such a world, which of the objects existing there would my utterance denote? It

Our question must not be confused with that of a cross-world identification of English. The latter can
be answered by stipulation; we could, e.g., simply say that w be a world in which English is spoken on
Twin Earth. But we could not decide by stipulation what is to be the case on counterfactual assumptions
already made.
cannot be Nixon, because Nixon does not exist there. Neither can it be anything else, because "Nixon" is absolutely rigid.

We have, indeed, arrived at a straightforward contradiction generated by the following four assumptions:

1. In every context, ”Nixon” does not denote just any old object, but always the same one, namely Nixon.
2. Nixon exists in every possible context.
3. Nixon does not exist in all possible worlds.
4. Every possible world (in which at least one speaker exists) is the world of some context.

Kaplan escapes this dilemma by giving up assumption (2). Thus, according to him, the analysis of logically proper names as denoting an existent object in every context is wrong and the intuition that ”a exists” be a priori true is erroneous. Rather, in Kaplan’s view, that a logically proper name denotes a certain object is a given fact of the language under consideration; and this need not be so in all possible contexts but can be imported into them:

"I see here a reaffirmation of the importance of ... the distinction between what exists at a given point and what can be ‘carried in’ to be evaluated at that point, though it may exist only elsewhere. My ‘Circumstances of Evaluation’ evaluate contents that may have no native existence at the circumstance but can be expressed elsewhere and carried in for evaluation. What is crucial to the puzzle about ”Nixon” is that my ‘Contexts of Use’ are also points of evaluation.” (Kaplan 1989, p. 613)

According to Kaplan, then, the character of a logically proper name ”a” would have to be described thus: $\|a\| (\langle i \rangle) = a$ for all $c$ and $i$ in the domain of $\|a\|$, where $a$ must only exist in the actual context $c_0$ but not necessarily in the other possible contexts. However, for him the demand that only those contexts be relevant for a language in which this language exists amounts to analyzing how the expressions of the language get their intensions and hence their extensions in a given context; and this analysis he classes as belonging to metasemantics, as he calls it, not to semantics, which is what he is interested in.

This classification does not convince me, though. Our distinction between recursive semantics and the empirical theory of reference as introduced in section 3.1 (p.94) roughly corresponds to Kaplan’s distinction between semantics and metasemantics. I have already explained the differences and connections between these two areas. Given that, recursive semantics does presuppose an empirical theory of reference. This does, however, not mean that recursive semantics reduces to an empirical theory of reference; recursive semantics can lean on the empirical notion of reference without clarifying it.

---

9 These are our indices.
3.3 What is a Possible Context?

I do, indeed, see the Nixon story differently. There is nothing to be done about the first and third of the above assumptions, because it was presupposed that we are dealing with a logically proper name denoting a not necessarily existing object. But inconsistency also vanishes when we give up the fourth assumption and as possible context worlds for ”Nixon” only admit such worlds where the word ”Nixon” in its given interpretation and, consequently, the object Nixon itself, both exist.

This way out of the dilemma is not implausible. Certainly, logically proper names are strange expressions and one can already dispute whether natural language contains any of them at all, or whether they are not just a philosophers’ invention. If, however, one accepts that they do exist and that there are thus things I am directly acquainted with and cannot be mistaken about and that I can name these things, what is it then supposed to mean that such names are used in contexts in which these things have never existed? If I call this one definite pain ”Nixon”, how could an utterance of ”Nixon” in a world in which I have never experienced that pain denote it? I think it is much more plausible to say that, in such a context, an utterance of the symbol ”Nixon” simply is not, and cannot be, the utterance of the same word in a semantic sense.

Moreover, this way out has no counter-intuitive consequences. For it is important to realize that the dilemma exclusively arises for absolute names, because only for these have we assumed that the actual denotation is part of the meaning and hence of the word itself. Only in the case of absolute names, then, does the demand that the character function be only defined for such contexts in which the language exists have the further-reaching consequence that in each possible context world its denotation must exist, too. But no objection should be arise out of this consequence, because we will later argue that ordinary proper names, like natural kind terms, are not absolute, but hidden indexicals.

Here, finally, a further disagreement with Kaplan and even an advantage of our way out become apparent. According to Kaplan, ordinary proper names, like logically proper names, have constant characters. And since, of course, the second assumption, viz. that its denotation exists a priori, is implausible in the first place, it is only logical for Kaplan to altogether drop this second assumption. However, with his equal treatment of ordinary and logically proper names he can no longer account for the epistemological peculiarity of logically proper names; he cannot reconstruct the difference between proper names that do not induce a priori existence, and logically proper names for which this is at least plausible – due to the direct acquaintance with the denotation. On our account this difference will, however, show in the distinction between covertly indexical and absolute proper names.

Analogous observations apply, mutatis mutandis, to predicates. They reveal that Kaplan’s disapproving attitude need not worry us; on the contrary, they corroborate our claim that a context is only to be admitted as a possible context of utterance in a given language if the language exists in the world of that context.
3.4 The Meta-Language Objection

Let us now return to the definition of the character of "water":

\[ \| \text{water}_E \| (c) (i) = \{ x \mid x \text{ consists of the same liquid as those things that form the causal origin of the English usage of "water" in } w_c \}. \]

It is conspicuous that this account of the meaning of "water" itself contains the word "water". There is a standard objection against semantic rules of this kind and we will have to address it now. The objection says that such rules amount to re-interpretations of object-linguistic expressions as meta-linguistic and that this is inadequate; I will thus call this objection the meta-language objection.

In our special case the argument runs as follows: Once we claim that "water" roughly means the same as "that which is called ‘water’ in English", we would also have to say that a sentence like "Water quenches thirst" means the same as: "That which is called ‘water’ in English quenches thirst". Intuitively, however, the former is an object-linguistic assertion about water, whereas the latter is a meta-linguistic assertion about "water"; so our semantic explanation cannot be right.

The meta-language objection can be traced back to Frege who, in his 1879 "Begriffsschrift", had analyzed the identity sign as a relation between names of objects (Frege 1879, §8), thus solving the problem of informativity for identity statements. But in his 1892 article "On Sense and Reference", he criticizes his earlier analysis:

"What is intended to be said by \( a = b \) seems to be that the signs or names ‘\( a \)’ and ‘\( b \)’ designate the same thing, so that those signs themselves would be under discussion; a relation between them would be asserted. But this relation would hold between the names or signs only in so far as they named or designated something. It would be mediated by the connexion of each of the two signs with the same designated thing. But this is arbitrary. Nobody can be forbidden to use any arbitrarily producible event or object as a sign for something. In that case the sentence \( a = b \) would no longer refer to the subject matter, but only to its mode of designation; we would express no proper knowledge by ist means." (Frege 1892, pp. 56.)

Putnam (1988) presents another version of the meta-language objection: If – briefly borrowing Putnam’s original example – one says that "elm" means "that which is called ‘elm’ in English”, one would also have to claim that it is analytically true that an elm is what is called an “elm” in English. Putnam takes this to be absurd: "no philosopher ... has ever maintained that it is analytic that elms are called by the name ‘elm’” (p. 27). Even though, he thinks, what is called what in a particular language is very important to speakers of that language, it is not analytic:

"An important part of the purpose of the notion of \textit{meaning} is precisely to \textit{abstract away from} the phonetic shape of the name. To say that the phonetic shape of the name (\textit{elm'}, or \textit{Ulme'}, or \textit{orme'}) is essential to the meaning is to confuse precisely what we want to abstract away from in meaning talk." (Putnam, 1988, p. 27)
A particularly frequent form of the meta-language objection says that, according to ‘meta-linguistic’ semantic theories, sentences from different languages can never be synonymous and that it would thus be impossible to correctly translate from one language into another – a consequence nobody would seriously try to defend. I will call this particular version the translation argument. Putnam (1988, p. 27) uses it to clarify his own position in the following way:

Intuitively, sentence (5) means the same in English as does sentence (6) in German:

(5) I like drinking water.
(6) Ich trinke gerne Wasser.

However, following our definition of the character of ”water”, (5) would have to be analyzed as (7) and, analogously, (6) as (8):

(7) I like drinking what is called ‘water’ in English.
(8) Ich trinke gerne das, was im Deutschen ‘Wasser’ heißt.

But the German translation of (7) is not (8), but (9):

(9) Ich trinke gerne das, was im Englischen ‘water’ heißt.

Obviously, (9) is not synonymous with (8); thus (7) does not mean the same as (8) either, so that our semantic theory seems to assign different meanings to (5) and (6).

The translation argument can also be combined with Putnam’s first objection: If ”water” means ”that which is called ‘water’ in English”, then not only would ”water is what is called ‘water’ in English” have to be an analytic truth, but so would its German translation: ”Wasser ist das, was im Englischen ‘water’ heißt”. But this seems to be a particularly absurd result, since one can hardly claim that this sentence must strike any competent speaker of German as true.

The translation argument also often comes in an individualistic form where it says that, according to meta-linguistic theories of meaning, speakers of different languages can never be attributed the same beliefs: If a speaker of English uttered (5), one would want to say that, ceteris paribus, he thereby expresses the same thought as a speaker of German when uttering (6); but, on a meta-linguistic analysis one would have to ascribe them the distinct thought contents (7) and (8).

The meta-language objection – and especially in its translation argument version – counts as a striking objection against any attempt to explain meanings of expressions by reference to the expressions themselves.10 This is why this initially inviting strategy of solving the problem of informativity is hardly ever pursued any further. In what follows I will, however, try to show that the analysis advocated here is not among the meta-

---

10 This is why it is normally only mentioned briefly and by the way. Cf., besides Putnam (1988, p. 27), e.g., Loar (1986, p. 101) and Salmon (1986, p. 72). A slightly more extensive discussion can be found in Burge (1979, pp. 96-99).
linguistic theories of meaning to which the above objections would apply. It does not imply that a sentence like (5) is to be interpreted as a statement about "water" and not about water. It makes clear that it is not at all absurd but even, on one reading, correct to construe the sentence "Water is that which is called ‘water’ in English" – and its German translation "Wasser ist das, was im Englischen ‘water’ heißt" – as an analytic truth. And it does not only allow for (5) and (6) to be synonymous but, in a sense, even explains this synonymy.

Let us start with Putnam’s first point. Why does it appear so absurd to Putnam that it is analytic that water is whatever is called "water" in English? He seems to take it that by "meanings" intensions are meant; in any case, he seems to object to the following semantic rule: A body of liquid \( x \) is in the extension of "water" in a world \( w \) if and only if \( x \) in \( w \) is called "water" in \( w \)’s English. A semantic rule like that would make being called "water" in English an essential property of water: in every world it would hold that the water in it is exactly what is there called "water" in English. And this is indeed absurd; for it just contradicts Putnam’s insight that "water" has the same referent in every world, viz. \( \text{H}_2\text{O} \). Of course, there could be worlds in which the origin of the English usage of "water" is different from that in the actual world, \( \text{XYZ} \) instead of \( \text{H}_2\text{O} \), say. Similarly, there could be worlds in which English had developed in quite a different fashion and in which the speakers of English called beer "water", or had yet another word for water.

But then it is also clear that this is not our account of the meaning of "water". Two points are crucial here:

The first point to note is that, in our character definition, we do not speak of the English in \( w_j \) or in \( w_c \), but of English as it is at the moment considered. The English language is a complex and certainly pretty vaguely individuated object in constant development, and hence different at different times; and it also is highly contingent, and hence different in different possible worlds. But the character function can only be sensibly understood as describing a particular (semantic) state of English as determined beforehand.\(^{11}\) In our definition English must therefore be construed as such a state of English, occurring as a fixed parameter in both the definendum and the definiens, and not as considered at different indices or contexts. In particular, this means that, when we talk about English in different contexts, we thereby do not mean English as it could have been, but precisely the English of a previously determined state.

Later, this point will become noticeable again. I have already mentioned it here because it meets the criticism of our definition which is nourished by the intuition that linguistic signs are, after all, arbitrary; this is also the intuition called upon in the above Frege quotation. To be sure, the English language could have used a different word for water, or "water" for something else. But the actual present state of English could not; otherwise it would just not be its actual present state.\(^{12}\)

---

\(^{11}\) Thus a semantic change of English must be described by a change of the character function over time. We will later return to this point (cf. VVV pp.113f).

\(^{12}\) Of course, "actual" must be taken as an indexical here, not as redundant.
The second crucial point is that, according to our rule, water in \( w_i \) is not whatever in \( w_i \) is called "water" in English in the above, precise sense, but that which is like the stuff which in \( w_c \) is at the origin of the English usage of "water". The whole point of our rule was to both capture Putnam's intuition that "water" is a rigid designator referring to the same stuff in every world and his intuition that "water" is a hidden indexical, and thus a context-dependent expression. In no way does our semantic rule for "water" say that it is an essential property of water to be called "water" in a given state of English; the indistinguishable XYZ might be called so instead. It only says that, whenever we speak of water in this world, water is in all worlds identical with whatever we talk about here.

In other words: According to our semantic rule, the description "that which is called 'water' in English" must be interpreted referentially, not attributively. On its attributive reading, the sentence "Water is that which is called 'water' in English" is not analytic indeed.\(^{13}\) On its referential reading, however, it is analytic – at least if our account of the hidden indexicality of "water" given in section 3.2 is correct. For on this reading, the sentence clearly brings out the context dependence of "water", ascribing, in any context \( c \), self-identity to whatever in \( c \) is called "water" in English.

So, as far as the intension of "water" is concerned, the word "water" itself is not essential for it. To this extent, Putnam's accusation that one should not turn into a component of meaning what one wants to abstract from, misses its target. And we leave Frege's basic intuition underlying the meta-language objection intact: one would want to talk about things, not about words; on the referential reading, one does talk about the thing itself, whether one uses "whatever is called 'water' in English" or simply "water".

The toughest argument against meta-linguistic approaches is, no doubt, the translation argument. However, it will also lose its power once certain hidden ambiguities are removed; and we will finally be able to show that the English sentence (5) and the German sentence (6) are indeed synonymous on our theory.

So let us take a closer look at the translation argument and at sentences (5) - (9). The first confusion arises because the argument mixes two notions of translation. Our question was whether (5) and (6) are identical in meaning, or synonymous. The most sensible way to construe the notion of synonymy within our theory is obviously as having the same characters. We are therefore only interested in translations inasmuch as they are character-preserving; expressions of one language are to be translated by target-language expressions with the same characters. But if Putnam – quite plausibly – claims that only (9) is a translation of (7), but not (8), he thereby uses a different, stricter notion of translation. According to it, translations must apparently satisfy a character-level version of Carnap’s intensional isomorphism\(^ {14} \) – character isomorphism, as it were, when each part of a complex expression of the source language must be

\(^{13}\) There is a hidden subtlety here. For, strictly speaking, Putnam does not talk about this sentence but of the sentence "Water is called by the name 'water' in English" – see the first of the above quotations. Putnam’s sentence is only equivalent to the attributive reading of our example, not to its referential reading. Putnam is therefore quite right in categorizing his example as not analytic.

\(^{14}\) Cf. Carnap (1947), pp. 56-64.
translated into a corresponding part of a target-language expression with the same character.\footnote{It should, moreover, be remarked that literary translations more or less require direct speech in one language to be translated as direct speech in the other language – and thus (7) as (8) and not (9). [This shows that there is yet another criterion for translations, viz. sameness of comprehension requirements across the audiences of different languages.]

Granting this stricter notion of translation, Putnam is right; "‘Wasser’" and "‘water’" as well as "English" and "das Deutsche" do not even have the same extensions, let alone the same characters. But the meta-linguistic theory of meaning does not collapse by simply applying a stricter criterion of translation; for the latter still leaves open whether (7) is synonymous with (8), and with (9), in the weaker sense of being identical in character. After all, one would not want to say that "bachelor" and "unmarried man" are not synonymous just because "bachelors are neurotic" is the correct translation of "Junggesellen sind neurotisch", whereas "unmarried men are neurotic" is merely a paraphrase and not a translation.

But even if one ignores the stricter criterion for translation, the impression remains that (8) and (9) are not synonymous. Here another, already familiar confusion comes into play, viz. the ambiguity between the referential and the attributive readings of the descriptions "das, was im Deutschen ‘Wasser’ heißt" and "das, was im Englischen ‘water’ heißt" in (8) und (9). The impression is clearly due to the attributive reading. If, however, one applies the referential reading – which is the only one that accords to our theory –, intuitions about the understanding of (8) and (9) become shaky.

Up to now, my sole aim has been to deprive the translation argument of its initial plausibility. But it is still an open question whether, on their referential readings, (6) and (5), or (8) and (7) or (9) are synonymous. I will thus go one step further and show that our definition of the character of "water":

\[ \text{\text{\textit{water}}}_E \text{\textit{(c)}}^i = \{ x \mid \text{in w}_i, x \text{ consists of the same liquid as those things that form the causal origin of the English usage of "water" in } w_c \} \]

and the corresponding definition of the character of "Wasser":

\[ \text{\text{\textit{Wasser}}}_D \text{\textit{(c)}}^i = \{ x \mid \text{in w}_i, x \text{ consists of the same liquid as those things that form the causal origin of the German usage of "Wasser" in } w_c \} \]

define the same function and that therefore (5) and (6) have the same character.

At first glance, it would seem as if the arguments in the previous section implied that the characters of "water" and "Wasser" could at most coincide in their common domain, the contexts in which both English and German exist; only in contexts in which both the German usage of "Wasser" and the English usage of "water" exist does it make sense to ask whether the two have a common causal origin. Hence sameness of

\[ \text{\text{\textit{water}}}_E \text{\textit{wasser}}_D \text{\textit{(c)}}^i = \{ x \mid \text{in w}_i, x \text{ consists of the same liquid as those things that form the causal origin of the English usage of "water" in } w_c \} \]

And the corresponding definition of the character of "Wasser":

\[ \text{\text{\textit{Wasser}}}_D \text{\textit{(c)}}^i = \{ x \mid \text{in w}_i, x \text{ consists of the same liquid as those things that form the causal origin of the German usage of "Wasser" in } w_c \} \]

define the same function and that therefore (5) and (6) have the same character.

At first glance, it would seem as if the arguments in the previous section implied that the characters of "water" and "Wasser" could at most coincide in their common domain, the contexts in which both English and German exist; only in contexts in which both the German usage of "Wasser" and the English usage of "water" exist does it make sense to ask whether the two have a common causal origin. Hence sameness of
character among "water" and "Wasser" only seems to be possible under this additional restriction. We will, however, soon see that this restriction is actually empty.

But let us first restrict our attention to contexts in which both English and German exist the way they really are – as we have pointed out above. In the actual world – or so we assume – typical cases of English "water" and of German "Wasser" consist of the same stuff, viz. H\textsubscript{2}O. But we could be wrong. It might be the case that the causal origin of "water" is a different stuff than the causal origin of "Wasser", so that English would be related to German in quite the way as it is to Putnam’s Twin English. By allowing for the possibility of such error, however, we seem to admit that there are context worlds in which the causal origin of the English usage of "water" and that of the German usage of "Wasser" are distinct. This implies that "water" and "Wasser" differ in their characters. Have we thereby proved ourselves wrong?

No; the chain of conclusions is defective. For there is a crucial respect in which the present relation between English and German differs from that between English and Twin in English in Putnam’s story. There it was part of the story that there had been no contact between Earth and Twin Earth until one day Earthian visitors landed their space-ship on Twin Earth to find that the liquid that is called "water" there is not water. Things would have been different if contact between Earth and Twin Earth had been made long before anybody had been able to distinguish H\textsubscript{2}O from XYZ. If a cultural exchange between Earth and Twin Earth had started at that time, the causal-intentional complexes of English and Twin English would have been so intertwined that bodies of XYZ would have become paradigm cases of applying the English term "water", just like bodies of H\textsubscript{2}O would have become paradigm cases for applying the Twin English term "water"; given their otherwise identical features, English and Twin English would thus have merged into one language. If then some day one had discovered that the rivers and lakes on Twin Earth are filled with a different stuff than those on Earth, one would only have drawn the conclusion that "water" obviously is not a proper natural kind term but that it rather works like Putnam’s "jade" example: Just like "jade" denotes two distinct materials, viz. jadeite and nephrite, "water" would then denote two distinct substances that happen to occur in distinct regions of the universe.

Now the point is that today there does exist such traditional cultural exchange between English and German. We translate German "Wasser" into English "water", and vice versa. And since this practice of translation is well-established, our usage of "water" is, as it were, tied to the German usage of "Wasser", our typical cases of applying English "water" having been extended by those samples that for Germans are typical cases of applying "Wasser" – and vice versa. In the problem case where the stuff that satisfies the conditions of German "Wasser" within the German language area, is different from the stuff satisfying the conditions of the English usage of "water" within the English language area, it would turn out that both "water" and "Wasser" relate to the same two distinct substances.\footnote{The symmetry of the exchange is crucial here. For if only German had been translated into English, then the English usage would be tied to the German, but not the German to the English.}

We may thus conclude that the characters of "water"
and "Wasser" produce the same extensions in at least those contexts in which both languages exist the way they do nowadays.

These considerations may appear paradoxical. For their upshot is that, in the beginning, translation can never achieve synonymy in the sense of character identity and that the latter is only established by a long-term mutual practice of translation. But at closer inspection, this is not at all paradoxical. It just seems to be an inevitable consequence of Putnam’s observation that many of our predicates are hidden indexicals; and the claim only applies to them.

Now the above-mentioned restriction turns out to be empty too; the characters of "water" and "Wasser" do not only have the same course of values on their common domain, they also have the same domain, and are thus identical indeed. For we are concerned with the English and German usages of "water" and "Wasser" today, i.e., at a time when both have been translated into each other for a long time and when the causal-intentional complexes pertaining to English "water" and German "Wasser" have long melted into one. However, this means that the present state of English as such could not have existed without being tied to German – and vice versa; thus, in every context in which present-day English exists, present-day German exists, too – and vice versa.

Finally, these reflections also bring out that, just like the sentence "Water is whatever is called ‘water’ in English”, the sentence "Water is whatever is called ‘Wasser’ in German" is analytically true, on its referential reading anyway. The intuitive reservation one might feel against this consequence vanishes once one recognizes that in this connection analyticity in a language only means truth as determined solely by the character function of that language, and that it does not mean that every competent speaker must immediately realize the truth of an analytic sentence; claiming the latter would indeed be above anybody’s linguistic competence. We will get back to this point and discuss it in depth in sections 3.8 and 3.9.

We may safely conclude that the meta-language objection has in every respect been rejected. Does that mean that our theory cannot be characterized as meta-linguistic? No. Its meta-linguistic character becomes particularly obvious when we consider our account of subjective meaning in terms of the diagonal rather than the usual entire character function. For according to the above character definition the diagonal of "water" can roughly be described as "that which is called ‘water’ in English". Again, "in English" refers to the state of English under consideration; but otherwise the description must be read attributively –the diagonalization turning referentially read descriptions into attributive ones. And what we managed to deny in the case of the referential reading is now true of the attributive reading: the causal-intentional complex of being called ‘water’ including the word ‘water’ itself, is now part of the informational content and not just of the contextual conditions of the description.

However, this meta-linguistic approach is not to be criticized as such but only if it leads to undesired consequences; and these we have so far managed to completely reject. Moreover, jumping ahead a bit, it may be noted that our final explication of subjective meaning will not be based on the above definition of character but on a further modification to be given in section 3.9, after having left the present debate of the
problem of informativity for a discussion of the problem of internalism. There we will be able to find yet another justification of the meta-linguistic nature of our account. The next few sections will, however, be devoted to the question of how our findings concerning the meaning of 'water' can be generalized and carried over to other predicates and thus in what way natural kind terms differ from other concepts.

But let us first record an observation immediately pertaining to the above discussion. In our rejection of the translation argument we obviously made essential reference to present-day English and present-day German; it would not have worked if we had considered a time at which English was still separated from German, as it is from Twin English – if such a time there ever was. Here a need already established on p. VVV 108 VVV shows up again: the character function of a language must be relativized to a particular time (and a particular possible world) even though our notation has not made this relativization transparent. We may therefore describe semantic change by a temporal change of the character function. Semantic change is not our topic; but as it would be an interesting topic, I would at least like to mention two aspects emerging in the case of covertly indexical natural kind terms.

First, our discussion has brought out that the character of such a natural kind term already changes, so to speak, by travelling, i.e., by an extended reference space of the respective language community; for thereby the set of paradigm cases simply changes.

Secondly, it is clear that the character of a natural kind term changes as the relevant differentiation capacities of the language community become finer (or are lost). Let us, e.g., suppose that $\text{H}_2\text{O}$ and XYZ can only be distinguished by electrolysis and by no other previously known procedure. It is then conceivable that, in the case of English $E_0$ as before the discovery of electrolysis, there would have been contexts in which, for whatever reason, the English speakers have had massive contact with portions of XYZ. These XYZ portions would then pass as paradigm cases of "water". In such a context, "water" thus denotes both substances. English $E_1$ after the discovery of electrolysis can also be imported into such contexts. Since, however, behaviour under electrolysis has become one of several criteria for making out paradigm cases of "water" – which does not mean that it is an analytic condition on water – all these new portions of XYZ can no longer end up among the paradigm cases of "water". And thus, in such a context, "water" still denotes water in $E_1$, and not XYZ.

So the character of our covertly indexical natural kind term at least partly depends on how the extension of our reference space and the refinement of our distinguishing capacities interact. Most certainly, scientific theorizing also determines our distinguishing capacities – which leads me to a final remark concerning the philosophy of science:

One of Putnam’s (1975) crucial motivations, to be sure, had been to return from an empiricist or operationalist philosophy of science to a realistic one. To this aim, he comes up with a host of examples substantiating his claim that, contrary to non-realist doctrine, scientific and other concepts do not change with our theories about the things denoted by these concepts. This justification seems to be largely accepted meanwhile.
But does that mean that the authors criticized by Putnam\textsuperscript{17} have simply been wrong, that the view that the meaning of scientific terms depended on theories in which they appear is simply mistaken? Not necessarily. For when Putnam talks about meaning, he is talking about intension; and, as far that goes, we agreed with him. But if, on the opposed view, meaning is character – or the diagonal of the character, which we have identified with subjective meaning – he would no longer be right; for if the above observations are correct, the character of a natural kind term does change, even if its intension remains unchanged. I do not want to analyze whether the philosophers Putnam criticizes did have such an escape in mind or if it would have been acceptable to them; but I did want to remark that our framework opens up the possibility of reconciling the opposing views on this important matter.

3.5 Essentiality Conventions and Hidden Indexicality

We have so far clarified and defended our definition of the character of "water" as developed in section 3.1. In this section and the next one, we are going to discuss how the observations on "water" can be generalized to other predicates. Let us therefore briefly summarize these observations.

We had made out two factors determining the intension of the English word "water": the usage of "water" on the one hand, and on the other hand the way the world is. The English usage of "water" – at a given time \( t \) – restricts the intension of "water" by determining which objects in a context world \( w_c \) count as typical cases for applying "water" in English. Thus, e.g., at the present time the criteria for typical cases of applying "water" do not just consist in their being, under certain circumstances, transparent, liquid and thirst-quenching, but also in their behaving like \( \text{H}_2\text{O} \) under electrolysis; in present-day English "water" could neither refer to beer nor to ice cream nor to new portions of XYZ. In a sense, the English usage of "water" thus offers a descriptive content of "water"; however, this descriptive content is not to be reconstructed as an intension but as a condition on possible character functions for "water".

Within these boundaries, the intension of "water" is only fully determined given the condition of the world, i.e., the actual kind of liquid of which the typical cases of applying "water" consist. And precisely because a range of variation still remains, "water" is a hidden indexical.

We have followed Putnam in taking it to be intuitively clear that this range of variation does indeed exist. But we have not touched on the question of how it comes

\textsuperscript{17} Putnam (1979) himself gives no names, but it is clear that he chiefly adddresses logical empiricism, as supported by Rudolf Carnap and many others, but also positions like Thomas Kuhn’s and Paul Feyerabend’s.
about. The easy answer is, of course, that it is based on the English usage of "water" too. But we have yet to focus on the aspect that is responsible for it: the range of variation obviously derives from assumptions on which the English usage of "water" rests – assumptions as to what kinds of properties are essential for being water, that it is its physical structure, and not its surface properties that decide whether something is water, that, briefly, "water" is a stuff name in English; we will refer to this aspect as the English essentiality convention for "water". More precisely, the range of variation is due to the fact that the English essentiality convention for "water" only determines which kind of property is essential for water – viz. its physical structure – and that it leaves it to the world to decide which property this is going to be in the end – e.g., being H2O.

There are thus two aspects of the English usage of "water" that determine the character of "water". On the one hand it determines, for each context world, which things in that world may count as typical cases for applying "water". On the other hand the English essentiality convention allows to project an intension, i.e., an extension for every index world, from these typical cases in the context world.

It seems that, for each predicate, an essentiality convention must exist that fixes its context-dependent or context-independent intension, thus immediately affecting its character. This observation now suggests the following general schematic character of English unary predicates:\footnote{From now on, the subscript "E" will be attached directly to the English character function – and rightly so: after all, in the end we would like to get hold of the entire character function for the whole English language. Since we have so far only been concerned with a few words, it was still possible to mark their linguistic origin by a subscript on them.}

\[
\ll P \ll_E (c) (i) = \{x \mid x \text{ in } w_i, x \text{ possesses those properties which, according to the English essentiality convention for } "P" \text{ in } w_c, \text{ are essential for those objects' being } P \text{ to which } "P" \text{ is typically applied in } w_c\}.
\]

It is only awkward, but not difficult, to generalize this scheme to n-ary predicates.

The scheme as such, of course, lacks any substance and is even circular. For on the one hand, the character of "P" is supposed to yield, for a given context c, the property expressed by "P" in c; but on the other hand this very property is mentioned in the definiens in these very terms. Moreover, this property itself is one of the properties that are essential for it. However, none of this depreciates the scheme. We have already seen in which way its instantiation with "water" is informative. And in the following sections it will turn out to be a practical guide once we start explaining its scope and illustrating it with other examples.

First, however, we need to fix our terminology. For our use of "essential" in the above scheme somewhat deviates from usual practice. In fact, three different usages of "essential" must be distinguished.

In its first common usage, it expresses a relation between objects and properties: A property Q is essential for an object a if, and only if, a could not have existed without
being $Q$, i.e., if the sentence "if $a$ exists, then $a$ has property $Q$" is necessarily true. One also says that an object is constituted by its essential properties. For instance, both being human and having the very biological parents he or she has is essential for each single human being.

Second, "essential" is used as a property of properties: a property $Q$ is essential if, and only if, it is essential for each object to which it applies, i.e., if the sentence "any $Q$ is necessarily $Q$" is necessarily true. For instance, being human is an essential property; every possible human being is necessarily a human being. Essential properties in this sense are also called constitutive; they constitute objects. The first usage could be reduced to the second one if it were only essential properties that could ever be essential for an object. However, we can leave it open whether this is so or whether there are rather properties that necessarily apply to some of their instances and contingently to others.

The above character definition for predicates does, however, not make use of either of these meanings of "essential". Rather, we have used "essential" in a third way, viz. as a relation between properties: Property $Q$ is essential for property $P$ if, and only if, an object can only be $P$ if it is also $Q$, i.e., if the sentence "every $P$ is $Q$" is necessarily true. That consisting of H$_2$O is necessary for being water thus only means that nothing that does not have the molecular structure H$_2$O can be water. If $Q$ is essential for $P$ and $P$ is essential for $Q$, then – and only then – may we say that $P$ and $Q$ are the same property. Thus, given that nothing but consisting of H$_2$O is essential for being water, being water and consisting of H$_2$O are one and the same property.

In the case of water and H$_2$O, two notions of essentiality coincide. On the one hand, consisting of H$_2$O is essential for being water. On the other hand, consisting of H$_2$O is an essential property simpliciter; no single portions of water or H$_2$O could exist without consisting of H$_2$O. This coincidence has created a considerable amount of confusion, as we will see. But an immaterial coincidence it is just the same. For example, it is initially plausible – and we will return to this – that being male and unmarried is essential for being a bachelor. And even if every single bachelor is essentially male, he is not essentially unmarried or a bachelor; for being unmarried and being a bachelor are simply no essential properties. Similarly, in our world being infected by a certain virus is essential for suffering from measles. But suffering from measles is not an essential property – which inter alia shows in the fact that whoever does suffer from measles only does so for a short while during his or her existence.

The next thing to do is to give a more precise account of the essentiality convention for "water", partly for illustration. We had said that it determines that, in any context $c$, the properties essential for being water are structural properties; the character of "water" then assigns to every context $c$ and every index $i$ the set of those $x$ from $w_i$ with the same material structure as the typical cases of applying "water" in $w_c$. This formulation seems to presuppose that the typical cases of applying "water" in $w_c$ happen to have a certain material composition because otherwise one could not speak of the same structure. But what if this presupposition is not satisfied? It is certainly not the case that in such a context "water" would have to be regarded as meaningless. Rather, in such cases the English essentiality convention for "water", as we understand it here,
reacts in a flexible way, adjusting as far as possible its presuppositions to the new situation. Two cases, already discussed by Putnam (1975), pp. 240f. can be distinguished.

On the one hand, the presuppositions of the essentiality convention for "water" can insofar be violated as the typical cases of application are not of a uniform structure but fall under two or more categories each of which is characterized by its own internal structure. For instance, it could have turned out that water occurs both as H2O and as XYZ. Or we may imagine a context in which the further development of physics reveals that there are two distinct kinds of H2O portions. In the first case there would then be two sorts of water. And in the second case consisting of H2O would still be essential for being water; but there would then be two sorts of H2O and thus also of water. The most obvious modification of the assumption that water be one substance would then be that it happens to consist of two substances. And the English essentiality convention for "water" would then say that water is characterized by a disjunctive essential property, that consisting of one or the other of the two substances is essential for being water. Precisely this seems to have happened in the case of "jade" – mentioned by Putnam (1975, p. 241) – when it was discovered that the typical cases of applying "jade" fall into two classes of quite distinct minerals, jadeite and nephrite.

On the other hand – and this is the more drastic violation of assumptions underlying the essentiality convention for "water" – the context could be such that the typical cases of applying "water" are not of one or a small number of uniform material structures. Chaos could have started immediately below the surface features of water. Then the essentiality convention for "water" would presumably declare these surface features the essential properties of water. Or the sub-atomic situation could be completely chaotic. Then, according to the essentiality convention for being water, it would still be essential to consist of hydrogen and oxygen atoms; but, being essentially characterized only by their chemical properties, hydrogen and oxygen themselves would not be substances.

All these examples and considerations are supposed to show that the English essentiality convention for "water" does not depend on the correctness of the assumption that water is a single substance. Rather, the convention includes different, more or less remote fallback positions that can be activated according to contextual demands. In the above general character definition this multiplicity of possibilities of reacting is taken to be hidden in the talk about the properties which, according to the English essentiality convention, are essential for being \( P \) in \( w_c \). Of course, with more and more bizarre contexts, the question of how English speakers would react with their essentiality conventions becomes an increasingly vague and speculative matter. But then this does not do any harm to our argument; up to a point, we do have remarkably clear intuitions about our essentiality conventions, including their fallback positions.

What, then, does the above general scheme say about the hidden indexicality of predicates? From an abstract point of view, the situation is relatively clear. According to the scheme, contextual dependence on a given \( c \) can only arise via the context world

---

19 In a sense, this possibility has become reality by the discovery of isotopes.
Accordingly, a predicate is a hidden indexical if, and only if, its character assigns two distinct extensions to an index and two distinct context worlds; to prove hidden indexicality thus always boils down to finding an example for such a situation.

The discussion so far suggests that this situation obtains precisely when the essentiality convention for the predicate does not by itself determine the properties that are essential for the property expressed, but only the kind of properties that are – thus leaving it to the context world to determine which properties of that kind are actually essential. This is indeed so. But it should not be taken to mean that the essentiality convention always only informs about what kind of properties are essential. It is, e.g., certainly analytic that a vixen is a fox (in a gender-neutral sense); so the essentiality convention for ”vixen” already fixes this without any contextual help, and thus for all contexts. This example can even be strengthened: It is arguably analytic that vixens are precisely female foxes; in that case the essentiality convention by itself thus not only lays down necessary, but also sufficient conditions for ”vixen”. Nevertheless, this determinateness of the essentiality convention does obviously not result in a non-indexicality of ”vixen”. On the contrary, the predicates ”fox” and ”female” are hidden indexicals, too. And since this is so, the indexicality of ”vixen” is compatible with the analytic truth of ”Vixens are foxes” and, given the second assumption, the analytic equivalence of ”vixen” and ”female fox” even necessarily implies the indexicality of ”vixen”.

The indexicality of a predicate ”P”, then, only goes away if an equivalence of the form ”All Ps, and only they, are Q” is analytically true, where the more or less complex predicate ”Q” denotes the property it expresses in a non-indexical way, i.e., if the essentiality convention for ”P” already fixes all properties that are essential for being P. Cases in point are certainly mathematical predicates like ”prime number” as well as geometric predicates like ”triangular” that may also apply to empirical objects; we will see that many examples are in fact rather problematic.

If the essentiality convention for ”P” is so specific that it renders ”P” absolute, then it should be possible to eliminate the reference to the context world wc from the above character scheme. This is indeed the case. For, first of all, the reference to the essentiality convention must then no longer be relativized to wc. And second, the reference to the typical cases of applying ”P” is vacuous; the typical cases of applying

\[ \text{This only means, though, that the scheme does not cover overtly indexical predicates like "past", "native" etc.} \]

\[ \text{It holds quite generally that the analyticity of the universal implication "all } P \text{ are } Q \text{" together with the indexicality or non-indexicality of one of the predicates do not bear on the indexicality or non-indexicality of the other predicate. For example, if we take it that, analytically, coins are flat metal pieces of money, it is then also analytic that coins are flat. Nevertheless, "coin" is indexical, because "metal" is, whereas "flat" is absolute, because such directly perceptible forms are devoid of any hidden nature. As to the reverse case in which "P" is absolute and "Q" indexical, I have not been able to find an example though.} \]
"P" are then simply objects that have property P and they no longer have the regulating function of serving as starting points in the quest for P’s nature.\textsuperscript{22,23}

3.6 Other Examples

Although the rather abstract findings of the preceding section may serve as a guiding line, when it comes to in concrete cases, they will turn out to be of limited help only. But then the discussion of particular cases too easily leads out of bounds: essentiality conventions come in many shapes, thus provoking ever new considerations, they are often vague and indeterminate, making the discussion indecisive, and they form a web in which the essentiality convention of one predicate refers to that of the others. Moreover, the most interesting cases, e.g., those involving phenomenal or other psychological predicates, lead right into hot philosophical debates. Having already extensively discussed the stuff name "water", I will therefore only briefly touch upon other, philosophically rather less tricky examples.

It seems to be widely accepted that natural kind terms whose essential properties lie in their referents’ internal material structure are all hidden indexicals – Putnam’s standard cases being physical stuff names like "water", "gold" and "aluminium", biological kind names like "tiger", "cat", "beech" and "lemon", but also verbs like "grow". And rightly so. For a natural kind term "P" would only then be non-indexical if the essentiality convention for "P" already completely revealed the internal structural properties. This would, however, mean that exact knowledge of the properties that are essential for being P would have become part of the English usage of "P"; and this in turn would mean that the English language community would have to possess final knowledge of the nature of Ps. But this is not the case; presumably, we can always be wrong about the nature of the world and, in any case, our scientific theories only approach it without any guarantee of full and final coverage.

On the other hand, little agreement is found in the literature as to how far hidden indexicality extends beyond natural kind terms.\textsuperscript{24} Putnam himself holds an extreme

\textsuperscript{22} It follows that, in the case of absolute predicates, it is unnecessary to restrict the domains of their characters in the way discussed in section 3.3; their essentiality conventions fix their intensions even in contexts in which the language considered does not exist.

\textsuperscript{23} At this point another interesting consequence of the above character scheme fits in: in the case of covertly indexical predicates "P" it is part of the English \textit{a priori} (i.e., true in all contexts in which English exists) that most cases of applying "P" (= most objects that speakers of English think are P) are indeed P; this is the only possible outcome of the regulative function of the cases of applying "P". This not too implausible consequence does, however, not carry over to non-indexical predicates, for which one may have expected it even more.

Davidson (1983) arrives at similar conclusions, albeit on quite different paths, viz., via his theory of interpretation that crucially involves the so-called ‘Principle of Charity’. I will not go into the relation between my reasoning and Davidson’s.
position; he claims indexicality for almost all predicates. In particular, he claims that even words that denote artefacts – like "pencil", "table", "plate" – are indexicals (1975, p. 242). Let us see whether we can reconstruct this claim within our framework and discuss Putnam’s arguments for the indexicality of "pencil" – the example he goes into most deeply (pp. 242f.).

First of all, Putnam assumes that the opposite of covertly indexical predicates are predicates that can be defined by a conjunction (or a bundle) of properties and are thus synonymous with descriptions; these he also calls monocriterial words. Moreover, he assumes that only for monocriterial predicates, but not for indexicals, analytic equivalences can be found – i.e., general statements of the form "Ps are Q" that are not only necessary but also a priori (though he talks about metaphysical and epistemic necessity in this connection).

These assumptions reveal how strongly Putnam is influenced by the very views he is attacking. At any rate, the considerations above already make clear that neither of these two assumptions are tenable. Our discussion of "vixen", which Putnam uses as an example of a descriptive, monocriterial and hence non-indexical predicate makes this clear: "Vixens are female foxes" just turned out be an analytic definition or equivalence the definitendum of which is thus as indexical as its definiens. Hence Putnam’s assumption that only absolute predicates can be defined must be restricted in the following, trivial way: only absolute predicates can be defined without reference to indexical predicates. But let us stick to Putnam’s assumptions for a moment, in order to be able to follow his line of reasoning. It goes like this:

If "pencil" were an absolute predicate, then following these assumptions, the meaning of "pencil" would allow for the formulation of analytic truths about pencils, like "Pencils are artefacts" and "Pencils are standardly intended to be written with"; indeed, given sufficiently many such statements, they would have to add up to an analytic equivalence. Now Putnam tries to undermine the purported analytic truth by showing that one of these statements, "Pencils are artefacts", is not an a priori truth. To this end, he constructs the following example: Let us imagine that one day we found that the typical cases of applying our term "pencil" were live organisms that reproduce, grow up, etc. If this is a possible context world for the English word "pencil", in that context the sentence "pencils are artefacts" cannot be true; thus, Putnam’s first thesis reads, this sentence is no a priori, and a fortiori not an analytic truth.

Putnam takes this argument to already prove the hidden indexicality of "pencil". However, the argument only becomes complete by a second thesis, also supported by

---

24 See the survey in Abbott (1989).

25 We are thus led leads to a further criticism of an example of Putnam’s: We cannot define "bachelor" as "unmarried man" – at least not if we understand "man" as a biological predicate. For otherwise the indexicality of "man" would carry over to "bachelor". But if, on Twin Earth, we were to find an exact image of our society including all our customs, but then discovered that we are dealing with robots rather than biological organisms, we would still want to talk of bachelors; that is, we would go on referring to these robots that play the same role over there as our unmarried men as "bachelors". This, anyway, is the kind of story told by Schwartz (1980, pp. 188-189) in his – to my mind plausible – defense of the absoluteness of "bachelor", which then cannot be defined in terms of indexical predicates.
Putnam: that pencils are necessarily artefacts, if they are artefacts in the actual context world; in other words, that the world of imagination – as an index world – would then not be a possible world in which pencils are living beings but only one in which organisms of a certain specis are mistaken for pencils. Now the hidden indexicality of "pencil" does follow\(^\text{26}\) – and, by analogy, the hidden indexicality of predicates denoting artefacts in general: given any context \(c\) and index \(i\), "pencil" denotes just those objects which, in \(w_i\), are of the same nature as whatever is called "pencil" in \(w_c\) – where "nature" must not be understood as internal biological or physical structure, but in a wider sense, so as to cover having been produced artificially.

Putnam’s arguments in favour of the indexicality of "pencil" and other artefact terms have been criticized most notably by Schwartz (1978). To begin with, Schwartz points out that the argument does not go through without the second thesis, which looks suspect to him; that is, he questions the necessary truth of "pencils are artefacts" at one context and that of "pencils are organisms" at another one. If our paradigm cases of "pencil" are artefacts, does that really exclude the possibility of naturally grown instead of artificially produced pencils? Could there not be, in another possible world, objects that, e.g., grow like fruit on trees or, as Putnam has it, spawn like fish and at the same time would have to be classified as pencils? And if our paradigm cases of "pencil" turned out to be live organisms, would we then really not regard objects in a different world with the same shapes and function as our pencils as pencils? Would we say instead that anything can only be a pencil if it has the same biological structure as our pencils – even if they may have mutated to a completely different form and function? Couldn’t things stand with pencils just like they do with sponges? After all, our paradigm cases of "sponge", in the sense of washing utensil, consist both of artefacts and of grown, if dead, organisms – so that thus sponges can be either.\(^\text{27}\)

Schwartz himself concedes that, in the end, intuition stands against intuition and that intuitions are not so clear and unequivocal as in the case of "water" or other true natural kind terms. But his – and my – intuition tends against Putnam and towards the view that only surface features like form and function are essential for pencils, i.e., that they supply the essentiality convention for "pencil" with content. If this is so, Putnam would simply have focussed on the wrong, irrelevant kind of example; he had better considered his other example, "pencils are standardly inteded to be written with". But even if one cannot follow Schwarz’s intuition, the discussion still shows what kind of considerations are are needed to find out whether "pencil" is a hidden indexical or not.

---

\(^{26}\) One needs to be careful here: The fact that "pencils are artefacts" is necessary but not analytic only implies the hidden indexicality of that sentence and thus of at least one of the two predicates occurring in it, but not necessarily of "pencil". The latter does not even follow from the fact that in one context"pencils are artefacts" is necessarily true while in another one "pencils are organisms" is; strictly speaking, this would require an additional premise, e.g., the certainly plausible assumption that, in both contexts, necessarily artefacts are not organisms. But the hidden indexicality of "pencil" also directly follows from the construction of the example; for there we face an index world in which "pencil" has two distinct extensions, depending on whether we take our actual world or that world itself as the context world.

\(^{27}\) The example is due to Abbott (1989), p. 280.
From the fact that the form and function properties essential for pencils are surface properties Schwartz does indeed conclude that these features are analytic for pencils and that thus ”pencil” is not indexical but absolute. What lies behind this conclusion, though, is Schwarz’s bipartition into ”natural kinds” and ”nominal kinds” plus the view that only natural kind terms, i.e. words that designate some underlying nature to be scientifically explored, are hidden indexicals. It is, however, by no means clear that this extreme opposite of Putnam’s view is correct. I would like to discuss this question now with reference to function terms, by which I mean predicates whose essential properties are properties of function.

What is it supposed to mean that a function term ”P” is indexical? In analogy to natural kind terms, this would just have to be the case if the essentiality conventions for ”P” only determined that the properties that are essential for being P are functional properties without, however, specifying exactly which of them are essential. The semantic rule for ”P” would thus not simply be: \( \ll P \ll_E (c)(i) = \{ x \mid \text{in } w_i, x \text{ has function } F \} \), but rather: \( \ll P \ll_E (c)(i) = \{ x \mid \text{in } w_i, x \text{ has the same function as those objects to which English ”P” is typically applied in } w_c \} \), where this function may change from one context world to the next. The upshot of this would be that the English-speaking community regards ”P” as a function term, without having settled the question of what the precise function of Ps is.

This is certainly a possibility that cannot be dismissed right away, as the following example shows: Let us imagine that, among the excavations of a long-gone culture, we found loads of similarly, but strangely shaped objects which – perhaps following a still existing oral tradition – we call ”schlamis” and which in that culture – that much we believe to know from our investigations – must have served for some definite purpose which is, however, completely obscure to us. We thus identify schlamis by their shape, but only for lack of better criteria; for it is not shape, but function, that counts in the end. Thus something can only be a possible schlami if it has the same function as our typical schlami cases from the excavations; it would appear, then, that ”schlami” is a case of a covertly indexical function term.28

So if Schwartz solely relied on the thesis that only natural kind terms could be hidden indexicals, his conclusion would not be correct. But in the case of ”pencil”, as in the case of most function terms, it must be conceded that Schwartz is right. For ”pencil” could only turn out to be a hidden indexical if the English-speaking community could ever be ignorant or even wrong about the function of pencils. Is it conceivable that none of them has ever used a pencil for writing, nor ever intended to do so, but only erroneously believed all others did? In view of the fact that everyone has acquired this belief mainly from everyday writing practice involving pencils, this question must, I think, be answered in the negative. Rather, this belief seems to be part of common knowledge without which the English usage of ”pencil” would not be the same. It is

28 Putman (1982) uses similar arguments to the effect that words like ”sloop”, ”game” or ”joke”, though not indexical for us, are hidden indexicals for a Martian anthropologist who is not acquainted with our habits. The same point is made by Kornblith (1980).
thus analytic that pencils are standardly intended to be written with and thus "pencil" appears to be non-indexical indeed.\textsuperscript{29}

Empty predicates form another class of interesting cases for our character scheme. Again, some of them are hidden indexicals and some are absolute.

A case in point is "unicorn"; there are no unicorns in our world. As Kripke (1972), p. 763, wrote, unicorns do indeed form a mythological species. Like "tiger", "unicorn" is therefore a covertly indexical natural kind term; the nature of unicorns is contingent on the actual paradigm cases of "unicorns". However, since there are none, nothing would be essential for unicorns; no possible object could be said to have the same nature as these non-existing cases of application. Hence unicorns are impossible objects; they do not exist in any possible index world. This conclusion holds in general; any covertly indexical empty predicate is necessarily empty.

We nevertheless can imagine context worlds in which "unicorn" is not empty, in which it thus turns out that the unicorn myth is not a fairy-tale but a legend that can be traced back – maybe via a long oral tradition – to an extinct species. As Kripke emphasizes, unicorns exist in such context worlds not because there are things that match the usual description of unicorns, but only because our usage of "unicorn" has a real causal origin. And, of course, if there are unicorns in the context world, they also exist in many index worlds. All these modal statements immediately fall out of our account of hidden indexicality by our character scheme, and they are in full accordance with what Kripke has said about unicorns (and which, at the time, still sounded so strange that he felt obliged to give further explanations in the addenda).

Let us make a comparison with another empty predicate, "witch". Not so long ago, there used to be many paradigmatic cases. Had "witch" been a hidden indexical or even a natural kind term, most of these cases would have to have been witches, whose nature would have to be determined by the essentiality convention for "witch". The fact that, according to our present understanding, there are no witches could then only be explained by a change of the character of "witch". A better explanation would be that the essentiality convention for "witch" is already specific enough to exclude the existence of witches. It is thus natural to say that, analytically, witches are just women in alliance with the devil. This does not mean that "witch" could not be indexical. Rather, "witch" would be as indexical as "woman"; and we better not go into the question of whether "devil" is a hidden indexical. Anyway, this already suffices to derive, as we do nowadays, the non-existence of witches from the non-existence of the devil. And it opens up the possibility that, unlike unicorns, witches do not exist only contingently; if the devil can exist in other index worlds, then there could also be witches in those worlds. Be that as it may, we have seen that, in general, absolute predicates can also be contingently empty.

Absolute predicates can, however, also be empty in another way. Take round squares. There are no round squares, just as there are no witches. Round squares even necessarily do not exist, just like unicorns. And there are not even any conceivable

\textsuperscript{29} We better not go into the question of what distinguishes pencils from other writing utensils. It is only too clear that we could smuggle in hidden indexicality again by answering that it is their graphite leads.
contexts in which there would be round squares. "Round square" is simply not
indexical, and it is analytically true that they do not exist.

Fictional predicates too are among the analytically empty predicates, according to
our theory. If, e.g., for one of his novels Stanislaw Lem invents kurdels ???, gigantic
animals with phantastick inner thoughts and feelings, then it is equally analytic that there
are no kurdels. For, like "unicorn", "kurdel" is necessarily empty. Moreover, it is also
part of the English usage of "kurdel" that kurdels are fictional, i.e., that they form a
species that Lem made up for his science fiction story; their non-existence is not to be
discovered or rejected. There thus is no conceivable context in which the English usage
of "kurdel" could still refer to existing beings. But this means that "kurdel" is not only
empty in the actual context, but necessarily empty in all contexts, i.e., analytically
empty.

On a subjectivel level, however, things may be different. A reader could be
uncertain as to whether Lem reports his own personal experiences; and if kurdels are
unknown to him, now that he hears about them, he need not know that they are
creatures of fiction. Using the tools to be developed in section 3.9, we will also be able
to account for the subjective certainty or uncertainty about the existence of kurdels. But
this is quite consistent with there not being such open-endedness on the collective
level of the English-speaking community; I believe that there one can only say that "kurdel"
is analytically empty.30

An unbearable consequence of this is that all fictional predicates like "kurdel" or
"sandworm" are synonymous. I conclude that that our theoretical framework is still not
adequate for treating fictional discourse. This should not come as a surprsie, though.
For even our extensive discussions of possible contexts and possible indices with all
their phantastic cases are only there to give an account of how we, as subjects and as
linguistic communities, talk about reality and what we believe about reality. And it
would indeed not be right to treat the beliefs that cows are mammals and that kurdels
are mammals on a par, if the former is made under the assumption that cows really exist
and the latter is based on our knowledge that kurdels are fictional. The treatment of
fictional beliefs is not our topic; suffice it to say that our theory is, as it were, too
reality-biased to account for fictional discourse.

3.7 Hidden Indexicality and Rigid Designation

30 There could, of course, also be uncertainty on a collective scale. I have just presented unicorns as a
case in point where it is unclear whether our usage of "unicorn" rests on legends or on fairy-tales.
However, whether that account is correct, is itself not so clear. One would, rather, have to admit that the
borderline between fictional and non-fictional speech is vague and that "unicorn" just exemplifies this
vagueness. But none of that alters the fact that in our framework the clear cases of fictional and non-
fictional predicates must be accounted for in the way I have just done.
It was, of all people, its discoverer Putnam who associated several misleading ideas with hidden indexicality. In the previous section we have already criticized two wrong assumptions of his. But we have yet to address a more serious misconstrual. For up to now I have only talked of the hidden indexicality of predicates, thereby consciously avoiding any talk of predicates as rigid designators. However, in Kripke’s writings we only find the concept of a rigid designator, not that of hidden indexicality, even when he is going on about predicates. But Putnam speaks as if this is the same phenomenon:

"It should be clear, however, that Kripke’s doctrine that natural-kind words are rigid designators and our doctrine that they are indexical are but two ways of making the same point." (Putnam, 1975, p. 234)

It is thus important and instructive to finally clear up how these two concepts relate to each other.

Kripke initially introduced the concept of a rigid designator for individual terms, to wit in a purely intensional framework without any reference to characters; there an individual term $\alpha$ is a rigid designator if, and only if, $\alpha$ denotes the same object in each possible world (cf. Kripke 1972, p. 269). Within the character framework, we reconstructed this definition by saying that $\alpha$ is a rigid designator if, and only if, for any context $c$ for which $\ll \alpha \ll(c)(i(c))$ is defined, there is an object $x$ such that $\ll \alpha \ll(c)(i) = x$ for all indices $i$ for which this expression is defined (cf. p. VVV 28). As to whether $\ll \alpha \ll(c)(i)$ is defined if the object $x$ does not exist in $w_i$, Kaplan and Kripke give different answers. Kripke would say that $\alpha$ does not denote anything at such $<c,i>$. According to Kaplan’s understanding, on the other hand, a rigid or – as he prefers to call it – directly referential individual term is completely index-independent and, if it denotes any object at all in a given context, it denotes that object at every index $i$ even if it does not exist in $w_i$. We have not committed ourselves to either of these possibilities, because it would not have made a difference then (see, however, the discussion in section 3.3); but the matter will become relevant soon.

According to Kaplan, index-independence or rigid designation is the decisive feature of deictic expressions. He defines them as context-dependent and not index-dependent expressions, combining this with the empirical claim that simple indexical individual terms of natural language – such as "I", "you", "this", etc. are deictic in this very sense. The same is true of demonstrative descriptions, which Kaplan analyzes with the aid of the dthat-operator: $\ll V \alpha \ll(c)(i)$ was $\ll \alpha \ll(c)(i(c))$ and thus rigid. All these expressions were rigid, because the context already supplies the object denoted, thus leaving no room for any variation over indices.

The fact that, in the standard cases of indexicals, context-dependence always implies index-independence may have motivated Putnam’s claim that his thesis of the indexicality of natural predicates and Kripke’s rigidity thesis are two sides of a coin.

\[\text{Cf. Zimmermann (1991), p. 164, who makes the stronger claim that all lexical items are rigid or absolute, i.e., that lexemes can never be both context-dependent and index-dependent. If our analysis of predicates is correct, this thesis turns out to be untenable; we will soon see that some predicates are both context- and index-dependent.}\]
But it is by no means obvious whether this claim carries over from individual terms to predicates, or how it does so. What is obvious, rather, is that a direct application of the above definition of rigid designation to predicates does not make much sense: If one were to say that a predicate \( P \) is a rigid designator just in case for any context \( c \), there is an extension \( E \) such that \( \|P\|(c)(i) = E \) for all indices \( i \), then a rigid predicate would, of course, have to have the same extension in every index world. There are almost no predicates for contingent objects that satisfy this condition – and certainly the above examples from Putnam and Kripke do not.

The insight that the concept of a rigid predicate cannot be defined in this simple way has led to basically two alternative reconstructive attempts in the literature. According to the first one, due to Cook (1980), a predicate is rigid if, and only if, its intension represents an essential property. The second one takes it that the denotation of a predicate is not a set of individuals but an abstract object, viz. a kind or a property, and that a predicate is rigid if, and only if, it denotes the same kind or the same property in every world. This position has, e.g., been advanced by Donnellan (1983, p. 90f.), Boër (1985, p. 133f.), and Linsky (1984). Let us take a closer look at these alternatives:

Cook’s point of departure is the following account of Kripke’s rigidity, which is equivalent to the one given above: An individual term \( \alpha \) is rigid just in case all contexts \( c \) satisfy: if \( x = \|\alpha\|(c)(w) \) for some possible world \( w \), then \( x = \|\alpha\|(c)(w') \) for all possible worlds \( w' \) in which \( x \) exists. The immediate parallel is: A predicate \( P \) is rigid just in case all contexts \( c \) and all objects \( x \) satisfy: if \( x \in \|P\|(c)(w) \) for some possible world \( w \), then \( x \in \|P\|(c)(w') \) for all possible worlds \( w' \) in which \( x \) exists. Thus if \( P \) is rigid, no \( P \) can exist without being \( P \); that is, \( P \) is rigid if, and only if, in every context it expresses an essential property in the non-relational sense given on p. VVV 117. That this explication of rigid predicates matches Kripke’s intentions is also attested by the following quotation:

"It should be clear ... that ‘pain’ is a rigid designator of the type, or phenomenon, it designates: if something is a pain it is essentially so, and it seems absurd to suppose that pain could have been some phenomenon other than the one it is" (Kripke 1972, p. 337).

The adequacy of Cook’s generalization of the concept of rigidity also shows in the fact that, according to it, an individual term \( \alpha \) is rigid just in case the predicate "to be = \( \alpha \)" is rigid.\(^{32}\)

Boër (1985), p. 132, however, points out that one of Kripke’s crucial arguments does not remain valid under Cook’s account of rigidity: Kripke claims that the factual truth of identity statements among rigid designators already implies its necessary truth (and its factual falsity implies its necessary falsity). In the case of rigid individual terms, this is immediate. But Kripke (1972, p. 331ff.) repeatedly uses this argument as applying to predicates, too. However, on Cook’s reconstruction, it becomes invalid: if

\(^{32}\)Cook’s generalization does not work starting out from Kaplan’s rigidity, according to which an individual term is rigid if it denotes the same thing in all index worlds, no matter whether it exists there; it is at least not clear how this concept can be transferred to predicates in Cook’s fashion. Since we are going to follow Cook in the end, I take this to speak against Kaplan’s view and in favour of Kripke’s.
"P" and "Q" are rigid predicates in Cook’s sense and have the same extension (at c and) in w, it does by no means follow that they have the same extension (in c) in every w'. This can be made clear by the following reasoning:

Let x be an object that does not exist in w and that lies in the extension of "P" but not of "Q" in w'. This situation violates the conclusion that "P" and "Q" have the same extension in every w', while being compatible with the premisses: "P" and "Q" could well have the same extension in w and every object which (in c) lies in the extension of "P" or "Q" in some world may well lie in the extension of "P" or, respectively, "Q" (in c) in every other world in which it exists. A case in point may be "sponge" as a function term in the sense of "washing utensile" versus "natural sponge" as a biological predicate. It is not implausible to assume that every sponge is essentially a sponge and that every natural sponge is essentially a natural sponge. In some worlds – e.g. in ours until a while ago – both predicates are co-extensional. But they are obviously not co-intensional.

Before deciding whether this result speaks against Cook or against Kripke, we must take a look at the second account. Donnellan (1983, pp. 90-95), too, claimed that "P" and "Q" both being rigid predicates and having the same extension in the actual world is not enough for the identity statement "P is Q" being necessarily true. Unlike Cook, however, Donnellan takes predicates to be names for abstract entities, for substances, species, etc. – all of which Donnellan subsumes under the general concept of a kind; and, in analogy with individual terms, he defines a predicate as rigid if it denotes the same kind in every world.

If, however, one understands a statement of the form "P is Q" as being true (at c and) in a world w precisely when (at c) in w the specimens of kind P are also specimens of kind Q and vice versa, we again face the above problem: "P" and "Q" could rigidly designate two distinct kinds that only happen to have the same specimens in our world. In order to prove Kripke’s argument valid even for predicates, one would thus have to construe a statement of the form "P is Q" as a claim about the identity of two kinds or, as, e.g., Boër (1985) and Linsky (1984) propose, about the identity of two properties. Two kinds or properties would then have to count as identical if they are exemplified by the same specimens or, respectively, have the same extension in every world. Then the truth of "P is Q" and the rigidity of "P" and "Q" do indeed imply the necessity of "P is Q".

The principal reservation against this view has already been mentioned by its proponents (cf. Boër 1985, p. 130, Donnellan 1983, p. 91, and also Schwartz 1977, p. 38): this reconstruction of rigid designation does not match Putnam’s and Kripke’s distinction between natural and non-natural kind terms. For, according to it, not only predicates like "water", "tiger", or "lemon" would have to be regarded as rigid designators but also predicates like "bachelor"; if taken as a name of a property, "bachelor", too, has the same designation in every world, viz. being a bachelor. Not all predicates are rigid, though, as Linsky emphasizes; complex predicates like "colour of the sky" do not necessarily denote the same property in every world. But this is just not the way to get the classification that Putnam and Kripke were after.
Cook’s reconstruction, on the other hand, certainly captures the ontologically more interesting distinction. It thus seems to me to be the more sensible and, given our framework, only possible account of rigid designation.

However, even Cook’s account does not match Putnam’s intentions. For one thing, Putnam would call ”grow”, ”have measles”, etc. natural, indexial predicates, while they are not rigid; growing and having measles are not essential properties. For another thing, ”pencil” is likely to express an essential property – nothing which is a pencil could exist without being a pencil – but we above concluded that ”pencil” is probably not indexical.

But then this result is not at all surprising. If our above account is right, Putnam intends his indexicality to be a property of the characters of predicates; but rigid designation is a property of their intensions. It is thus clear from the outset that the two features do not coincide and that examples can be found for all four combinations of them.

These observations also show that the concept of rigid designation is not at all relevant to the description and explanation of the interesting phenomena that Putnam and Kripke are going on about and that we have discussed at length. The fact that identity statements like ”Water is H\_2O” and also implications like ”tigers are animals” are necessary without being analytic, the fact that such necessary statements can still be informative – all this is captured by our semantic rules and by applying diagonalization to them. But this has nothing to do with rigid designation but only with the fact that the predicates involved are hidden indexicals.

3.8 Predicates and the Problem of Internalism

As explained in section 3.2, our analysis of the characters of predicates developed so far makes a certain solution to the problem of informativity available; with it we can explain why some utterances that express necessary truths are nevertheless informative in English. For even if an utterance is necessarily true, the sentence uttered need not be an apriori truth. If it is not, its diagonal is informative – and so is the utterance, because we have identified its informational content in English with this diagonal.

In section 1.4 a complementary aspect of the problem of informativity was mentioned, which can also be accounted for in this way: how is it possible that necessarily false statements can be believed by rational persons? If someone seriously and honestly asserts: ”Water is not H\_2O”, we can say that the content of his belief – the subjective meaning that ”Water is not H\_2O” has for him – just isn’t the impossible proposition that the utterance expresses in the actual context, but only the diagonal of the sentence uttered; clearly, the sentence is not a priori false and hence its diagonal is informative, not contradictory.
On the other hand, in discussing Burge’s arguments in section 1.3, we have already seen that it may actually make sense to ascribe a person a belief in an analytically false statement without thereby accusing him of irrationality. This is something we can still not account for. For if John seriously and honestly claims "Jack has arthritis in his thigh" and if we take it that it is analytic in English that arthritis is an inflammation of the joints only, then this utterance not only expresses the impossible proposition; the sentence uttered is even an apriori falsehood of English, because not being a disorder of the thigh is an essential property of arthritis in all contexts. But then even the diagonal of "Jack has arthritis in his thigh" is contradictory; and if that diagonal is supposed to represent John’s belief content, we would be forced to accuse him of irrationality.

Analytic sentences too call our treatment of the problem of informativity in question. Although a sentence like "arthrits is an inflammation of the joints" is analytic and thus uninformative in English, it may well be informative for individual persons. We have already seen, in our account of Burge in section 1.3, that neither this case nor the previous one can be explained by assuming that the persons in question are no competent speakers of English who could therefore be left out of consideration; as we said there, a more extensive discussion can be found in Burge (1979, pp. 79-82). It cannot be denied, of course, that the grasp of language in the cases at hand deviates from the ideal norm; but such deviations do not license us to deny the persons in question the status of English speakers. However, as it stands, our definition of subjective meaning has no room for that kind of individual imperfectness of linguistic knowledge.

That it cannot solve the problem of internalism as arising in the cases discussed by Burge therefore does not come as a surprise; using the diagonals of the sentences uttered the doppelgangers’ belief contents in Burge’s gedankenexperiment cannot be described as being identical. For if, in the actual context, Fred says "I have arthritis in my thigh", he utters an English sentence; but if he makes that utterance in the counterfactual situation described by Burge, he produces a sentence of a different language in which "arthrits" is a general term for rheumatism and thus not for the same disease as in English. Clearly, not only do the propositions expressed by the respective utterances differ, but so do the characters and the diagonals of the sentences uttered; surely, in English the essential properties of arthritis are not the same as they are in the counterfactual language. But the internal identity between Fred and his his counterfactual counterpart, then, cannot be captured by the diagonal, because the latter only relates to the respective languages that Fred does speak; the fact that, in the actual context, his linguistic understanding diverges from the official one, thus becoming identical to that in the counterfactual context, is not taken into account.

Burge’s examples force us to conclude that we were wrong in reducing the concept of subjective meaning to the character function of a subject to his belief set and the

---

33 This is so whether or not we take "arthrits" to be a covertly indexical predicate. Given the considerations of the preceding section, we should, however, accept that "arthrits" is indeed a hidden indexical – because both "joint" and "inflammation" are indexical. But this point is irrelevant to the present train of thought.
language he actually speaks. The relation between objective and subjective meaning thus appears to be more complicated than we have so far been assuming.

However, before drawing any further conclusions from this insight, we will once more turn to Putnam’s examples; for they make it clear that our present solution to the problem of internalism not only fails because of the possibility of incomplete knowledge of a language but also in the ideal case of perfect command. To see this, it is important to keep different variants of Putnam’s Twin Earth story apart. First, we may – unlike Putnam\textsuperscript{34} – think of Twin Earth as a counterfactual counterpart of our Earth; here again, two cases must be distinguished:

In the first case we consider English at a time $t$ before the arrival of modern chemistry, and we assume that Oscar is not only a competent but even a perfect speaker of English$_t$ – by which we mean that all other speakers taken together have no better command of English$_t$ than him, i.e., that he is acquainted with all paradigm cases, that he has the best methods of identifying water available at time $t$ at his disposal, that he exactly knows the relevant essentiality convention, etc. Now, if Oscar utters the sentence ”This is water” in the actual world, he thereby expresses another proposition than he would if his utterance were to take place on counterfactual Twin Earth. We may, however, assume that he speaks the same language in that counterfactual situation as he does in the actual world, viz. English$_t$; and that thus the character of his utterance is the same in both situations. And since we had assumed Oscar to have a perfect understanding of English$_t$ and, in particular, of the word ”water”, it therefore seems to be intuitively adequate to describe the subjective meaning that ”this is water” has both for him and for his identical counterfactual counterpart in terms of the diagonal of this character.

However – still taking Twin Earth to be a counterfactual state of Earth – the second case, where we are looking at Oscar and English at the present time $t'$, is different. For then we would have to assume that English$_t$ did not exist on Twin Earth. The usage of ”water” in the language spoken by Oscar on Twin Earth would have to involve quite different criteria for finding paradigm cases of ”water”: portions of liquid could be called ”water” there if they behaved like XYZ under chemical analysis – and not, as in English$_t$., if they behave like $\text{H}_2\text{O}$. We could still, though, take Oscar to have perfect command of English$_{t'}$, but if we want to stick to the assumption that he is internally identical in the counterfactual situation, we are forced to say that he is wrong about the language he speaks there and that, in that sense, he does not perfectly master it. And, as expected, in this case the diagonals of the sentences uttered cannot adequately account for Oscar’s subjective meaning; the character of ”This is water” in English$_{t'}$ is different from the character that this sentence has in counterfactual Twin English$_t$.

So far it still appears as if the cases for which our present reconstruction of subjective meaning does not work can all be traced back to incomplete linguistic knowledge. But this impression is mistaken. For let us now turn to the third case, in which – as in Putnam’s original story – Twin Earth no longer features as a counterfactual state of Earth but as a different planet within the same possible world,

\footnote{See \textit{VVV} footnote 6 \textit{VVV} of section 3.2 above.}
resembling Earth in almost all respects. Let us keep in mind: Even if we consider Earth and Twin Earth at a time \( t \) before the arrival of modern chemistry, the character of "water" in English would still not be identical to the character of "water" in Twin English. The reason, already given in section 3.1 above, is that the intension of covertly indexical predicates in a language only vary with the possible world, but not with the language area. In context worlds in which Earthen English speakers have a different substance at the origin of their usage of "water" than Twin English speakers, the character of English "water" thus yields a different intension than does the character of Twin English "water". So if we imagine that some time in 1750 earthling Oscar uttered the sentence "there is enough water" and so did his Twin Earth counterpart, then Oscar's utterance would not only express a different proposition than the one expressed by Twin Oscar; the diagonal of the English sentence differs from the Twin English one's too. And it does not matter whether we take Oscar and Twin Oscar to have a perfect command of their respective languages; they do not speak the same language anyway – even if the usages of "water" including the relevant essentiality conventions are completely identical.

Apart from all this, English and Twin English would not even be the same language if, both on Earth and on Twin Earth, the paradigm cases of "water" consisted of \( \text{H}_2\text{O} \)-molecules. For as long as there is no communication between the two linguistic communities, there are always possible worlds conceivable where distinct substances formed the causal origin of the respective usages of "water"; and in these worlds the intensions of English and Twin English "water" would differ.

Even in this third case, the fact that subjective meaning cannot be adequately captured in terms of the characters of the sentences uttered is ultimately due to incomplete knowledge of language – albeit a very peculiar kind that has nothing to do with Burge’s cases and must therefore be kept apart from them. For, in a sense, even Oscar – who knows English perfectly in that he knows it as well as all other speakers taken together – does not really know English, because he does not really know which language English is – in the sense that worlds can always be imagined in which there are languages that for him are qualitatively indiscernible from English without being identical to English. For in order for a language to be English in another possible world it is not enough to contain the same syntactic expressions used in the same way; rather, it is also necessary for that language to have the same history as English has in the actual world; these properties of origin, which we have already described in some detail (and in all their vagueness) in section 3.1, are essential properties of English. In a strict, metaphysical sense, then, even a perfect speaker of English only knows his language if he knows these very properties of origin, or else if he knows – in the same strict sense – another object via which he can uniquely identify English. The most likely candidate for the latter would be he himself, English being the language that he speaks. However, in section 2.2 we have already extensively argued that nobody can be expected to know his own essential properties; and perfect speakers are even less likely to be acquainted with the essential properties of origin of their language. And, at any rate, even if in Putnam’s example we took Oscar and his twin to have such knowledge, we could no
longer claim that they were internally identical; their belief sets would then not contain the same worlds.

We can now draw the following conclusion: According to our analysis of predicates, the method of reconstructing subjective meaning of utterances by the diagonal of the character of the sentence uttered fails because knowledge of the characters of predicates must not be equated with the semantic knowledge competent speakers of the relevant language may be expected to have. For one thing, a competent speaker may have an imperfect knowledge of his language in quite an ordinary sense, without thereby being excluded from his language community; and, for another thing, there is a metaphysical sense in which he is bound to not know his own language – a sense that becomes relevant in the case of covertly indexical predicates. Hence Putnam’s Oscar turns out to raise the deeper variant of the problem of internalism than does Burge’s Fred. It seems that even the notion of a character function of a language is still too objective or metaphysical to be of use for an internal or individualistic concept of meaning. A modification of our theory is thus called for.

3.9 Formal Character and Subjective Meaning

Let us, at this point of our investigation, recall what originally motivated our view of subjective meaning. We started out from the observation that subjective and objective meanings of utterances normally do not coincide and explained this difference as due to the context dependence of objective meaning plus the incompleteness or even incorrectness of the subjective assumptions about the actual context. This idea has led us to distinguish two components in the interpretation of utterances by individual persons: on the one hand, the person’s semantic knowledge – which, in particular, consists in the knowledge of the character function of the language in question – and on the other hand his empirical beliefs, his assumptions about the nature of the context and the world in which it is located – which we intended to account for in terms of belief sets. Using the diagonal of the character of a sentence, the information contained in a sentential character could then be placed on a one-dimensional or propositional level. This, finally, made it possible to combine it with further empirical information by intersecting the diagonal with the belief set. We thus arrive at the following, rough picture: the more complete and correct the subject’s empirical assumptions about the context are, the more closely do subjective and objective meanings of utterances resemble each other.

However, if we now directly apply this view to predicates, then according to what we have said, subjective and objective meaning can only diverge in the case of indexical, i.e., context-dependent, predicates. The empirical component would then consist of beliefs about the properties that are essential for the property expressed by the predicate in the actual context; and, as explained above, the semantic component would cover knowledge of English in both a metaphysical and an ordinary sense. We have just
seen that this partition does not work: competent completely speakers never know the English language in a metaphysical sense, and they rarely do in the ordinary sense. Still, this diagnosis suggests a solution of the dilemma that falls back on our original motivation for a distinction between contextual knowledge and purely linguistic knowledge, but draws the line differently: an account of the problem of internalism for predicates by reconstructing the subject’s knowledge of his own language as part of his contextual knowledge, thereby reducing his semantic knowledge. I will now present this solution in detail.

In order to account for this reduced form of semantic knowledge, we will presently introduce the notion of a formal character of an expression in a language – to be distinguished from the earlier concept of a character, to which we will from now on refer as the objective character of the expression in that language. For the time being, the new concept will only be defined for (unary) predicates. Recall that, if $L$ is a language and "$P"$ is a one-place predicate of $L$, the objective character of "$P"$ in $L$ had been defined as:

$$\| P \|_L (c)(i) = \{ x \mid \text{in } w_c, x \text{ possesses the properties which, according to } L's \text{ essentiality convention for "} P \text{" in } w_c, \text{ are essential for those objects' being } P \text{ to which the } L\text{-predicate "} P \text{" is typically applied in } w_c \}.$$ 

Now, following the above suggestions, the formal character of "$P"$ no longer presupposes the true nature of $L$, but leaves it open. We must therefore not refer to $L$ as a given parameter but instead introduce a variable ranging over possible languages and fixed by the context. The formal character of "$P"$ is only supposed to contain the information that the extension of "$P"$ in context $c$ and at index $i$ depends on the context language $l_c$ in the way indicated – whatever $l_c$ may look like. This information is the minimal strutral semantic knowledge every competent speaker must be taken to have with respect to the predicate "$P"$. Since the context language $l_c$ can be defined as the language that $s_c$ speaks in $w_c$ at time $t_c$, we need not introduce a new contextual parameter here; for reasons of readability, we will still frequently refer to the language spoken by $s_c$ in $w_c$ at $t_c$ as $l_c$.

Thus the formal character of a predicate "$P"$ no longer treats "$P"" as an expression of a fixed historical language $L$ with a given meaning. On the other hand, we do not want to define the formal character of mere noises. We thus assume that we are talking about syntactically identified expressions, where "$P"$ is part of a fixed syntactic structure $S$ and classified as a unary predicate.

So the picture we are finally arriving at will look like this: A concrete historical language $L$ is characterized by various features. As a first component, there is its syntactic structure $S$. A second component is its formal character function defined on all expressions which are well-formed according to $S$ and describing, as we are going to

---

This description is still somewhat sloppy; how precisely it is to be understood will be addressed on p. VVV 143.
put it, the structural semantic properties of \( L \). The third component of \( L \), which may be dubbed the content component, is formed by the already extensively described ‘usages’ associated with \( S \)’s lexical items.\(^{36}\) The fourth and final component of \( L \) is its historical situation, the spatio-temporal location of \( L \). As is clear from various examples above, the first two components do not determine the third one, and neither do the first three determine the fourth; only all four components together fix the objective character of \( L \).

Any complete knowledge of \( L \), whether in an ordinary or a metaphysical sense, would have to include a complete knowledge of the components of content and location. We have, however, seen that neither must be taken for granted. We will therefore only assume speakers of \( L \) to master the first component of their language, its syntactic structure \( S \). Their structural semantic knowledge can then be described by the formal character function for \( S \). And all knowledge about the other components will be subsumed under the speakers’ contextual knowledge.

The definition of the formal character of a unary predicate ”\( P \)" in a syntactic structure \( S \), then, reads:

\[
\ll P \ll^S(c)(i) = \{ x \mid \text{in } w_c, x \text{ possesses the properties which, according to } l_c \text{ 's essentiality convention for } "P" \text{ in } w_c, \text{ are essential for those objects' being } P \text{ to which } "P" \text{ is typically applied in } l_c \text{ in } w_c \text{ – where } l_c, s_c \text{’s language in } w_c \text{ at } t_c, \text{ has } S \text{ as its syntactic structure}\}.
\]

Instead of the "properties which, according to \( l_c \) 's essentiality convention for "\( P \)" in \( w_c \), are essential for …" we will simply speak of the "properties that are \( l_c, w_c \)-essential for …"; and, as before, we will abbreviate the locution that in \( w_c "P" \) is typically applied to something in \( l_c \) by saying that in \( w_c \) something is called "\( P \)" in \( l_c \). The diagonal of the formal character of, say, "This is water" in the syntactic structure \( S(E) \) of English can now be given the following formulation:

\[
\partial^{S(E)}(\text{this is water}) = \{ c \mid \ll \text{this is water} \ll^{S(E)}(c)(\iota(c)) = 1 \} = \{ d \mid \text{in } w_c, \text{ that which } s_c \text{ is pointing at in } w_c \text{ at } t_c \text{ has the properties that are } l_c, \text{ } w_c \text{-essential for those objects' being water that are called "water" in } l_c \text{ in } w_c \}\).
\]

Thus, on the level of formal character, the information the sentence ”this is water” conveys to a subject \( s \) is still so unspecific that it allows for contextual languages with

\(^{36}\) Up to now we have only talked about usages within a definite, spatially and temporally located language; but now the concept of a usage must be understood as abstracting from location, in the sense that, e.g., – before the discoveries of the respective chemical structures – the English usage of "water" is the same usage as the Twin English usage of "water".

\(^{37}\) The description "that which \( s_c \) is pointing at in \( w_c \) at \( t_c \)" must obviously not be taken as a general proposal for interpreting "this", but merely as an ad-hoc-interpretation serving the present example.
all kinds of semantic properties; contexts are, e.g., possible for $s$ where, in his language, "water" refers to beer, or to pencils. However, this changes as soon as the information in the belief set is taken into account; for the latter is supposed to contain, among other things, $s$’s assumptions about the semantic properties of his actual language. It will thus be part of the beliefs of most speakers of English that, in their language, the prototypical cases of "water" just aren’t long wooden sticks with graphite leads or brownish, slightly bitter portions of liquid but, rather, transparent and tasteless ones. Accordingly, contexts in which $l_c$ does not have these properties are not possible contexts for such speakers.

A person’s belief set is thus made to represent, among other things, the result of his language acquisition. Behind all this is the following, extremely simplified picture (which we only describe for the case of predicates):

There are two ways in which a person may acquire semantic knowledge about the expressions of the language spoken in his language community. For one thing, he learns by ostension: other members of the language community apply a predicate "$P$" to objects or events that he perceives, and he forms a concept by abstracting from his perception, associating it with the hypothesis that "$P$" stands for just this concept. So the person acquires the belief that the objects called "$P$" in his language have certain perceptable properties. As a rule, the subjective concept of being called "$P$", as developed by this person, will then not be the property objectively expressed by "$P$".\(^{38}\) For some of the predicates that are not hidden indexicals, though – as in the case of colour predicates like "being red" – it does not appear implausible that, after a sufficient number of application samples, the subjective concept does coincide with the objective one.

The other way for a person to acquire and extend a concept of being called "$P$" is language-internal: he hears and interprets utterances in which "$P$" occurs and is related to other expressions. If, e.g., he hears "$Ps$ are $Qs$", he may form the belief that the objects that have property so-and-so and are called "$P$" in his language, are also objects that have property so-and-so and are called "$Q$" in his language. In this fashion, a complex internal conceptual network emerges.

Although, in the present context, we are only interested in linguistic concepts, they are not the only ones. Humans also master concepts they have acquired by perception or in interaction with the corresponding objects, without having been supplied a name for them; and, surely, animals without language cannot be denied concepts either. For

\(^{38}\) Strictly speaking, the subjective concept of being called "$P$" can never be identical with the property objectively expressed by "$P$". For, first, subjective concepts correspond to diagonal intensions, and thus functions from contexts to extensions, whereas objective properties correspond to ordinary intensions, i.e., functions form indices to extensions. Second, it is not part of the objective property that the objects to which it applies are called "$P$"; it is, after all, also defined for indices at which the language to which "$P$" belongs does not even exist. In a looser sense, a subjective and an objective concept can still coincide. First, every context corresponds to an index, and thus to every property, i.e., function on indices, there corresponds a subjective concept, a function on contexts. Second, the objective concept can be restricted to the subject’s belief set; and within the latter, the subjective concept and the function on contexts determined by the objective intension can approximate each other and, in principle, even be identical.
instance, a person can have a concept of a mirabelle without knowing that, in his language, such objects are called "mirabelle". As a rule, concepts of this kind are merely subjective concepts; the person masters the objective concept of a mirabelle as little as he masters the objective meaning of "mirabelle" (in English, in the actual world), precisely because he does not know the essential properties of mirabelles, with or without that term.

At this point the question naturally crops up as to what entitles us to call such a subjective concept a concept of mirabelles. Why shouldn’t it rather be a narrower concept of yellow or ripe mirabelles, or a wider concept of mirabelles-or-greenages?

This question has been extensively discussed by, among others, Fodor (1987, chapter 4), under the heading of the "disjunction problem". But it is not our problem here. It is only too clear that we cannot give an exact account of a person’s internal concept or his belief set; in either cases, we would have to be able to precisely model the distinctions made by the subject. When characterizing a concept as a concept of mirabelles, we just make a rough de re ascription from the outside, guided by the fact that the concept has primarily been formed on the basis of ripe or unripe mirabelles, but not of greenages.

It is also clear that such non-linguistic concepts form the indispensable foundations underlying all linguistic concepts. To be sure, some concepts may have been formed on purely innerlinguistic grounds; most speakers, for example, will never have even made remote contact with plasma and therefore have formed their concept of plasma only via books. But this cannot be the case for all concepts. Indeed, in ostensive learning the linguistic and the non-linguistic concept are formed together, as one single concept.39 That two things are brought together here becomes particularly apparent if one bears in mind that it is very well possible for one person to have both a subjective P-concept and a subjective concept of being called P without merging the two; it may well have happened that he has, on the one hand, made P-experiences and formed the relevant concept after them, and, on the other hand, much later built a concept of being called "P" on purely language-internal grounds, and maybe without making the connection between the two.40 – So much for the discussion of our view that a person’s belief set also represents his linguistic development.

It must be emphasized that our partition of competent speakers’ knowledge of meaning into the two components just described – the formal character of syntactically structured expressions and assumptions about their usage by the language community, as represented in the belief set – no longer contains a category corresponding to what is traditionally called semantic knowledge. The formal character only codes competent speakers’ structural knowledge, and any attempt to distinguish between semantic and empirical assumptions within the belief set appears to be neither promising nor theoretically fruitful. The very notion of a competent speaker, too, is granted a large range of vagueness this way. Whether one would want to say that a person knows the

39 Quine has repeatedly pointed out this double induction involved in ostensive learning; see, e.g., Quine (1974), §§ 10ff.

40 Nida-Rümelin (1993) describes such cases involving colours and colour terms and draws strong conclusions about the mind-body-problem.
meanings of the expressions of his language now wholly depends on how strict the requirements are that one would want to pose on the beliefs about the language as contained in the respective belief sets. This vagueness already pertains to the notion of a stereotype as used by Putnam in describing individual linguistic competence.

Losing a separate theoretical component of semantic knowledge in the traditional sense, however, does not strike me as an unwelcome result but, rather, as an inevitable consequence of recent ideas about the theory of meaning, as they can be found in the writings of Putnam, Kripke, Burge, and others; moreover, it is certainly also a reflection of Quine’s insistance on the inseparability of language and theory – even if Quine’s framework is otherwise hardly comparable to ours.

It must also be stressed that, with the concept of a formal character, we have finally arrived at the reconstruction of Stalnaker’s propositional concept announced in section 2.1. It is the objective character of a given language that Kaplan is exclusively concerned with. That Stalnaker’s propositional concept, like our formal character, also allows for varying languages has already become clear in our account of the relevant examples in section 1.4. And, according to my interpretation, after all the other modifications announced in section 2.1 have been done, this is the only point in which Kaplan’s character differs from Stalnaker’s propositional concept. However, our reconstruction of the propositional concept cannot quite be in accordance with Stalnaker, because we are going to use formal characters in our solution to the problem of internalism, whereas Stalnaker (1990) explicitly denies that his propositional concepts clarify this problem.

What, then, does our solution to the problem of internalism look like? We have, basically, already given the crucial ingredients. We may go on reconstructing an subjective meaning of utterences via the diagonal of the character of the sentences uttered – as long as we no longer refer to the objective character but to the formal character instead. If we now consider this diagonal on the background of the respective belief sets, we get exactly the internal identities we were after in Burge’s and Putnam’s examples:

If Fred, as a competent speaker of English, seriously and honestly utters the sentence “I have arthritis in my thigh”, then the diagonal of this sentence’s formal character in the syntactic structure $S(E)$ of English can be given the following, simplified account:

\[
\{c \mid \exists s_c \in \text{have arthritis in one’s thigh} \upharpoonright_{S(E)}(c)(i(c)) \}
\]

\[
= \{c \mid \text{in } w, s_c \text{ has that disease in his thigh which is called ”arthritis” in } l_c \text{ in } w_c \}.
\]

Exactly the same diagonal results for the sentence ”I have arthritis in my thigh” if uttered by Bill in another possible world and in another language $E’$ with the same syntactic structure as English but a slightly different usage of ”arthritiis”. Since we are following Burge in assuming that Fred has the same physical make-up in the counterfactual world that he has in the actual world, and since we further argue under the assumption that such sameness also implies a sameness of his beliefs in an internal
sense, we may therefore take it that his belief set is the same in the actual context \(c\) and in the counterfactual context \(c'\). In particular, both in \(c\) and in \(c'\) does it contain the belief to be someone whose language is such that typical cases of "arthritis" not only cover inflammations of joints but, e.g., also of painful diseases of the thigh. We therefore clearly also have: 

\[
B_{Fred,c} \cap \partial^{S(E)} (I have arthritis in my thigh) = B_{Fred,c'} \cap \partial^{S(E)'} (I have arthritis in my thigh);
\]

as desired, then, our theory ascribes the same subjective beliefs to both actual Fred and his counterfactual counterpart.

Putnam’s examples can be treated in a quite parallel fashion, and no matter whether we consider Twin Earth as an actual planet or as a counterfactual counterpart of our Earth. If Oscar and Twin Oscar utter the sentence "There is enough water", then in either case the diagonal of its formal character is – put rather briefly – the set of all contexts \(c\) in whose world \(w_c\) there is enough of that which in \(w_c\) is called "water" in \(I_c\). And since Oscar and his twin are supposed to be internally identical, we may again take their belief sets to be the same; in particular, both of them make the same assumptions about the usage of "water" in their respective language and about its spacio-temporal location. Thus, again, Oscar’s and his twin’s subjective beliefs turn out to be the same.

An attempt to meet the problems brought up by Putnam and Burge and to regard beliefs as internal psychological states, in order to be compatible with Stich’s (1978) autonomy of psychology, has also been made by White (1982). This is how his strategy compares to ours:

With each use of a word by a given speaker White associates a token of a functional state, where functional states are internal psychological states. A possible context of acquisition for a word is a pair consisting of a possible world plus a state token occurring in that world and associated with the word. An interpretation of a word is a set of possible contexts of acquisition for that word such that their state tokens are all functionally equivalent with each other, i.e., of the same state type; one may say that each interpretation of a word represents a possible subjective meaning. A possible context of acquisition for a word then determines the character of that word in Kaplan’s sense in that the possible world of the context provides the state token of the context with all those causal connections with the social and natural environment of the token that, according to the causal theory of reference, determine the character of that word. Finally, the partial character of a word relative to an interpretation of that word, is a function that assigns to each context of acquisition in this interpretation the character of the word it determines. How this conceptual apparatus is supposed to work is explained in detail in sections III and IV of White (1982).

White’s character corresponds to our objective character. But then White does not regard natural kind terms like "water" as indexical (cf. p. 356f.); he only localizes their hidden indexicality at the subjective level of partial character and not, as we have done, already on the linguistic level of objective character. Apart from that, his step from non-internal characters to internal partial characters corresponds to our abstraction of formal character from objective character; for the formal character is to represent the subjects’ internal semantic knowledge and fixes the objective character because the subject’s language \(I_c\) in \(w_c\) is a specific natural language \(L_c\).

Nonetheless, there are two further differences. For one thing, according to White, internal beliefs do not appear to be propositional attitudes; a subject’s belief that \(\phi\) and another subject’s belief that \(\psi\) are only internally identical if \(\phi\) and \(\psi\) have the same partial character. However, on our account, even the objects of internal belief are propositional – if one is liberal enough to categorize sets of contexts as propositional. This difference rests on the fact that the concept of diagonalization, which is crucial for us, plays no role for White.

For another thing we have a different way of modelling subjective meaning. According to White, for each word, it simply consists in a functional state type or an interpretation on which its partial character is then defined. However, our formal character is perfectly general and vacuous and does not determine any interpretation in White’s sense. We may, though, capture the subjective meaning of a word by restricting
Our concept of subjective meaning and the concept of a formal character raise two problems analogues of which we had already had to clear up in connection with the introduction of objective character. The first one concerns the domain of the formal character function. In our discussion of the objective character of an expression \( \alpha \) in a language \( L \), we saw that it is only defined for contexts in which \( L \) exists (cf. section 3.3). Analogously, we must now assume that the formal character of an expression \( \alpha \) in a syntactic structure \( S \) is only defined for such contexts in which \( S \) is realized, i.e., where there is a language \( L \) with a syntactic structure \( S \). Without this assumption the definition of a formal character would not make sense because – as in the case of the objective character – we cannot resort to a counterfactual interpretation. If there is no language \( L \) with structure \( S \) in \( w_\alpha \), then \( c \) contains no information as to what the semantic properties of such a language would be in \( w_\alpha \), if it existed in \( w_\alpha \).\(^{42}\)

Apart from that, we will again – as in the case of the objective character – construe the definition of formal character in the sense that \( \alpha \) need not really be uttered in \( c \); rather, \( \text{lloll}^5 (c) \) yields an intension that \( \alpha \) would have had if it were uttered by \( s_\alpha \) in \( w_\alpha \) at \( t_\alpha \) – and \( s_\alpha \) would sufficiently master the instantiation \( l_\alpha \) of \( S \) in \( w_\alpha \). The reasons behind this view are the same as the ones given in our discussion surrounding the objective character: otherwise there would be no way of defining sensible notions of logical consequence or synonymy on the level of subjective meaning.\(^{43}\)

---

\(^{42}\) This is even so if \( \alpha \) happens to be absolute in \( L \) and its objective character is thus also defined for contexts in which \( L \) does not exist. For that \( \alpha \) is absolute is a fact about \( L \) itself and cannot be read off its syntactic structure \( S \).

\(^{43}\) This would, of course, require a more precise account of how to transfer the concepts of logical consequence, synonymy, analyticity, aprioricity, etc. to the levels of formal characters and subjective meanings. However, that much is clear: If we require the formal and the objective characters to coincide at least in the case of logical constants – i.e., if only he who knows the relevant semantic properties of at least these words of his language counts as a competent speaker – we get the usual logical consequences on the formal character level – provided that we define \( \psi \) to be a logical consequence of \( \phi \) just in case \( \psi \) is true at the designated context of each model in which \( \phi \) is true; if, however, we were to define the formal character of \( \alpha \) as a partial function that is only defined for those contexts where \( \alpha \) is actually uttered, such a definition of logical consequence would not be possible.
However, it now turns out that the definite description "the language with syntactic structure $S$ and spoken by $s_c$ in $w_c$ at $t_c$", which featured in the formal character function – and which we had abbreviated by "$l_c$" – is not only meaningless if there is no language with syntactic structure $L$ in $w_c$. Difficulties also arise if there are two or more instantiations of $S$, as in the story about Earth and Twin Earth. Two cases must be distinguished here. The unproblematic case is that in which $s_c$ happens to speak one of these instantiations in $w_c$ at $t_c$; for then $l_c$ just denotes that instantiation. But the case in which $s_c$ does not happen to have command of any of the instantiations of $S$ in $w_c$ is problematic; for, again, the question of which of these languages $s_c$ would speak in $w_c$ has no clear answer. It must therefore be noted that the formal character $||S||$ is only defined for contexts $c$ for which we have: (i) there is exactly one instantiation of $S$ in $w_c$ at $t_c$, or (ii) if there is more than one instantiation of $S$ in $w_c$ at $t_c$, then $s_c$ masters exactly one of these instantiations in $w_c$ at $t_c$.

The second problem for our account of subjective meaning concerns its metalinguistic nature. As far as objective character goes, we had been able to meet the relevant objections in section 3.4. But we left open the question of how the individualistic variant of the translation argument could be countered. Having given a new account of subjective meaning, we must finally face this question.

The relevant version of the translation argument was this (cf. p. VVV 107): If John – a speaker of English without any knowledge of German – seriously and honestly utters sentence (5), one would intuitively want to ascribe the same belief to him as to German speaker Hans – who himself does not speak any English – when he utters sentence (6):

(5) I like drinking water,
(6) Ich trinke gerne Wasser.

According to our theory of subjective meaning, however, the two speakers can never have the same belief set; for the set (5a), the diagonal of (5), is only a superset of John’s belief set, but not of Hans’s, and vice versa with the set (6a), the diagonal of (6) (where we have given a simplified account of the analysis of ”water” and, for ease of exposition, ignored those of ”trinken”, ”drink”, ”gerne”, and ”like”):

(5a) \{ $c \in w_c, s_c$ likes drinking that which is called ”water” in $l_c$ in $w_c$ \},
(6a) \{ $c \in w_c, s_c$ likes drinking that which is called ”Wasser” in $l_c$ in $w_c$ \}.

Unless it coincides with logical equivalence, synonymy is practically non-existent on the formal character level. But then we would still like to be able to define a notion of synonymy on the subjective meaning level. Thus, one may suggest that two expressions $\alpha$ and $\beta$ are synonymous for a subject $s$ if, and only if, the restriction of $\alpha$’s formal character to $s$’s belief set coincides with the restriction of $\beta$’s formal character to $s$’s belief set: this would then mean that $s$ takes his language to be one in which $\alpha$ and $\beta$ have the same objective character. This definition, too, would not be possible if the formal characters of $\alpha$ and $\beta$ did not have the same domain.
Hans thus ascribes to himself the property of liking to drink what is called "Wasser" in his language, whereas John ascribes to himself the property of liking to drink what is called "water" in his language; and these two properties are obviously distinct.

It is, however, not clear whether it is intuitively desirable to ascribe the same belief set to Hans and John, i.e., to really regard them as internally identical. Strictly speaking, anyway, speakers of English and of German – unlike the twins in Putnam’s and Burge’s stories – can never be internally identical, simply because they deal with different phonetic utterances in their different languages. So the intuition behind the translation argument seems to be concern the level of de dicto ascriptions rather than that of internal belief content.

In section 1.3 we have already made clear that de dicto ascriptions are not internal ascriptions; this was the inevitable consequence from the examples given by Kripke, Putnam and Burge. In our framework, we may adopt the following truth-conditions for de dicto belief ascriptions:

A sentence of the form "a believes de dicto that φ" is true if there is a translation φ' of φ into a’s actual language such that the diagonal of φ’’s formal character is a superset of a’s belief set.

This definition is motivated by the usual assumptions about de dicto ascriptions: it satisfies the extended disquotational principle and it allows for substitution of synonymous expressions – by which we here mean: expressions with the same objective character. Since (5) and (6) are translations of each other and thus, as explained in section 3.4, have the same objective character, it turns out that, according to our theory, John and Hans have the same belief in a de dicto sense. The individualistic version of the translation argument thus again misses its target.

But maybe there is an intuition beyond de dicto according to which John and Hans are internally identical in a certain, if not quite so strict, sense. One may have the possibility in mind that John and Hans have exactly the same ideas and beliefs about the world and its objects and that they only differ in their beliefs about their respective languages; if one could abstract from the latter, they would otherwise have to consider the very same worlds or contexts as possible. Without further ado, this kind of internal identity can obviously not be captured on our approach. And I do not want to decide here whether it would at all be desirable to account for it. It is, however, clear that such a reconstruction would have to presuppose a way of meaningfully distinguishing between speakers’ linguistic and non-linguistic beliefs. An argument against our conception could thus only be constructed if this possibility was proved to obtain and if, moreover, our framework was shown to be unable to account for this proof.
Chapter 4

Proper Names

In the preceding chapter we have seen that the problem of informativity largely vanishes once we take Putnam’s observation seriously that predicates may be hidden indexicals. However, in order to solve the remaining aspects of the problem of informativity as well as the problem of internalism, we had to pass on to an even more abstract notion of character, according to which the intension of predicates does not only depend on the world, but also on the language of context. The diagonals of such characters then not only model incomplete world knowledge but also the subjects’ incomplete linguistic knowledge; these two aspects turned out to be crucial for the concept of subjective meaning.

We would now like to investigate how the strategies developed so far can be put to use in a semantic account of proper names. In section 1.1, we have already discussed both Kripke’s main thesis that names are rigid designators having constant functions as intensions and the problems it poses for a theory of subjective meaning. If we want to solve these problems by way of diagonalization, as we did in the case of predicates, we can only succeed if proper names are indexical too, with characters that are not constant. This is the question to be settled here.

To begin with, there is the undeniable fact that every name can be the name of not only one, but many – and in principle: arbitrarily many – objects; there are lots of Johns and, depending on the context, “John” denotes different individuals. It is initially plausible to conceive of this fact as one kind of indexicality of names: which object a name \( N \) refers to depends on who utters \( N \) at which place and time, and to whom and with which intentions. This view has, e.g., been expressed by Burge (1973), who treats names as demonstratives or, in the terminology introduced earlier, as open indexicals; according to him, "John" must be read as "that John". But this view is controversial; Kaplan (1977, section XXII) outright rejects it, defending the thesis that the multiple usage of names is a form of systematic ambiguity or homonymy. He argues that uses of "John" with different denotations are not different usages of one name but of different homonymous names.

A detailed discussion of the question of whether names are overtly indexical or ambiguous in detail will be given in sections 4.5-7. There we will find that it is surprisingly intricate and that there is reason to answer it either way; the final answer can only be given within a a general semantics of noun phrases which we will, however, not go into. But let us, for the time being, postpone this question. In any case, we will have to find out whether names are also hidden indexicals. For if we follow Kaplan and regard names as ambiguous, our strategy can only help explain the informativity of, say, "Hesperus = Phosphorus" by a hidden indexicality in these names. And if we agree with Burge, the open indexicality of names alone cannot be enough.
For as we will see in section 4.6, only because a lot of other objects – like my cat or my electrician – are or can be called "Hesperus" or "Phosphorus" – does this open indexicality induce informativity of "Hesperus = Phosphorus" via diagonalization; and this will turn out to be an unattractive account of the problem of informativity.

I will therefore first ignore the question of whether names are to be treated as open indexicals or as ambiguous expressions, making the provisional assumption that each name has only one usage and denotes at most one object. Given this assumption, the other question, viz. whether names are hidden indexicals, is easier to discuss; and the positive answer I am going to give will be easier to explain.

4.1 The Character of Names and Their Hidden Indexicality

In his "Naming and Necessity" (p. 298-303), Kripke goes some way towards drawing a positive picture of how the reference of proper names is determined:

"A rough statement of a theory might be the following: An initial baptism takes place. Here the object may be named by ostension, or the reference of the same may be fixed by a description. When the name is ‘passed from link to link’, the receiver of the name must, I think, intend when he learns it to use it with the same reference as the man from whom he heard it.” (p. 302)

Starting with a kind of baptism in which the name N is attached to an object x, a causal network of uses of the name N thus develops; and these uses refer to x, at the origin of the network, precisely because they all partake in that causal network. In other words, each new use of N is referred back to the initial baptism situation via at least one causal chain in that network, and it therefore denotes this object, x. Kripke speaks of a causal chain of communication in this connection; Devitt (1981), p. 29, uses the narrower term "designating-chain” to stress the specific nature of such causal chains; Lerner and Zimmermann (1984), p. 18, speak of traditions (of names). I prefer to go parallel with my terminology for predicates and will from now on speak of the usage of N when relating to a name N, or, more precisely, of the English usage of N when talking about the English language1; what is behind this terminological shift will become clear presently. The gist of Kripke’s causal theory of reference, then, is this: The referent of the English name N is the origin of the English usage of N (given the

1 It may appear that names that have been introduced by one linguistic community are almost automatically adopted by the other linguistic community so that names are common property not pertaining to any particular language. This is indeed often the case, but there are also clear counter-examples: Germany’s most westerly city is called “Aachen” in German; in English it is also sometimes called “Aachen”, but usually it is called ”Aix-la-Chapelle” – which it is never called in German. So ”Aix-la-Chapelle” is a name that belongs to English but not to German. It is not only because we are interested in the semantics of a single language, that names are regarded as parts of particular languages, but also because of such examples.
above provisional assumption, there is only one English usage of \( N \); and since names are rigid designators, \( N \) refers to this origin not only in the actual world but also in all counterfactual situations or index worlds in which that origin exists.\(^2\)

As far as the exact nature of this usage of \( N \) is concerned, two conditions can already be found in the above quotation: Every subsequent use of \( N \) must be partly caused by an earlier use of \( N \) in a specific way; and every subsequent use of \( N \) must be made with the intention to use \( N \) with the same referent as it had in earlier uses. Such locutions are obviously in need of further explanation; accordingly, intensive attempts at more precise accounts have been made.\(^3\) We will, however, not follow these attempts in detail. For the task of giving a detailed account of the causal intentional complex that makes up the German usage of \( N \) is again part of the empirical theory of reference, which must be distinguished and can be separated from our proper goal, recursive semantics (cf. VVV p. 94).

Our present problem is different – viz.: What does Kripke’s causal theory of reference of proper names say about their characters? Or, in other words: Which rule of context dependence of proper names can be derived from the causal theory of reference? The answer is quite straightforward.\(^4\) Kripke, though, only talks about the actual context; he did not yet seem to be aware of the need to consider other possible contexts.\(^5\) But he does talk generically, as it were, about the actual context so that his claims may be taken to relate to all possible contexts. If we now transfer what we have just called the gist of Kripke’s theory to all context worlds \( w_c \), we then immediately obtain a definition of the objective character of a proper name \( N \) in English:

\[
\| N \|_E (c) (i) = \text{that object which is at the origin of the English usage of } N \text{ in } w_c \\
\text{(if it exists in } w_l)\]

I am going to keep to this definition; it will allow for a reasonable solution of the information problem as applied to proper names. However, in order to make it plausible, more detailed explanations will have to be given. In this section I want to give a more precise account of the English usage of \( N \), thereby explaining how the definition implies a hidden indexicality of \( N \). In section 4.2 it will then have to be investigated how the definition relates to problematic kinds of names, viz. empty ones and descriptive ones. In section 4.3 I am going to discuss whether names can be assigned a descriptive content on the basis of our semantic rule, and I will once more

---

\(^2\) This does, of course, not mean that, in the index-worlds, this origin must exist as the origin of a usage of \( N \); it can also exist there without being talked about.

\(^3\) The principal references are Devitt (1981), chapters 2 and 5, and Evans (1973) and (1982), chapter 11. Kaplan’s most detailed account is in his (1990).

\(^4\) Still, the only place where I have found the following step made explicit is Lerner and Zimmermann (1984).

\(^5\) He does frequently talk about what may have turned out to be the case, and hence about epistemic possibility. But the idea of representing epistemic possibility via possible contexts only originated with Kaplan (1977) and Stalnaker (1978).
refute the meta-linguistic objection and related suspicions of circularity. Only then, in section 4.4, can we return to the problems of informativity and internalism.

But let us first take a closer look at the above definition of the characters of proper names. The first observation is that, according to it, proper names are rigid designators. This just follows from the fact that the definiens does not (or only parenthetically) depend on \( w_i \).

It is, furthermore, clear that our arguments about the domains of characters, given in section 3.3, can be directly taken over. So the character of \( N \) in English is only defined for such contexts in which the English language, or at least the English usage of \( N \), exists. At that point the question cropped up whether this definition of the functional domain implies that the denotation of \( N \) exists in every context, thus rendering the sentence ”\( N \) exists” an English a priori. In the case of logically proper names, which are absolute, we had conceded and, against Kaplan, also accepted this consequence. In the case of ordinary proper names, however, this consequence would be fatal. It will not arise, though – as we have said – if ordinary proper names are hidden indexicals. We are now going to pursue this crucial point.

In order to do so, we will have to further clarify the nature of the English usage of \( N \). Given the above definition, the target of explanation has moved, however. Before, we were after a sharper version of Kripke’s rather sketchy picture, a more precise account of what the English usage of \( N \) actually is. Now we are facing the task of explaining more precisely what is to be understood by the same English usage of \( N \) in different possible contexts. These are two different tasks. For, surely, not all causal and intentional details of the actual usage of \( N \), and not even the participation of all those persons that are actually involved in it, are essential for the existence of the English usage of \( N \). So the description of these details is only relevant in mastering the first task, not for the second one.\(^7\)

The motivation behind my terminological change can also be found here. After all, for a causal chain or a causal network, all causal connections therein are essential; if one were to change anything, one would get another chain or another network. Talking of causal chains or networks in the above character definition would therefore have induced wrong associations. It is thus more appropriate to speak of traditions or usages in that definition; they are less tied to the specific succession of events.

What, then, in the sense of our present, second task, is to be understood by the English usage of ”Aristotle”? In the case of ”water” we had answered the corresponding question in section 3.2, when we discussed the identity conditions of the causal-

---

\(^6\) We are, of course, talking about the individual \( N \), not about the name \( N \). Here it becomes clear that normally I am using ”\( N \)” autonomously, designating the name ”\( N \)” itself, and only occasionally do I use it to talk about the denotation \( N \) of the name. No misunderstandings should arise out of this.

\(^7\) This change in perspective, too, have I only found explicitly in one place, viz. Kaplan (1990). There Kaplan develops a rather different conception of words, which he dubs the ”common currency conception”. According to it, words, including names, are not phonetic or graphical symbols that materialize in utterances but rather something which is at least reminiscent of our usages. Towards the end, on p. 118f., he asks questions he takes to be legitimate and important about the essential properties of words and names in his sense; he does not arrive at any positive claims, though.
intentional complex of the English usage of "water". We may now adopt these findings here, with a few minor changes.

As to water, we said that the English language community (at a given time) has certain accepted methods of identification at their disposal; in a civilization of experts, like ours, these are going to be the scientifically most advanced methods developed by our water-experts. It is these accepted methods that decide what does and what does not count as a typical case of applying "water". As we have just made explicit in section 3.5, the English essentiality convention for "water" is then used to extrapolate from these typical cases the extension of "water" in all index worlds. Apart from the essentiality convention, then, the English usage of "water" chiefly consists of accepted methods of identifying water. The point was that the typical cases, the causal base of the English usage of "water", were neither part of nor uniquely determined by this usage; different substances with different essential properties may equally well pass the identification tests, thereby being indistinguishable from water for us. This was the key to our solution of the problem of informativity as applying to predicates.

In the case of Aristotle\(^8\), we also have experts on whom we, as laymen, rely. They are those persons who still used to know Aristotle personally and to be in close contact with him.\(^9\) If we took the baptism metaphor literally, only those persons who had been present at Aristotle’s baptism, would be the experts. But then metaphor must not be taken. Devitt (1981), section 2.8, rightly emphasizes the importance of the fact that the causal network of uses of a name is ‘multiply grounded’ in the object denoted. It would then rather seem that our "Aristotle"-experts had many encounters with an object which, due to their "Aristotle"-expertise, they took to be Aristotle and which they called "Aristotle", or to which they applied the Greek word into which our word "Aristotle" has developed; these encounters then correspond to the typical cases of applying "water", as determined by accepted methods. For the most part, they have encountered one and the same object; then this object is Aristotle – just like most typical cases of applying "water" are water. The fact that they were sometimes mistaken and took the wrong one to be Aristotle, cannot change this.

There is, of course, no guarantee that they mostly took one and the same person to be Aristotle. It may also have been two or more objects that they were continuously and

---

\(^8\) It should be clear that the considerations to follow are completely independent from Aristotle’s celebrity. It is easily seen that they also apply at least to all names of concrete objects, like persons, mountains, celestial bodies, streets, boats, teddy bears, historical and private events, whether they are famous or only known to a small community. Our aim has already been reached once it is shown that such names are hidden indexicals. I will therefore ignore names of numbers, sets, and other problematic objects.

\(^9\) The fact that the "Aristotle"-experts have all been dead for a long time, is not crucial here. It only poses an additional epistemic problem that we always have with historical objects, but not with objects that still exist. In the case of predicates too, it is imaginable that the experts have all died out; it could be the case that only the predicate has been handed down.

The fact that the "Aristotle"-experts cannot be members of the English language community – simply because it had not even existed at the time – is no problem either. Of course, one linguistic community can lean on the other one and thus completely rely on foreign experts.
systematically mistaking for each other. It is not entirely clear how such a situation should be accounted for. The presupposition underlying our present discussion thus also implied that, most of the time, our "Aristotle"-experts had been dealing with the same object.

As "Aristotle"-laymen, we then take that person to be Aristotle whom our experts have already taken to be Aristotle. At any rate, the English convention says that this name must be granted this tradition. This aspect of the English usage of "Aristotle" is particularly striking in that Aristotle has been dead for so long and has had such historical impact since. But it is really the less important aspect. The crucial aspect we rely upon is our experts’ ability to identify and recognize Aristotle.

The above characterization is fully in accord with Evans’s (1982), ch. 11, account of using names. Evans speaks of "proper-name-using practices". What we, following our discussion of predicates, have called experts and laymen, for him are producers and consumers. He distinguishes three phases: the early phase in which there are practically only producers; the midpoint when there are both the producers and lots of consumers who have acquired the use from either other consumers or, in the end anyway, from the producers; and, finally, the late phase where the name tradition only survives in consumers and is possibly fading away. Evans describes all three phases in detail. But for him too, the early phase is the crucial one; according to him, the use of a name is grounded in the producers’ abilities of recognition and identification.

Finally, just like the usage of "water", the English usage of "Aristotle" includes an essentiality convention. It serves the ontological task of determining what kind of object Aristotle is; for the talk about the origin of the usage of "Aristotle" does not by itself answer this ontological question. For instance, I do not think that a human being is

10 The following quotation from Evans (1982), p. 382, in which he also draws the parallelism with natural kind terms, is particularly instructive:

"It seems reasonable to suggest that what makes it the case that an ordinary proper-name-using practice involving the name ‘NN’ concerns a particular individual is that that individual should be known to the producers in the practice as NN. It is the actual pattern of dealings the producers have had with an individual – identified from time to time by the exercise of their recognitional capacities in regard to that individual – which ties the name to the individual. The information circulating in the practice will normally provide good evidence for which individual it is that has been recognized from time to time as NN by the producers in the practice, for much of it will be a trace of some encounter between a producer and an individual, although countless equally relevant identifications of an individual as NN will leave no permanent trace upon the practice.

"To see the actual pattern of recognition and identification by the producing members of the practice as the fundamental mechanism whereby ordinary proper names are endowed with a reference is to see a parallel between those names and the ordinary words we have for natural kinds of things and stuffs. It is an essential feature of the practices associated with terms like ‘elm’, ‘diamond’, ‘leopard’, and the like that there exist members – producers – who have a de facto capacity to recognize instances of the kind when presented with them. I mean by this an effective capacity to distinguish occasions when they are presented with members of that kind, from occasions when they are presented with members of any other kinds which are represented in any strength in the environment they inhabit. This recognitional capacity is all that is required for there to be a consistent pattern among the objects which are in fact identified as elms, or whatever, by members of the speech community – for all the objects called ‘elms’ fall into a single natural kind – and no more is required for a natural-kind-term practice to concern a particular natural kind." (Italics as in the original)
identical with the spatio-temporal region that he covers during his lifetime, nor with that spatio-temporal region plus its material filling – simply because, for the latter two objects, but not for human beings, spatio-temporal location is essential. But maybe it is not implausible to say that a human being has exactly the same causal effects as the filling of the spatio-temporal region it occupies. If so, then both the human being Aristotle and his spatio-temporal region might be the origin of the usage of "Aristotle". The English essentiality convention for "Aristotle" eliminates this and other indeterminacies by fixing that Aristotle is essentially a certain person, not a certain spatio-temporal filling. This point may be so obvious that one fails to notice that the English language community must decide on one out of a whole range of ontological possibilities. This is why I have mentioned it here, even though I am not going to pursue it.

As in the case of predicates, the crucial point is that Aristotle, the origin of the English usage of "Aristotle", is not himself part of that usage. If he were, the above rule would make the character of "Aristotle" constant, and the name absolutely rigid. His not being part of it, however, shows that "Aristotle" is a hidden indexical. For even our "Aristotle"- experts do not know Aristotle’s essential properties, so that their methods of identification, their abilities of recognition may, even collectively, completely fail. There are thus context worlds in which someone else appears to be Aristotle to them and, consequently, to us who are following them.

Such context worlds are easy to invent. Another sperm than the actual one may, e.g. have fertilized the egg from which Aristotle grew. Then not Aristotle, but another person would have existed instead, and we could assume that that person would have looked and lived like Aristotle. Or the evil midwife could have foisted a different child on Aristotle’s mother at his birth. Or, at a later time, there could have been a miraculous replacement never to be discovered, maybe as an intervention by the gods. In each such context world the same English usage of "Aristotle" would have had a different origin.

There is a complication here, though. The later such a replacement takes place, the more one would be inclined to say that there would not be exactly one usage of "Aristotle" in such context – which leads us back into the well-known and temporarily suspended problem. But if the replacement takes place before Aristotle’s historical deeds, one may say about such a context that the initial usage of "Aristotle" had disappeared and only the second one was handed down until today, so that today we would indeed have only one usage, which is the one we are familiar with; the latter would then not refer to our Aristotle but only to Aristotle’s substitute.\footnote{Cf. Evans (1982), who gives a detailed discussion of such a substitution on pp. 388f. In our case, Evans would say that there is only one usage of "Aristotle" that would have had no reference during the problematic period in which the experts had nearly equally relevant dealings with the two persons they were confusing. But he also holds it possible that in the end this usage only refers to the second person, which would be the case if the first person had left no significant traces of information among the experts.}
According to the above character definition and the explanation as to the English usage of "Aristotle", stories like this imply the covert indexicality of "Aristotle". What do other authors say about this?

Kripke (1972) seems to take the origin itself as part of the name's usage or of the causal chain of communication; this is suggested by various remarks of his as to whether it could have turned out that Hesperus is not Phosphorus. But the matter is not clear, because when he explains that Hesperus and Phosphorus cannot be distinct, he only does so with reference to other index worlds, and never to other context worlds.

Devitt (1981) seems to think alike. But then, he did not ask the question that we are pursuing here either, since it was of no theoretical interest for him; in his attempt at solving the infomation problem for identity statements (in section 5.5), he employs what he calls designating-chains that include their causal beginnings.

However, the discussion in Evans (1982, S. 388ff.), already mentioned in footnote 11 above, makes it clear that Evans would have allowed distinct objects to underlie one and the same usage of a name. On the other hand, there is no indication that he takes theoretical advantage from this result in the way we do.

According to Lerner und Zimmermann (1984) too, one name tradition can have distinct origins, and they explicitly note the relevance of this fact in applying diagonalization to solve the problem of informativity (cf. p. 37); but then they do not go into the identity criteria for name traditions.

Finally, Kaplan is a special case. In section 3.3 we had already stated his quite distinct views of (1977) and (1989), according to which it is a metasemantic question what reference a name has, the semantics of the name stating that it has this reference in all possible contexts. In (1990) he seems to have changed his mind. There, on pp. 118f., he calls it an open and substantive metaphysical question whether it be "possible that a name which in fact names a given individual, might have named a different individual" – albeit without arriving at any conclusions in this question.

However, as we have seen, our view of this matter is clear. And our reasons are the same as those given in section 3.2, relating to predicates: it allows for a distinction between necessary and analytic identity statements and a solution of the problem of

---

12 E.g., when he says, on p. 308: "... in a counterfactual world in which 'Hesperus' and 'Phosphorus' were not used in the way we use them, as names of this planet, but as names of other objects, one could have had qualitatively identical evidence and concluded that 'Hesperus' and 'Phosphorus' named two different objects" (my italics).

13 His definition of designating-chains, given on p. 29, is rather unequivocal: "The chain underlying my first use of 'Nana' begins with Nana at her naming ceremony; it runs through my perception of that ceremony; from then on its my ability thus gained to use 'Nana' to designate her. I shall call such a causal chain a 'd-chain', short for 'designating-chain'." But then in section 5.4 he also discusses more ambivalent cases.

14 On p. 24 they say: "Which individual x stands at the beginning of a tradition T, is a contingent property of T."

15 As already mentioned in VVV footnote 7 on p. 150f., Kaplan takes names to be something like our usages or name traditions.
informativity as to proper names. But these theoretical conclusions will only be drawn in section 4.4.

At first glance, however, the fact that the English usage of "Aristotle" does not contain its origin as a part, seems to obscure its individuation: What, then, is holding it together as the one usage it is? In the case of predicates, where ambiguities are rare and could thus be ignored, this was mainly done by their own phonetic or graphical form. But this option is not open to us, if later we want to take into account that many objects may bear the same name. And it cannot be the persons partaking in the usage either. In the case of predicates (cf. p. VVV 96), we had already made it clear that for centuries the English language community could have consisted of persons that are quite different from any actual ones. Not even the actual "Aristotle"-experts are essential for the usage of "Aristotle". Maybe, e.g., Aristotle’s numerous aunts had been the experts for him, because he did not grow up in his parents’ home and his uncles had all been dead. In a different context world, though, his aunts could all have died instead of his numerous uncles, who would then have been the experts. And now, we have said, not even Aristotle himself, the origin of that usage, is supposed to be essential for it any more. What then, does constitute the one English usage of "Aristotle"?

It is, as we have said, the experts’ methods of identification. More precisely, if the English language community exists in a context world, if an object called "Aristotle" exists there, and if they, via their experts (who do not have to be among them) on most encounters take this object and nobody else to be our actual Aristotle, then, and only then, the English usage of "Aristotle" exists in that context world.16

This sounds cryptic. How are foreign persons in a foreign world supposed to take somebody to be our actual Aristotle? But there is no secret here. I can easily imagine counterfactual circumstances where I confuse somebody with my friend Madeleine; I am then in a different world and mistake somebody, who may only exist there, for someone from the actual world. If I can imagine that, I can also imagine that such a mistake not only happens to me but to many or even most people, and not only on that one occasion, but on many or even most occasions. We thus have arrived at the case described above, where a different object underlies the German usage of "Aristotle".

In a sense, it is still the actual Aristotle who holds the English usage of "Aristotle" together, for, of course, to a large degree it is him who shapes the methods of identifying and the abilities of recognizing him. For the reasons given, however, only these methods may be taken to individuate the usage, not Aristotle himself.

Of course, this individuation of the English usage of "Aristotle" is unremovably vague. A single expert’s recognizing powers are already loose and hard to account for. Forming a collective sum does make this ability more reliable but no more determinate. It is notoriously vague how many encounters count as most, or how many diverging

---

16 It must be borne in mind that this explanation is made under the assumption that there is only one English usage of "Aristotle". It will have to be slightly modified once we are going to drop this assumption in section 4.5.
expert opinions are still tolerable. But all this rarely creates practical problems. And it does not suffice to discredit the theoretical role we want to attribute to such usages.

We will, in the course of our discussion, meet further examples, putting us in a position to get a sharper view of the usage of proper names. But let us first reflect on how these ideas relate to two kinds of names that are generally taken to be problematic.

4.2 Empty Names, Descriptive Names

As to empty names, we may directly follow up on what we have said about empty predicates on pp. VVV124ff., where we had distinguished between contingently, necessarily, and analytically empty predicates. For names, the first option is out; there are no contingently empty names, i.e., names that denote nothing in a given context, but that do denote something in other possible index worlds. This does not just follow from their rigidity; after all, that a name is rigid only means that it has the same denotation in all index worlds in which it has a denotation at all, which is compatible with its having no denotation in some index worlds. It rather follows from the above character definition as a whole; if the definite description in its *definiens* does not apply to any object in the given context, there is nothing then that could be identical with something from a different index world. If, e.g., we arrive at the conclusion that Homer did not exist – which does not mean that there was no single person to have written the *Iliad* and the *Odyssey*, but only that our usage of "Homer" lacks any historical foundation – then Homer necessarily does not exist; he could not have existed then.

All empty names, then, are necessarily empty. But they need not be analytically empty. This follows from their hidden indexicality – and it is in turn an argument for treating them as hidden indexicals. Let us consider the purported planet Vulcan that was supposed to be responsible for the observed irregularities in planet Mercury’s orbit. When Einstein realized, with the aid of his general theory of relativity, that these irregularities are a measurable effect of the spatial curvation generated by the mass of the sun, it became clear that Vulcan does not exist; and no possible object could then have been Vulcan. But our context world could have been Newtonian and the irregularities could have been due to a celestial body that was hard to discover. Then this celestial body would have been the origin of our usage of "Vulcan". This is one of many ways in which Vulcan could have turned out to exist after all. The fact that Vulcan could not have existed, even though its existence is *a posteriori* thus receives a direct explanation in our theory.

The same is true of all names about which we have realized or about which it is yet to be discovered that they actually denote nothing: mythical creatures, whose existence our ancestors believed in, objects that were alleged on account of an indirect but misleading access, maybe objects of collective hallucination, etc., thus Zeus and all the

---

17 Devitt (1981, p. 175) aptly calls them "failed names".
Greek gods, Atlantis, Vulcan, and Nessie, and maybe also King Arthur and his Round Table, whose historical origin has not been completely cleared up, and so on.

Finally, there are also analytically empty names. In analogy with empty predicates, one may first think of artificial examples, e.g., as one may give names to impossible objects, like Meinong's round square or the largest prime number, which would then not denote anything in any context. But since there is normally no need to talk about such impossible objects, we usually do not have names for them either.

There are, however, more natural examples that correspond to the other sort of analytically empty predicates. For, according to our theory, fictional names are empty too. Let us take the name "Sherlock Holmes", for example. If we were to apply the above picture, we would have to know who the "Sherlock Holmes"-experts are and what are their methods of identification. This is a strange question. In a sense, the prime "Sherlock Holmes"-expert is Sir Conan Doyle, of course; and asking him for his methods of identifying Sherlock Holmes would have been misplaced because he was the first to know that Sherlock Holmes does not exist, which is why he could not identify him. But if it is already part of the "Sherlock Holmes"-experts' knowledge that "Sherlock Holmes" does not denote any existing object, then, on our account, there is no possible context in which "Sherlock Holmes" denotes anything. This means that "Sherlock Holmes" is not only necessarily, but even analytically empty.

A fictional object need not have a single creator. Maybe Little Red Riding Hood was not made up by one person. But somewhere, sometime this fairytale came into being and to the people at that time and place, the "Little Red Riding Hood"-experts, it was clear from the outset that they were telling an invented story about a fictional object. And again, if the experts already deny her existence, "Little Red Riding Hood" cannot denote anything in any context.

As in the case of predicates, the fact that fictional names are analytically empty seems to be an unwelcome result. For example, the absurdity that all fictional names are synonymous, again follows. Moreover, intuitively one would have thought that Sir Arthur Conan Doyle somehow described at least one possible world; in contradiction to this we now get that, even seen from possible contexts, he only described the impossible.

In section 4.4 we will be forced to introduce a formal character of names, quite like we have done for predicates. On the basis of it, these consequences can be softened. We will thus get that the formal character of "Sherlock Holmes" is trivially distinct from that of "Little Red Riding Hood". Thus distinct fictional names have distinct subjective meanings, and the possible subjective uncertainty as to whether a story is about an actual object or about a fictional one, can always be accounted for in theoretical terms. Indeed, our account completely coincides with Stalnaker's (1978), as presented on pp. VVV 48f.

None of this, however, will solve the problems of fictional discourse. The intuition that "Sherlock Holmes" and "Little Red Riding Hood" not only subjectively, but also objectively in English, mean different things, will not be captured. The same is true of the intuition that Dr. Jeckyll is necessarily but not analytically identical with Mr. Hyde.
Finally, our apparatus will not cover the subjective belief that (fictional, not possibly actual) Little Red Riding Hood is (fictionally) eaten by the wolf.

Abstractly, the reason for this failure is again obvious. Our theory only talks about the actual world, about what may subjectively or collectively be taken to be the actual world, and about counterfactual variations of that theme; and reference in fictional discourse simply lies beyond this range of possibilities. How this situation could be improved, I am not going to discuss here.

Let us now turn to descriptive names. They have mainly attracted interest because they appear to contradict Kripke’s rigidity thesis and to fall into the scope of allegedly obsolete descriptive approaches. But to me they seem to be somewhat less problematic.

Let us take a standard case: "Jack the Ripper" is considered a descriptive name because it is synonymous with the description "the person who committed certain women murders in London in 1888"; this is precisely how the name had been introduced. However, this synonymy is quite in accord with our theory of names. Experts’ knowledge and identification procedures will normally be rich and based on many encounters. But they can also be very thin and only comprise a single feature characterizing the bearer of the name. Then the method of identification can be captured by a short description. I believe that "Jack the Ripper" is such a case; the above synonymy holds because the descriptions (1) and (2) are also synonymous:

(1) the person who is the origin of the English usage of "Jack the Ripper",
(2) the person who committed certain women murders in London in 1888.

This claim is only doubtful because definite descriptions are always subject to an ambiguity between a referential and an attributive reading. If we explain "Jack the Ripper" by (1), then, according to the above character definition, (1) must be read referentially; if "Jack the Ripper" is explained as a descriptive name by (2), then, one might object, (2) could also be read attributively. But this objection is illegitimate; "Jack the Ripper" does not seem to have an attributive reading. This can be seen from counterfactual statements like the following:

(3) If William II had murdered these seven women, he would have been Jack the Ripper.\textsuperscript{18}

This sentence can only be true on an attributive reading of "Jack the Ripper", which I think it does not have; it is false on any account. (As always, this must not be confused with the fact that William II could have turned out to be Jack the Ripper.) Thus, like other names, "Jack the Ripper" is covered by our theory.

The same is true of "Vulcan", apart from that it is empty – albeit, as we have just seen in an unproblematic way. "The Morning Star" and "The Evening Star" can also be

\textsuperscript{18} Examples of this kind can be found in Evans (1982, p. 60), where they are meant to illustrate the same point as here.
categorized as descriptive names in that sense. There are, however, more problematic cases. The most famous one is presumably due to Kaplan (1978, p. 241):

(4) Newman 1 = the first child to be born in Munich in the twenty-first century

In this case, "Newman 1" does seem to have both a referential and an attributive reading. If, e.g., on the day in question, the first child to be born in Munich, is born at 00.03 and the second one at 00.05, and the second one’s mother then says:

(5) Had I given birth to my child three minutes earlier, my child would have been Newman 1,

I think she can thereby express something true. Granting this, our approach to names does not apply to "Newman 1". But maybe this should rather be taken as indicating that "Newman 1" is not a real name, but just an abbreviation for a definite description indeed, which can be read either referentially or attributively.

4.3 The Descriptive Content of Names and the Meta-Language Objection

So even if I reject any attributive readings of names, our explanations of the above definition of their characters allow for a positive comment on the much-discussed descriptive content of names. The obvious view is that, even though every subject associates some descriptive content with each name that he uses, none of these are socially binding, so that the name as such lacks any descriptive content. Within our theoretical framework, though, the matter appears to be different. Here the diagonal of a name’s character represents the descriptive content that it has in a language. According to the above definition, in the case of an English name $N$, this is just the description "the origin of the English usage of $N"$, or in more detail: "the object such that, in most cases in which the $N$-experts encountered something they took to be $N$, it was that object". That $N$ itself satisfies this description is the English apriori for the name $N$. Of course, this apriori is pretty vacuous – which is to be expected. But this definite description

---

20 The second formulation highlights the parallelism with the English apriori associated with indexical predicates; cf. footnote 25 of the third chapter (p. 120). This apriori is not that which a subject a priori knows about $N$. We will return to these matters in section 4.4, in connection with the problem of internalism.
does satisfy the minimal requirements: due to the hidden indexicality of $N$, it is descriptive, and it is definite.

Both on a subjective and on a collective level, this *apriori* is, of course, enriched with many beliefs about the relevant object. Inasmuch as such beliefs are widespread, they form the stereotype associated with the name, as one could say in analogy with Putnam. But, as in the case of predicates, the stereotype need not be fully, or even half, true of the bearer of the name. This has become clear from Kripke’s (1972) criticism of Searle (1958), who had thought that the bearer of a name must always be that individual that satisfies most of the information associated with the name. But then not all the details of the causal chain of communication associated with a proper name belong to its descriptive content either; this information too, is *a posteriori* and can turn out to be false. It is only the *apriori* just mentioned that cannot be shaken and is thus completely tied to the name.

The kind of descriptive content of names that I have proposed here, i.e., something along the lines of "the object that is called $N$", has, of course, always been around. It is certainly not what one would typically take to be a descriptive content. In particular, however, its meta-linguistic nature seemed to render it inadequate or illicit. In the case of predicates, we have already managed to reject the meta-language objection in section 3.4; however, it is useful to rehearse the rejection as applied to proper names.

The intuition underlying the meta-language objection was that semantic rules in which the word to be interpreted occurs, amount to misconstruing statements about objects as statements about words (cf. p. VVV 106). The above semantic rule for names seems to make just this mistake. For, according to it, a sentence like

(6) Aristotle was a philosopher

is synonymous with

(7) Whoever is at the origin of the English usage of "Aristotle" was a philosopher.

It thus seems to equate an object-linguistic statement about Aristotle with a meta-linguistic statement about "Aristotle". However, just as in the case of predicates (cf. p. VVV 109), this impression is mistaken. Our definition of the character of proper names only implies that on one reading sentence (6) is synonymous with – has the same character as – (7), in which case the definite description occurring in (7) is understood referentially. But on such a reading an utterance of (7) is not about the English usage of "Aristotle"; it is not the English usage of "Aristotle" that enters the proposition expressed by (7) in a context $c$, but only the individual that forms the origin of the English usage of "Aristotle" in $c$ – because, on its referential reading, (7) – like (6) – is also true in index worlds in which the English usage of "Aristotle", and maybe even the whole English language, do not exist.

---

On pp. VVV 108f. we had discussed a different formulation of the meta-language objection, viz. that it be absurd to claim a sentence like (8) analytic:

(8) Aristotele ist that person who stands at the origin of the English usage of "Aristotle".

This seems to say that Aristotle is necessarily called "Aristotle" in English – which would be absurd, because Aristotle could obviously have been called differently in English. This variant of the objection can be countered in the same way. For, following our semantic rule, (8) is only analytically true if the description it contains is read referentially. On that reading, (8) is true at c and i if, in i, that person who, in c, is at the origin of the English usage of "Aristotle" is identical with that person who, in c, is at the origin of the English usage of "Aristotle" – which only means that Aristotle is identical with himself, and not that he is called "Aristotle" in every index world. On the other hand, the attributive reading of (8) captures the intuition that Aristotle could have been called differently; for on that reading we obtain that (8) is a priori true, but does not express a necessary proposition in any context, and is thus not analytic.

Yet another variant of the meta-linguistic objection was the translation argument (cf. p. VVV 107). We give an abridged version: Intuitively, (10) is the correct German translation of (9):

(9) Munich is pretty,
(10) München ist hübsch.

However, our theory analyzes (9) and (10) in the following fashion (re-translating "ist hübsch" into "is pretty"): 

(11) The object standing at the origin of the English usage of "Munich" is pretty,
(12) The object standing at the origin of the German usage of "München" is pretty.

Obviously, (11) and (12) do not appear to be synonymous; and if this is so, then, against all intuition, (9) and (10) would not be synonymous.

In our treatment of the translation argument in section 3.4 we had already pointed out that we always take synonymy in the sense of sameness of character and that therefore a translation already counts as correct if it is character-preserving (cf. pp. VVV 109f.). So the translation argument will have been defeated as soon as we are able to show that the definite descriptions in (11) and (12), on their referential readings, have the same characters, or – which amounts to the same thing – that our semantic rule

---

22 We ignore the complication arising if the object in question does not exist in w_i.

23 It should be clear that this origin is not just the 12th century market town, the historical origin of Munich. 13th and 20th century Munich equally underlie the present usage of "Munich".
assigns the same character to the English word "Munich" and to the German word "München".

Here again, the argument given on pp. VVV 110ff. immediately carries over. The present stages of English and German – which is all we are concerned with here – are so intertwined that no context worlds are possible where different objects stand at the origins of the English usage of "Munich" and at the German usage of "München". For, given established translation practices and the cultural exchange between the two language communities, the "Munich"-experts are the same as the "München"-experts, viz. at least all Germans, Englishmen, and other persons who have spent some, longer or shorter, time in Munich; to be sure, in an analogous way the paradigm cases of "Wasser" had become paradigm cases of "water", and vice versa. So the English usage of "Munich" and the German usage of "München" are identical these days, which is why English "Munich" and German "München" have the same intension in every context.

This identity claim appears implausible. But again, our modified concept of usage of $N$ must be noted. The German usage of "München" and the English usage of "Munich" have of course developed in distinct concrete causal networks, simply because "München" and "Munich" are distinct words and thus have disjoint sets of occurrences. However, as explained above, neither the phonetic or orthographic form of a name $N$ nor the specific causal network of its uses is essential for the identity conditions of the usage of $N$. The case of "Aristotle" had made it plain that we have the same usage of "Aristotle" today as the ancient Greeks, although we spell and presumably pronounce the name differently, and although our uses of the name occur at a much later time. We have simply adopted this usage from the ancient Greeks by historical mediation; and our usage is identical to theirs, because we rely on the same experts and their recognitional capacities. Now, the identity of a usage of $N$ cannot only be traced through a successions of language communities, where the later always rely on the earlier, but also through ramifications – which is why the German usage of "Aristoteles" is the same as the English usage of "Aristotle" – as well as more complicated causal interrelations – like, presumably, in the case of "Munich" and "München". In each case, it is the $N$-experts' knowledge that takes care of the identity of the usage of $N$; and it is inessential for the usage of $N$ – staying in the picture –which particular lay talk may be associated with it.

However, this does not mean that for any object there is only one usage. Cases diverging from that of "Munich"/"München" are conceivable too: Two language communities without any contact with one another, live on a tall mountain, one on the east side, the other in the west. In the first language the mountain is called "Tsaetnoum", in the second one it is "Tsewtnoum". On our account, "Tsaetnoum" and "Tsewtnoum" do not have the same character. For both languages could be spoken in worlds in which the origin of the usage of "Tsaetnoum" and the origin of the usage of

---

24 Kaplan (1990, footnote 14) also entertains the possibility that "München" be the same word as "Munich" – in his sense of "word".
"Tsewtnoum" are two distinct mountains, as well as in worlds in which the other language community does not even exist.

Now, what happens when an explorer visits both language communities to discover that by "Tsaetnoum" and "Tsewtnoum" they refer to the same mountain? He would surely translate them as "Tsaetnoum" and "Tsewtnoum". According to our theory, he would thereby not translate two expressions with identical characters but rather start the contact between the two language communities – with the possible result that the "Tsaetnoum"-experts become "Tsewtnoum"-experts, and vice versa, so that the two names would become identical in character.25

A remaining uneasiness about the above semantic rule may be due to the suspicion of circularity pronounced by Kripke (1972), p. 283f., who says that a semantic theory would be circular if it were to determine the reference of names by the rule that a name $N$ refers to the one who is called $N$; the answer that such a theory gave to the question "to whom does $N$ refer?" would simply be: "whoever $N$ refers to". He rightly remarks:

"Obviously if the only descriptive senses of names we can think of are of the form ‘the man called such and such’, ‘the man called ‘Walter Scott’’, ‘the man called ‘Socrates’’, then whatever this relation of calling is is really what determines the reference and not any description like ‘the man called ‘Socrates’’.” (p. 284)

He himself escapes this suspicion by giving a closer characterization of this relation of calling in terms of his causal theory of reference. We may also restore to this defense. With our slightly different account in terms of usages of names, we too have given enough substance to the relation of calling to scotch any suspicion of circularity.

But isn’t there another way in which our semantic rule is circular? On p. 160 we had given a slightly more extensive formulation, according to which at $c$ and $i$ "Aristotle" refers to the object such that in most cases in which the $N$-experts encountered something in $w_c$ they took be $N$, it was that object. This sounds circular in that the the definition of the reference of "Aristotle" already refers to Aristotle himself, the referent of the name in the actual context world. But this is a common feature of all semantic rules: In describing the meaning of the object-language in the meta-language, one uses meta-linguistic expressions with the same meanings as the object-linguistic ones. Then if meta-language and object-language are the same – in our case English – appearances of circularity finally arise. But this is unavoidable and usually taken to be harmless. So none of the objections against the meta-linguistic nature of our semantic rule for proper names proves right.

---

25 But then it could also happen that the respective cognitive accesses to the mountain remain associated with the names. Tsaetnoum would thus be this mountain as accessed from the east, and Tsewtnoum would be the mountain as accessed from the west. The two names would then retain their distinct characters, and the whole story would resemble the Hesporus/Phosphorus case which we will discuss immediately.
4.4 Proper Names and the Problems of Informativity and Internalism

The crucial justification for our semantic rule and the resulting hidden indexicality of proper names are due to the fact that they allow for a solution of the problem of informativity. What this solution looks like should already be more or less clear; we will still illustrate it with a few examples. Let us first consider sentence (13) and its diagonal (13a):

(13) Hesperus is Phosphorus,
(13a) \{c \in w_c, \text{ that object which in } w_c \text{ stands at the origin of the English usage of } "Hesperus" \text{ is identical with that object which in } w_c \text{ stands at the origin of the English usage of } "Phosphorus"\}.

The diagonal of (13) is informative if there are context worlds \(w_c\) where there is a different object at the origin of the English usage of "Hesperus" than is at the origin of the English usage of "Phosphorus", which can only be the case if the English usages of "Hesperus" and "Phosphorus" are distinct. We have seen that the fact that two distinct names enter the two usages is not by itself a reason that there be two distinct usages; after all, we have argued, the English usage of "Munich" and the German usage of "München" are the same usages. In the case of "Hesperus" and "Phosphorus", though, we are dealing with two distinct usages. This is so because, in both usages, the access to the object in the origin is determined rather precisely: Exactly those persons count as "Hesperus"-experts who know Hesperus, or its possible epistemic counterparts in other context worlds, from observing the evening sky, whereas "Phosphorus"-experts are exactly those persons who know Phosphorus, or its possible epistemic counterparts in other context worlds, from observing the morning sky. The overall assumption is that there is exactly one object underlying the usage of "Hesperus", i.e., that there not be different celestial bodies to be first seen on different evenings; and likewise for the usage of "Phosphorus". But it is not taken for granted that one object underlies both usages. There are therefore conceivable context worlds where, unlike in the actual world, the origin of one usage is distinct from the origin of the other usage.\(^{26}\) This is precisely why the above diagonal (13a) does not cover all contexts and is therefore informative.

This remains so until today, even though sentence (13) has been generally accepted for a long time. The reason is that the two accesses are still strongly associated with the respective names. So if, surprisingly, it turned out tomorrow that our astronomers had made an enormous mistake and different celestial bodies are to be seen in the morning

\(^{26}\) Kripke (1972), pp. 306f. and pp. 310f., describes exactly such situations in order to substantiate the claim that (13) is known \textit{a posteriori}. 
and in the evening, we would not simply say that two names, impossible to disentangle, now stand for two objects rather than one; we would rather draw the more specific conclusion that, after all, Hesperus is different from Phosphorus.  

By the same token, the diagonal of (14) is, on the other hand, not informative in English:

(14) Hesperus is the Evening Star.

For the English usages of "Hesperus" and "the Evening Star" are not distinct.

Let us now look at an equally famous though somewhat different example, viz. sentence (15) and its diagonal (15a):

(15) Cicero is Tullius,
(15a) \{c \in w_c, \text{that object which in } w_c \text{ stands at the origin of the English usage of "Cicero" is identical with that object which in } w_c \text{ stands at the origin of the English usage of "Tullius"}\}.

Is the English usage of "Cicero" the same as the English usage of "Tullius"? We had said that a usage of \(N\) is individuated via the identification methods associated with the name. So if the two usages were distinct and could thus possibly have different origins, the English usage of "Cicero" would have to include different methods of identification than the English usage of "Tullius". But this is not plausible in this case. To be sure, most people will associate the properties for which their denotation became famous with the name "Cicero", not with "Tullius"; but this is no reason for different methods of identification. Rather, it must be assumed that the experts in "Cicero" or "Tullius" had no different accesses to Marcus Tullius Cicero but used both names interchangeably right from the beginning. Things would be different if, e.g., the man had led a double life and had become famous in one of his roles as "Cicero", and as "Tullius" in the other one. Then we would have had the same situation as with Hesperus and Phosphorus. But this not having been the case, we must admit that, according to our theory, sentence (15), like sentence (14), is analytic in English and thus not informative.

At this point it becomes once more apparent that the concept of character used so far does not do justice to all intuitions surrounding the notion of subjective meaning. For even if it is not an unwelcome result that (14) and (15) are analytic in English, one would still say that both sentences can well be informative for individual persons.

27 See also our discussion on pp. VVV 98f. as to why, even today, " water = H\textsubscript{2}O" is not analytic, although the identity is well-known today.

28 It must be noted that so far we are only concerned with informativity in English. As to individual subjects and their a priori knowledge, there is, of course, no difference between (13) and (14). This will soon be accounted for.

29 Searle (1958) had conceded this consequence, too, though for different reasons.
However – as in the case of predicates – this concept of subjective informativity is not captured by the objective character of names defined so far.

The current approach does not solve the problem of internalism either; for it delivers no consistent belief content for a person like Kripke’s Pierre. Let us remember: Pierre utters both sentence (16) and sentence (17) seriously, honestly and with sufficient competence of both English and French – but without knowing that "London" is the English translation of "Londres"; given the present state of our theory, his belief set would then have to be a subset of both (16a), the diagonal of (16), and (17a), the diagonal of (17):

\[
(16) \quad \text{Londres est jolie}, \\
(16a) \quad \{c \mid \text{that object which in } w_c \text{ stands at the origin of the French usage of "Londres" is pretty in } w_c \}\), \\
(17) \quad \text{London is not pretty}, \\
(17a) \quad \{c \mid \text{that object which in } w_c \text{ stands at the origin of the English usage of "London" is not pretty in } w_c \}\).
\]

But what was true of the English usage of "Munich" and the German usage of "München" also holds for the English usage of "London" and the French usage of "Londres": due to the exchange between the two language communities, they are one and the same usage. So "London" and "Londres" have the same objective character, and consequently the diagonals of (16) and (17) are not consistent with each other; the intersection of (16a) and (17a) is empty.

In the case of proper names, then, the problem of internalism as well the subjective informativity of sentences like (14) and (15) give rise to the conclusion that we have already drawn in the analogous situation involving predicates: the objective character of a name is not suitable for modelling subjective meaning; it cannot capture the fact that competent speakers do not have to perfectly master their language – or in this case the usages of names of their language – or, at any rate, not to a degree of perfection that prevents the above problems from arising. Apart from the concept of an objective character, then, for names too do we need a concept of formal character that abstracts

30 In the following we are not only ignoring tense but also the detailed analysis of the predicate.

31 Lerner and Zimmermann (1984, p. 37f.) also propose to solve the problem of informativity by diagonalisation and a hidden indexicality of names. They, too, arrive at the conclusion that what we have called the problem of internalism cannot be solved by that strategy. But then they take this to be a reason for giving up the method of diagonalization altogether and to resort to a theory involving structured propositions in the sense of Cresswell and von Stechow (1982). Lerner and Zimmermann are not treating Kripke’s "London" example but a case that is analogous to Kripke’s "Paderewski"-example. I will only come to the Paderewski-case in section 4.7, because it presupposes the possibility of having one name denote different objects. This again necessitates a more detailed study of the semantics of noun phrases in general, which is why, even in section 4.7, I will not be able to demonstrate the solution to the Paderewski-case; but it will become completely clear that such a solution is available within the diagonalization framework.
away from the language to which the name belongs. The rule for the formal character of a name $N$ then reads:

$$
\ll N \ll_\text{S} (c) (i) = \text{that object which in } w_c \text{ stands at the origin of the usage of } N \text{ in } l_c \text{ – where } l_c \text{ is } s_c \text{'s language in } w_c \text{ at } t_c, \text{ and of syntactic structure } S.
$$

Given this, the diagonal of (15) reads:

(15b) \{c \mid \text{that object which in } w_c \text{ stands at the origin of the usage of } "\text{Cicero}" \text{ in } l_c \text{ is identical with that object which in } w_c \text{ stands at the origin of the usage of } "\text{Tullius}" \text{ in } l_c \}.\]

(15b) is almost trivially informative; after all, context languages are possible in which the words "Cicero" and "Tullius" are used in completely different ways than they are used in English. On the formal character level, no characteristics of the usages of "Cicero" and "Tullius" have yet been specified; the diagonal of (15) could even contain contexts in whose language "Cicero" referred to a rooster and "Tullius" to a month. However, the range of possible context languages crucially narrows down once we consider the diagonal of the formal character on the background of the speaker’s belief set, because then all his assumptions about the usages of $N$ in his languages must be taken into account. But then it will scarcely happen that, for any name $N$, a speaker knows the usage of $N$ in full detail; this becomes even more improbable, the more extensive and confusing the usage of $N$ is, or the more spatially and temporally remote its origin. Even if the speaker himself is one of the $N$-experts, he need not fully know the usage, i.e., the methods for identifying $N$; for these consist of the collective, not the individual, expert knowledge. Thus, even though Pierre is a "Londres"-expert, which is the same as a "London"-expert, his methods of identifying London are so limited that he would buy any city as the origin of the usage of "Londres" if only one of its quarters looks right to him. A speaker can only have perfect knowledge of the usage of $N$ if its object of origin has not been observed more closely by anybody but the speaker himself – which will presumably only happen with small, short-lived objects that therefore do not bear public names. In that case, one could say that the speaker knows the essential properties of the usage of $N$ – which does, however, not mean that he also knows the essential properties of its origin.

The formal character thus helps solving the problem of internalism for Kripke’s Pierre because Pierre makes such different assumptions as to the French usage of "Londres” and the English usage of "London” that he does not even know that they have the same origin. The formal account of this result, however, adds a little complication concerning the formulation of the diagonal; talking of the language of $s_c$ obviously does not make sense in the case of multi-lingual speakers like Pierre. The problem is solved if we refer, not to the one context-language $l_c$, but to two context-languages $l_1^c$ and $l_2^c$; the diagonals of (16) and (17) then read:
4. Proper Names

(16b) \{c \mid \text{that object which in } w_c \text{ stands at the origin of the usage of } "\text{Londres}" \text{ in } \mathcal{L}_c \text{ is pretty in } w_c \},
(17b) \{c \mid \text{that object which in } w_c \text{ stands at the origin of the usage of } "\text{London}" \text{ in } \mathcal{L}_c \text{ is not pretty in } w_c \}.

This finally allows for the following account of Kripke’s puzzle: In section 3.9, we have already given a definition of belief \textit{de dicto}, according to which a subject believes \textit{de dicto} that φ if there is a translation φ' of φ into the, or an, actual language of the subject such that the diagonal of the formal character of φ' is a superset of the subject’s belief set. Applying this definition, we conclude that Pierre believes \textit{de dicto} both that London is pretty and that London is not pretty, without thereby ascribing him an internally inconsistent belief – because (16b) and (17b) are consistent with each other. This is exactly how Kripke (1972) has always wanted the case to be accounted for.

Even so, with the formal character of proper names as it has been explained so far, we still have not arrived at a comprehensive and adequate account of their subjective meaning. To that end, we must finally give up the unrealistic assumption that every name can designate at most one object. This will lead to unexpectedly complicated consequences concerning the objective and formal characters of names; they are also novel in that there was no need to discuss them with reference to predicates.

4.5 Names as Ambiguous Expressions

At the beginning of the chapter I had already mentioned that there are two ways of interpreting facts such as the one that many people are called "Hans", or many places "Kirchberg". The ambiguity approach says that we are dealing with many different homonymous words here. The indexical approach, on the other hand, claims that we are always dealing with the same word "Hans", or "Kirchberg", which, however, depending on the context, rigidly refers to different objects. Of course, on this view, the contextual feature with which the reference of names varies, is not just the context world but a more local, speaker- and hearer-related property of the context, which we will go into later; in any case, it will have to do with the speaker’s intentions and causal connections or with the hearer’s expectations. This kind of context dependence of names would thus be a form of overt indexicality, which would then have to be added to the hidden indexicality discussed above.

In the following, we will first deliberate the arguments for and against the ambiguity approach, and in section 4.6 we will address the pros and cons of the indexical theory. It will turn out that the ambiguity theory is more suitable for describing the objective meaning of names, whereas the indexical theory better accounts for their subjective meaning. Thus far, these two results can be reconciled. But then the indexical approach leads to extensions that will be discussed in section 4.7 and that will finally go beyond the present work.
A prominent supporter of the ambiguity theory is Kaplan (1977, section XXII, and 1990). Kaplan’s intuition is this: If a name, or more precisely: the phonetic or orthographic form of a name, is uttered, then, in a sense, it obviously depends on the context, on the circumstances of the utterance, what this utterance refers to. But this should be taken to mean that the circumstances determine which word, which of many different homonymous names, had been uttered, and not that they determine the content of a single context-dependent word. The context-dependence in question is not something that pertains to the character, the linguistic meaning of the expression uttered, but a dependence that must be clear before semantic description can even begin. Why, nobody would suggest that the word "bay" be indexical because, depending on the context, one may use it to talk about a plant or a landscape, or that a noun phrase like "red apples and pears" would be indexical because, depending on the context, it must be analyzed as [(red apples) and pears] or as [red [apples and pears]] and interpreted correspondingly. Or, in Kaplan’s words:

“... semantics cannot tell us what expression was uttered or what language it was uttered in. This is a presemantic task.” (1977, p. 559)

"The contextual feature which consists of the causal history of a particular proper name expression in the agent's idiolect seems more naturally to be regarded as determining what word was used than as fixing the content of a single context-sensitive word. Although it is true that two utterances of "Aristotle" in different contexts may have different contents, I am inclined to attribute this difference to the fact that distinct homonymous words were uttered rather than a context sensitivity in the character of a single word "Aristotle". Unlike indexicals like "I" proper names really are ambiguous. The causal theory of reference tells us, in terms of contextual features (including the speaker's intentions) which word is being used in a given utterance. ... Some may claim that they simply use 'indexical' in a wider sense than I (perhaps to mean something like 'contextual'). But we must be wary of an overbroad usage. Is every ambiguous expression an indexical because we look to utterer's intentions to disambiguate? Indeed, is every expression an indexical because it might have been a groan?” (1977, p. 562)

Kaplan’s intuition that proper names are not indexical but ambiguous can be backed up by the following line of reasoning: If one wants to translate a sentence containing an ambiguous expression – for example, "My neighbour is sitting on the bank" – into a language that contains no expression with the same ambiguity, one must first determine in what sense the expression has been used in that connection. In some sense, the information needed here is certainly information about the context of utterance of the expression, but one would not want to count it among its semantics and include it in its character. But then a similar situation as with "bank" may arise with a proper name. Thus, e.g., the German sentence "Wilhelm war ein außergewöhnlicher Mensch" could be translated as "William was an extraordinary person" if it is about William the Conquerer, but not if it is about Wilhelm Busch; then only "Wilhelm was an extraordinary person" would be a correct translation. Before translating, one would thus have to know the relevant usage, or the relevant word. So far, this may not appear a very strong argument, because it can only be brought up in the case of the few proper names that are really translated, and not just adopted. However, in the next section we will derive a stronger argument in favour of the ambiguity approach by considering
translations; it will ultimately show that only with the ambiguity theory can we hold up the principle that translations must preserve the characters of expressions.

So far we have argued for the ambiguity of proper names by drawing on an analogy between names and other expressions that one would clearly call ambiguous. But, of course, there are also obvious differences between an ambiguous predicate like "bank" and an expression like "John". For, unlike expressions of other syntactic categories, proper names are not just accidentally but, in a sense, essentially ambiguous; that is, for an expression to merely belong to the category of proper names already implies that it can name objects of arbitrary number and (almost) arbitrary kind (where the arbitrariness is only restricted by possible descriptive components). So, in that respect, proper names are closer to demonstratives; and if one regarded names as indexical expressions like pronouns, their multiple usability would already be brought out on the character level. The ambiguity approach, on the other hand, treats the multiple usability of names like that of other ambiguous expressions, as something incidental: only those homonymous words of a phonetic form \( N \) then become arguments to the character function which happen to exist in the language and at the time considered. The fact that the very same name form \( N \) can always be used anew and in a different sense is thus not captured on the semantic level.

However, this need not be a disadvantage of the ambiguity theory. For one could also hold that the phenomenon observed is an example for the possibility of the creative use of language: name forms are word forms that can be used to generate ever new words, which is not true, or not true to the same degree anyway, in the case of other word forms. One may, to be sure, also use a predicate in a new sense, thereby creating a new word; but, unlike in the case of names, this is certainly not in the power of single individuals. But if the multiple usability of names is a form of creativity and not of indexicality, then it is quite adequate to not treat this point on the character level but within another component of linguistic theory.

If one now wants to give priority to the intuitions in favour of the ambiguity approach, the next thing to do is clarify the identity conditions of distinct homonymous names. Restricting ourselves for the moment to one world, the actual context world, it is clear that different senses of a phonetic form \( N \) may come into being because the uses of \( N \) occurring in this world belong to different causal-intentional networks; each such network constitutes a different name of form \( N \). But then, of course, we do not just want to consider actual, but also possible utterances, and we must therefore not only mark off names from each other within one context world but also identify them across different context worlds. This in turn makes it necessary to pass over from concrete nets to the usages they determine: distinct homonymous names are thus individuated by distinct usages. One may therefore simply disambiguate names by indexing them with the various usages they are involved in (cf. Lerner und Zimmermann, 1984, p. 24, for the same suggestion). We will use \( U, U' \) etc. to refer to usages; the following rule then accounts for the objective character of a proper name \( N_U \) in English:

\[
\lbrack N_U \rbrack_E (c) (i) = \text{that object which in } w_c \text{ stands at the origin of the English usage } U \text{ of } N.
\]
This rule only differs from the character definition given in section 4.1 by its disambiguation index. Whatever we said there thus carries over without further ado. In particular, the meta-language objection is not salient on any of its versions; and we still have it that proper names are hidden indexicals, which at least partly solves the problem of informativity. The crucial question is thus going to be whether the present strategy for solving the problem of internalism is still applicable. But before going into this, three points must be added arising in connection with the above rule.

First, as always, the functional domain of the above character must be given. \( \| N_U \|_E \) is only defined for contexts \( c \) in which the English usage \( U \) of \( N \) exists, for the reasons given in section 3.3 concerning to the partiality of the character of "water". On the other hand, \( \| N_U \|_E (c) \) must not be taken to imply that \( N_U \) is actually uttered in \( c \). Here again, we may simply adopt the interpretation proposed in section 3.3; there we had said that \( \| \alpha \|_E (c) \) denotes the content that the expression \( \alpha \) would have, if it were uttered in \( c \), so that strictly speaking we are not considering \( c \) itself but a context \( c' \), which is like \( c \), with the possible difference that \( \alpha \) is actually uttered in it – and with all resulting differences like, e.g., that in \( w_c \) at \( t_c \), \( s_c \) masters the language to which \( \alpha \) belongs to a degree that licences one to say that he made an utterance in this language. It must only be added that one of the resulting differences is that in \( w_{c'} \) at \( t_c \), \( s_c \) must be suitably connected with the usage \( U \) of \( N \).

Second, the principle "one usage, one name" that we have just used to individuate names, immediately leads to the question of how these individual usages themselves are to be individuated; this question presents itself under changed conditions now that we have given up the unrealistic assumption of our above discussion and allow an English name form \( N \) to be used in different senses and thus in different English usages of \( N \). We have seen that, along with the fact that we are dealing with an English usage in the first place, so that the name has some currency in the English language community, the \( N \)-experts' methods of identification suffice to individuate the unique English usage of \( N \). But now different English usages of \( N \) must be distinguished from each other. The following example reveals that the methods of identification alone do not suffice.

Let us imagine that two groups of hikers discovered a certain particularly beautiful resting place and that, by pure coincidence, both gave it the same name, "Hiker’s Haven". They thereby each start a usage of "Hiker’s Haven". Let us assume that the two groups never get in touch with each other and also only talk about the spot among themselves or with disjoint acquaintances, so that we are intuitively dealing with two different names and two different usages \( U \) and \( U' \). But then nothing prevents us from shaping the example in a way that the respective experts, that is, the members of the two hiking groups, use exactly the same criteria of identification, i.e., that the place was presented to them in the same way. In this case our present individuation of usages fails.

But then the way out of this difficulty is obvious enough: The two usages \( U \) and \( U' \) must be distinguished by the experts themselves. More precisely, a causal-intentional network of utterances of the form "Hiker’s Haven" only instantiates the usage of "Hiker’s Haven\(_c\)" in a possible context world if it starts out from the same experts with
the same methods of identification as does the usage of "Hiker’s Haven_1" in the actual world; and similarly for the usage of "Hiker’s Haven_2". What counts as the same methods of identification is, as we have already said, pretty vague. Similarly, the "same experts" need not be numerically the same experts; rather, we are dealing with some vague group identity that admits for some counterfactual variation in group membership.

Third, the question arises, as to when does an utterance of the name form \( N \) belong to a particular usage \( U \) of \( N \), i.e., when is it an utterance of the name \( N_U \)? Under our previous assumption that there only be one usage of \( N \), we could take this question easy, because then any utterance of \( N \) could principally be counted among this usage of \( N \). But now the question can no longer be dodged. It is, indeed, this question that Kripke answered tentatively and others more extensively (cf. footnote VVV 3, p. 149) by giving more detailed accounts of the expansion procedure of the causal network in which a usage of \( N \) is realized. I am still not going to discuss these accounts here. Let it only be said that, in any case, the speaker would have to link up to one of the existing usages of \( N \) (or start a new one), if one may at all talk of an utterance of the word \( N_U \) and assign a logical form to it. The speaker may fail to link up to the causal-intentional network of a usage of \( U \) and thus to utter a name altogether if, e.g., he takes two usages of \( N \) to be one; we are still going to present examples for that case.\(^{32}\) Or we may look at the above-mentioned possibility that the entire language community is mistaken in taking certain usages of a name \( N \) to be based on one and the same individual, whereas the truth is that there are two individuals that are permanently confused; here again the two usages of \( N \) do not form a single causal-intentional network; they do not belong to any usage of \( N \) and are thus not utterances of any name whatsoever.\(^{33}\)

It obviously follows that utterances containing such expressions do not express propositions and have no truth-values. This consequence does not appear unwelcome to me. But it already gives an idea of the problems arising for an account of subjective meaning as based on the ambiguity theory. For it seems implausible that such utterances have no meaning even to the persons who make or hear them. Let us thus turn to the true test case for an account of subjective meaning, the problem of internalism.

For this we will again consider Kripke’s Pierre and his utterances about London; given what we have said in the preceding section, it is clear that they must have a logical form of the following kind:

\[^{32}\text{This case is also discussed by Kaplan (1990, section 1.11) who arrives at the same result. Other, partly quite intricate examples showing that it is doubtful whether a speaker links up to an existing usage or not, can be found in Evans (1973).}\]

\[^{33}\text{As mentioned in VVV footnote 11 on p. 154, Evans would, in such a case, want to say that there is a certain name usage, but one without reference. But if, in the individual case, it is conceded that a speaker who is subject to such a confusion, fails to hook onto a name usage, one ought to say that, in the collective case, too, that a confused linguistic community fails to establish a name use. To that extent, it is true that a usage of } N \text{ has at most one object at its origin in a context-world.}\]
Pierre thus uses two expressions, "London" and "Londres", that belong to the same usage and are thus disambiguated by the same subscript, even though by their form, they are distinct words. But the two diagonals

\[(18a) \{c \mid \text{that object which in } w_c \text{ stands at the origin of the French usage } U \text{ of } "Londres" \text{ is pretty in } w_c \},\]

\[(19a) \{c \mid \text{that object which in } w_c \text{ stands at the origin of the English usage } U \text{ of } "London" \text{ is not pretty in } w_c \}.

of (18) and (19) are obviously not compatible with each other. Since the reference of "London" and "Londres" is crucially determined by the same usage \(U\), the same object is always considered, and as no object can both have and lack a property, there is no \(c\) that is both in (18a) and in (19a). (18a) and (19a) thus do not adequately account for the meaning that (18) and (19) have for Pierre.

We had already arrived at this point in our analysis of the London-case in the previous section. However, there we had seen that the problem can be solved if, as in the case of predicates, we define a notion of formal character for proper names by abstracting over the language under consideration. If we were to apply this procedure again, we would now obtain the following definition:

\[\ll N_U \ll^S (c) (i) = \text{that object which in } w_c \text{ stands at the origin of the usage } U \text{ of } N \text{ in } l_c \text{ – where } l_c \text{ is } s_c \text{'s language in } w_c \text{ at } t_c, \text{ and of syntactic structure } S.\]

It is, however, obvious that in the case at hand such a concept of formal character is of little use to us. For by referring to the usage \(U\) this function too becomes ‘rigid’ at the relevant place: even in contexts in which the context-language \(l_c\) is not English, \(N_U\) can only be interpreted via the very usage \(U\) of \(N\). Other usages of \(N\) do not come to bear here, because we are only looking at a particular use of \(N\), viz. \(N_U\).

Hence in sentences (18) and (19) the formal character of the respective occurrences of "London" and "Londres" would be the same and their diagonals would thus still be contradictory.

This does, of course, not show that it is principally impossible to develop an adequate concept of subjective meaning and solve the problem of internalism on the basis of the ambiguity theory. But then it is not obvious how this could be done within the framework of the approach I am following in the present work. For its basic idea had been to explain the gap between subjective and objective meaning by reference to individual subjects’ errors or uncertainties about contextual features that are relevant for determining the intensions of expressions. If, however, proper names are ambiguous and must be disambiguated in the way indicated, then the problematic cases do not involve error or uncertainty about properties of the context, but error or uncertainty
about the words that have been uttered, i.e., about form, not content. But modelling this would presumably amount to assigning different linguistic forms as arguments to objective and formal character functions. Later, when we have seen the ambiguity theory’s alternative, namely the indexical theory, I will be in a position to bring out this point more clearly.

4.6 Names as Overt Indexicals: First Variant

The indexical approach can be formulated in two ways. Let us first concentrate on the more obvious variant. It immediately emerges from our discussion of the problems that the ambiguity approach runs into in an account of subjective meaning; these problems are due to competent speakers’ and hearers’ error or ignorance as to which usage a given occurrence of a name form \( N \) links up to. Within our general framework, this suggests that the states of affairs determining which usage of \( N \) is actually at stake, should – against Kaplan – be taken as semantically rather than pre-semantically relevant properties.\(^{34}\) There would thus only be one word \( N \) which, depending on the context, could be used in different ways, as belonging to different usages of \( N \); the corresponding rule for the objective character of proper names then reads:

\[
\ll N \ll_E (c) (i) = \text{that object which in } w_c \text{ stands at the origin of the English usage of } N \text{ that is salient in } c.
\]

Here the adjective ”salient” marks those contextual features that help pick out whichever usage is at stake in context \( c \) out of a host of English usages of \( N \). For the time being, we can leave completely open which features these may be – the speaker’s intentions, the hearer’s actual or conventional expectations, or something else; we are going to return to this question in section 4.7.

It first seems as if this indexical analysis would already enable the objective character to give a solution to the problem of internalism and, more generally, an adequate account of subjective meaning. For instance – unlike the ambiguity theory – this approach has no problems with supplying the logical form of an utterance the

\(^{34}\) This is also how Lerner and Zimmermann (1984, section 4.1) formulate the indexical approach, and Recanati (1993) too seems to have something like this in mind when he says, on p. 120: ”The reference of a particular token of ‘Gareth Evans’ is the person who is related to the name-type ‘Gareth Evans’ by a name-convention operative in the context of utterance of this token.”

Almog (1981), like myself in the present work, also wants to show that indexicality extends to natural kind terms and proper names, and thus much further than usually assumed. Due to their multiple usability, he takes proper names as indexical in the present sense (cf. 367f.). In Almog (1984, p. 10) he has, however, been convinced by Kaplan that this multiple usability should better be described as ambiguity. Hidden indexicality of names, as analyzed here, makes no appearance in his work.
speaker of which uses a name without suitably linking up to an existing usage. Let us illustrate this by the following example inspired by Kaplan (1990, p. 108, note 13):

Kaplan’s mother has a G.P. named Dr. Shapiro while at the same time being treated by a specialist who also happens to be called Dr. Shapiro. Now, when Kaplan’s mother talks to her son about her medical treatment, she thus uses ”Dr. Shapiro” to refer to sometimes the one, sometimes the other doctor. If – counterfactually – we assume that Kaplan did not get it that two doctors by the same name are at stake, we could not assign a logical form to his utterance of, say,

(20) Dr. Shapiro is not a very intelligent man

within the ambiguity framework, because we would be unable to decide which ”Dr. Shapiro” word he is using. Consequently, we could not specify Kaplan’s subjective meaning of (20). Within the indexical framework, we do not have either of these problems. Here we would not have to disambiguate sentence (20); and its diagonal

(20a) \{c \mid \text{that object which in } w_c \text{ stands at the origin of the English usage of ”Dr. Shapiro” that is salient in } c, \text{ is not a very intelligent man in } w_c \}\n
seems to offer a good starting-point for an account of the subjective meaning the utterance has for Kaplan. For if we consider (20a) on the background of Kaplan’s other beliefs, we find that his belief set only contains contexts \(c\) in which there is only one doctor who is called Dr. Shapiro, treats his mother and is not a very intelligent man. Though the actual context is then no element of this diagonal, the latter seems to be pretty much what Kaplan wants to express with his utterance.

Apart from that, (20a) is also suited to describe the information an utterance of (20) conveys to a hearer who makes no assumptions as to who is being talked about – maybe because he is not even talked to and just overhears the utterance. For such a hearer, then, only contexts \(c\) are possible in which someone who is called Dr. Shapiro and is not very intelligent is talked about. Within the ambiguity framework, we would again have had to ascribe that hearer ignorance as to the word that has been uttered, and not as to its reference; and we would thus again have gotten into trouble if we had had to account for the subjective content the utterance has for him.

What about the problem of internalism, then? Given the above indexical semantic rule for names, the diagonals of the sentences

(16) Londres est jolie,
(17) London is not pretty

are the context-sets:

(16c) \{c \mid \text{that object which in } w_c \text{ stands at the origin of the French usage of ”Londres” which is salient in } c, \text{ is pretty in } w_c \}\,
At first, these two sets are compatible with each other. Their intersection is not empty, because there are possible contexts where English "London" is about another London – London, Ontario, for example – than French "Londres". But this reasoning already points to the snag of the matter. For in Pierre’s case, as Kripke describes it, we may take it that in a sense he does know which English usage of "London" is at stake, because he can distinguish it from all other actual English usages of "London"; the same is true of the relevant French usage of "Londres". But this means that Pierre’s belief set only contains contexts in which his utterance of "London" or "Londres" links up to the actual English usage of "London", or French usage of "Londres", whose origin in our world is the British capital. On that account, then, the intersection of (16c) and (17c) with Pierre’s belief set would, after all, again be the empty set and his belief would come out as inconsistent.

So the problem with the above rule is its reference to the actual speakers’ actual usages. For even if a subject knows the salient usage in the sense above, this does not at all mean that he knows its essential properties and is thus able to distinguish it from all counterfactual usages. In particular, Pierre knows neither the salient English usage of "London" nor the French usage of "Londres" well enough to know that they must always have the same origin; after all, he does consider contexts possible in which, by "London" and "Londres", he refers to two different cities. But these are not contexts in which only the actual English and French usages of "London" and "Londres", respectively, exist.

So again we must resort to the well-tried strategy and, along with the objective character, also define a formal character of names by abstracting from the language in question:

\[ \ll N \rrS (c) (i) = \text{that object which in } w_c \text{ stands at the origin of the usage of } N \text{ in } l_c \text{ that is salient in } c \text{ – where } l_c \text{ is } s_c \text{'s language in } w_c \text{ at } t_c, \text{ and of syntactic structure } S. \]

The problems we used to have with this on the ambiguity approach now, of course, vanish; the indexical rule for the objective character no longer makes reference to a particular usage which, in the ambiguity theory, stood in the way of abstracting from the given language.

The diagonals of the formal characters of (16) and (17) then are:

\[ (16b) \{ c | \text{that object which in } w_c \text{ stands at the origin of the usage of } "Londres" \text{ in } l^1_c \text{ that is salient in } c, \text{ is pretty in } w_c \} \],

\[ (17b) \{ c | \text{that object which in } w_c \text{ stands at the origin of the usage of } "London" \text{ in } l^2_c \text{ that is salient in } c, \text{ is not pretty in } w_c \} \].
Now their intersection with Pierre’s belief set is not the empty set any more, but rather contains Pierre’s belief that, by using “London” and “Londres”, he links up to two non-identical usages in his languages.

Having seen that the approach of treating proper names as overt indexicals leads to a more adequate account of subjective meaning than does the ambiguity theory, we must now return to objective meaning; maybe this approach again passes the comparison. However, at least two problems are causing trouble for it.

The first problem concerns the domain of the objective character. We are now in a different situation than with the ambiguity rule in that it does not seem to make sense to consider a name $N$ in a context $c$ in which $N$ is not uttered. The question of which usage of $N$ \textit{would be} salient if $N$ were uttered in $c$, could only be answered in some, but certainly not in all, cases. This means that if, instead of $c$, we consider a counterfactual context $c'$ that is like $c$ with the difference that $N$ is actually uttered in it plus all the differences resulting from that, then among the latter there will be the fact that the speaker uttering $N$ does link up to some usage of $N$, or at least tries to do so; but this does not by itself determine which usage that is going to be; and it is not clear how that could be decided in general. This problem did not arise for the ambiguity theory because there the usage was already given by the disambiguated word $N_U$. The fact that the characters of proper names are thus only well-defined for actual utterances then again leads to the same difficulties with logical consequence and synonymy as mentioned in section 3.3, which must therefore be seen as a drawback of the indexical approach. This point might appear irrelevant; but we will return to its impacts in section 4.7.

The second problem with the indexical rule for the objective characters of names concerns translation. So far we have assumed that a translation of a sentence $\phi$ of a language $L$ into a sentence $\phi'$ of a language $L'$ is correct if $\phi$ and $\phi'$ share the same objective character. The indexical interpretation of proper names, however, has the counterintuitive consequence that, e.g., sentence (22) cannot be regarded as a correct translation of (21):

\begin{align*}
(21) & \text{ Vienna is pretty,} \\
(22) & \text{ Viennes est jolie.}
\end{align*}

For according to the indexical rule, their objective characters yield different functions:

\begin{align*}
(21a) & \quad \ll_E (c) (i) = 1 \iff \text{the object which in } w_c \text{ stands at the origin of the English usage of ”Vienna” that is salient in } c, \text{ is pretty in } w_i, \\
(22a) & \quad \ll_F (c) (i) = 1 \iff \text{the object which in } w_c \text{ stands at the origin of the French usage of ”Viennes” that is salient in } c, \text{ is pretty in } w_i.
\end{align*}

Even if we choose to ignore the problem just discussed, viz. that characters of names are only partially defined so that two phonetically distinct names can never have the same character, we would still have two different functions in (21a) and (22a). There are, e.g., possible contexts where a speaker, who does refer to Austria’s capital
with his utterance of "Vienna", by "Viennes" still refers to the French town in the Rhone valley. But for such contexts the characters of (21) and (22) obviously do not deliver the same intension. So the problem is this: if proper names are taken as overt indexicals, then, unlike undeniably overt indexicals like "I" or "now", they do not necessarily have counterparts in other languages that are context-dependent in exactly the same manner.35 This is why in translation we cannot but presuppose a certain sense of the proper name – a fact which we already took as an argument in favour of the ambiguity theory above. The translation argument thus presents a serious difficulty for the indexical approach.

Finally, one point is still in need of clarification. If proper names are analyzed as overt indexicals in the way indicated, this sheds a new light on the question of their hidden indexicality. We had, as will be recalled, mainly used the problem of informativity to argue for the hidden indexicality of proper names. A sentence like

(13) Hesperus is Phosphorus

is not only informative for individual subjects, but in the English language too, which means that it could tell something new even to a person who has perfect command of English and thus knows the objective character of (13). In order to account for this case, the characters of "Hesperus" and "Phosphorus" must not be constant functions, because otherwise the diagonal of the objective character of (13) would not be a contingent proposition. However, if we regard proper names as overt indexicals, (13) becomes informative even if we do not take "Hesperus" and "Phosphorus" to be hidden indexicals. We could now interpret the concept of a usage of $N$ so as to include its origin; this would then amount to assigning the following character to a name:

$$\ll N \ll_E (c) (i) = \text{the object that stands at the origin of the English usage of } N \text{ that is salient in } c.$$  

This rule will yield the contingent proposition (13b) as the diagonal of (13):

(13b) \{c \mid \text{the object that stands at the origin of the English usage of } "\text{Hesperus}" \text{ that is salient in } c \text{ is identical with the object that stands at the origin of the English usage of } "\text{Phosphorus}" \text{ that is salient in } c \}.

But this would not be an adequate solution to the problem of informativity, though. For as we have already mentioned at the beginning of this chapter, according to (13b), (13) is, so to speak, contingent for the wrong reasons: "Hesperus is Phosphorus" is not informative in English, because quite different objects – my electrician, say, or the

35 Maybe there are actually other, undeniably indexical expressions that do not have character-equivalent counterparts in all other languages – perhaps because other languages do not always have means of expressing the relevant dependence. But then in such a case one would say that these expressions cannot really be correctly translated, and that their senses can only be approximated.
well-known matchstick company – carry the name "Hesperus" or "Phosphorus". Rather, intuitively the sentence is also – and even particularly – informative, if it is well known which usages of "Hesperus" and "Phosphorus" are salient, or if "Hesperus" and "Phosphorus" each had only one English usage. In order to do justice to this intuition, regarding them as hidden indexicals therefore remains justified and necessary.

Let us draw some tentative conclusions from our arguments for and against the indexical approach and the ambiguity theory. First, it seems unavoidable for an adequate account of subjective meaning to postulate some kind of overt indexicality of proper names; on the other hand, for certain tasks of objective meaning the ambiguity approach cannot be dispensed with either. It would thus be desirable to somehow combine these two aspects. This will appear possible once we return to the view introduced at the end of section 4.5.

If the ambiguity theory is more adequate for objective meaning, then the objective character function is defined on distinct homonymous names, distinct disambiguated words. But then, we said, individual subjects could not only be wrong about the interpretation of words but also about the words themselves. So the context-dependence described by the formal character function must not only tell which language the subject actually speaks in a given context, but also which word of that language he uses. Thus, on the indexical approach, the rule for the formal character of names must be taken to involve a double abstraction from the rule for objective character as formulated on the ambiguity approach: with respect to the subject’s language and with respect to the words of that language. A consequence of this is, however, that the objective and the formal character functions have different domains; the objective character is only defined for all realizations of a name form in the relevant language, whereas the formal character is only defined for the name forms themselves. But there is nothing wrong with this inequality; there are good reasons for assuming it.

Having drawn these conclusions, we could leave it at that – had there not been a word to be slightly shifted in the semantic rule for names. This shift will open up a host of new problems in view of which I feel unable to offer a definite conclusion. The semantics of proper names is, no doubt, surprisingly complicated.

4.7 Names als Overt Indexicals: Second Variant

The other possibility of describing the multiple usability of proper names as a form of overt indexicality emerges if, instead of starting out from the ambiguity theory and its faults, we directly draw on the analogy between proper names and demonstrative noun phrases. The view that names are to have the same logical form and the same semantic analysis as demonstrative descriptions has been advocated by Burge (1973). It can be understood as saying that, e.g., "London" or "Dr. Shapiro" are to be reconstructed as "that London" or "that Dr. Shapiro", respectively; names – or more precisely:
occurrences of names as individual terms or definite noun phrases – would thus be split into two components: a predicate or noun on the one hand, and on the other hand a demonstrative determiner. This makes it possible to account for Burge’s second motivation, viz. to develop, as far as that is possible, a unified theory covering all occurrences of names. For, as Burge points out, names are not only used as individual terms or noun phrases but also as predicates or nouns and, like these, can be combined into complex noun phrases. This is illustrated by the following examples:

(23) All Trothas are related to each other,
(24) Our Lucas gets on much better with the Joe from Munich than with the Joe from California,
(25) There are surprisingly many Spohns in Los Angeles,
(26) A certain Dreidinger called.

Burge therefore proposes to primarily regard proper names as predicates, which can become noun phrases, either by visible determiners – as in (23) to (26) – or by an invisible “that” – as in the examples discussed before.\(^\text{36}\)

It is clear that uses of names as those under (23) to (26) are not covered by the rules formulated so far. As a remedy within the ambiguity theory, one would have to assume that, apart from the individual terms of the form \(N_U\), every name form \(N\) would have a further usage \(U'\) that constitutes the predicative sense of \(N\); but then such a multiplication of words would certainly not be what Burge means by a unified theory. And whether the variant of the indexical theory given above can be reconciled with Burge’s proposal to analyze unmodified occurrences of names via an underlying name predicate, is questionable too; below we will, in fact, answer this question in the negative.

If Burge can adequately handle sentences (23) to (26), this would doubtlessly constitute a strong argument for his proposal. We are, however, chiefly interested in the question how the analysis of proper names resulting from Burge’s proposal relates to our above analyses; this shall be investigated in the sequel.

First, it must be clarified what the interpretation of a name predicate \(P_N\) is going to look like. As to this question, Burge himself says:

”A proper name is a predicate true of an object if and only if the object is given that name in an appropriate way”. (1973, p. 428)

\(^{36}\) This is not exactly Burge’s formulation. However, as his formalism diverges too much from the semantic rule format used here, I allow myself to not quote his ideas in spirit if not in letter.

\(^{37}\) These additional words \(N_U'\) would then presumably be what Kaplan (1990) calls generic names.
Roughly, this means that a name predicate $P_N$ applies to an object $x$ just in case $x$ is called $N$. Completing this formulation in terms of our earlier reflections we now obtain a precise semantic rule for the name predicate:

$$
\ll P_N \ll_E (c) (i) = \{ x \mid \text{in } w_i, x \text{ stands at the origin of an English usage of } N \text{ in } w_i \}.
$$

According to this definition, name predicates are only index-dependent, not context-dependent. Thus, e.g., the extension of the predicate "London" at an index $i$ – and at all $c$ – is the set of objects that are called "London" in English in $w_i$. But the rule needs further explanation.

First, the reference to English in the definiens must not be understood as to the present linguistic state, i.e., the state of English that we want to account for by our semantic rules and that we denote by the subscript $E$; rather it is counterfactual states of English that are intended here. After all, the rule is supposed to account for the already mentioned intuition that arbitrary objects can be given (nearly) arbitrary names. This is why in the rule I have explicitly talked of English in $w_i$.

Second, the predicate $P_N$ can only have a non-empty extension at indices where English exists and contains at least one usage of $N$. From a sentence "there is at least one London" it thus follows that the English language and the English word "London" exist. In our previous analysis it has always been important to show that the semantic rules proposed only lead to statements about objects and not to statements about words. In the case of name predicates, however, this metalinguistic aspect is unavoidable – it is just part of the content of the predicate. It will however vanish once we get to the semantics of "that $P_N$", i.e., of names as individual terms.

Third, suspicions of circularity may once again arise in connection with the above rule. For there the meaning of a name predicate $P_N$ is explained by reference to usages of $N$, where usages of $N$ are causal networks of uses of $N$ as individual terms; on the other hand we are now in turn analyzing the meaning of names as individual terms by

---

38 We are, of course, only concerned with the literal meaning of the name predicate, and not with metaphorical uses such as in "he is an Einstein" meaning "he has genius like Albert Einstein".

39 It is also conceivable to interpret $P_N$ as a covertly indexical predicate, i.e., by the rule

$$
\ll P_N \ll_E (c) (i) = \{ x \mid \text{in } w_c, x \text{ stands at the origin of an English usage of } N \text{ in } w_c \}.
$$

 Accordingly, an object would be an $N$ if, and only if, it is identical to one of the objects that are called $N$ in the context at hand. In other words: "to be $N$" then be interpreted as "to be one of those that are actually called $N$". Whether this interpretation makes sense can be tested by considering counterfactual locutions. Can one, e.g., understand "if I were an Eva-Maria ..." as "if I were one of those that are actually called 'Eva-Maria'? Or is this necessarily read as "if I were called 'Eva-Maria' (too)? What about constructions like "there are two Eva-Marias in our class; if I were an Eva-Maria too, there would be three" vs. "if I were an Eva-Maria, I would not take my place next to the other one'? But even if the second continuation should be possible, in view of the first continuation, one would at any rate need an interpretation of the name predicate where one may consider other individuals of this name in other index-worlds, so in any case one cannot do without the above rule interpreting $P_N$ as an absolute predicate.
reference to the corresponding name predicate, viz. "that \( P_N \)". But this only reflects
the explicit meta-linguistic nature of the predicate \( P_N \) that we have just noticed; just like
its alternatives discussed before, the present analysis is based on the empirical linguistic
nature of the usage associated with a name. Moreover, we had not expected earlier
definitions to be circle-free either in that they ought to offer an analysis or explication
that is illuminating and can be understood without that word.

Let us now return to the question already asked at the beginning: is it possible to
reconcile the earlier indexical semantic rule for names

\[ \ll N \ll_E (c) (i) = \mathrm{the\ object\ which\ in\ } w_c \mathrm{\ stands\ at\ the\ origin\ of\ the\ English\ usage\ of\ } N \mathrm{\ that\ is\ salient\ in\ } c. \]

with an analysis of names \( N \) as phrases of the form "that \( P_N \)", that is, to conceive of
them as the result of a semantic operation on the meaning of \( P_N \)? At first, this does not
seem to be possible. For obviously the word "that" cannot be understood as an operator
on the extension, the intension, or the character of \( P_N \), it rather appears as if it would
have to enter the definition of the character of \( P_N \) and, as it were, push the salience
condition beyond "stands at the origin" and immediately in front of "English usage of
\( N \)". There is a difficulty here that can presumably not even be overcome by more
sophisticated ideas about the compositionality of semantic rules.

Behind all this is a more fundamental problem lurking to the effect that the above
rule cannot be generalized to all demonstrative noun phrases of the form "that \( P \)". For if
\( P \) is not a name predicate, it does not make sense to define the meaning of "that \( P \)" as
above, by reference to salient usages; salience can in this case only concern the objects
themselves.\(^{40}\) So a more general semantic rule for demonstrative noun phrases ought to
be like this\(^{41}\):

\[ \ll \text{this } P \ll_E (c) (i) = \mathrm{that\ } x \mathrm{\ such\ that: } x \mathrm{\ is\ salient\ in\ } k \mathrm{\ and\ } x \in \ll P \ll_E (c) (i(c)). \]

Specialising this rule to name predicates then yields:

\(^{40}\) In this connection the following observation is of interest: Kaplan’s (1977) analysis of "that" –
including its analogous extension to complex demonstratives of the form "that \( P \)" – can be interpreted as
involving a parallel treatment of demonstratives and names, so to speak, under reverse signs, where
demonstratives are modelled on the ambiguity theory of names. For there Kaplan reconstructs every
logical form occurrence of "that" as being of the form "dthat \( \alpha \)", where \( \alpha \) stands for a contextually given
definite description that is supposed to code the mode of presentation of the object of reference. Strictly
speaking, then, he does not give a semantics for "that", but only the characters of various – so to speak,
contextually disambiguated – occurrences of "that". That the definite description \( \alpha \) acts much like the
disambiguation subscript on proper names is also confirmed by Kaplan’s remarks in (1989, p. 581): "On
this interpretation ‘dthat’ is a syntactically complete singular term, that requires no syntactical
completion by an operand. ... The description completes the character of the associated occurrence of
‘dthat’, but makes no contribution to content."

\(^{41}\) I have adopted this rule from Zimmermann (1991); see there, pp. 193f. and p. 195.
$\ll this P_N \ll_E (c) (i) = \ll N \ll_E (c) (i) =$
that object that is salient in $c$ and in $w_c$ stands at the
origin of an English usage of $N$ in $w_c$.

As the general formulation clearly reveals, the meaning of $P_N$ enters this rule in a
compositional way. Moreover, it still holds that names are rigid designators: in
determining the extension of $N$ only the context $c$, but not the index $i$, is relevant. One
also sees here that the demonstrative aspect of names (and of demonstrative noun
phrases in general) itself comes in two parts, namely, contextually given salience and
rigidity. Moreover, the above-mentioned meta-linguistic aspect has now disappeared:
only the object denoted by $N$ in $c$, not its name, enters the content of a sentence
containing a name $N$. On the other hand, we still have it that the definiens is not about
the actual state of English, but that the English language and, in particular, the English
usages of $N$ may vary from context world to context world. This point will become of
importance soon, in connection with subjective meaning.

One aspect of the above rules that has so far remained obscure is the rather vague
and general notion of salience. These rules, if they are to have empirical substance,
must obviously be supplemented by a theory of salience. For it is clear that the
descriptive component of demonstrative noun phrases, i.e., the predicate $P$, only
constrains the possible objects of reference, the actual determination of reference being
done by the demonstrative component, i.e., salience. However, elaborating a theory of
salience is a task which – like giving a precise account of the concept of a usage or of
calling – does not belong to the realm of recursive semantics but to an empirical theory
of reference. I will therefore make no attempt to answer the question of which
properties of context determine the reference of names and other demonstratives, and
will confine myself to some summarizing remarks on the relevant literature.

There are, roughly speaking, two different answers to the question of what
determines the reference of demonstratives.\footnote{See also Reimer (1992) for a survey.}
The first answer, which can, e.g., be found in Kaplan (1989), says that the crucial contextual parameter is the speaker’s
intention: if by his utterance of ”that”, ”that man”, or ”Fritz ” in a context $c$, a speaker
means a particular object $x$, then this makes $x$ salient in $c$. Another answer is given by
Wettstein (1984): By uttering a demonstrative, that object becomes salient which the
addressee may reasonably take to be salient; the qualification ”reasonably” indicates
that the hearer may rely on a body of conventional rules that the speaker must follow.

Normally both answers lead to the same result, because normally the speaker
reveals her intentions in accordance with the rules in question. There are, however,
special cases where the two approaches arrive at different results. For one thing, there
are the examples in which the speaker simply does not follow the rules: she utters ”this”
or ”Fritz ” and the information it puts at the addressee’s disposal in the situation of
utterance does not suffice to make exactly one demonstrated object or, respectively, one
Fritz salient. If, for example, I say to my neighbour: ”Fritz has unearthed your bulbs”
when there is no hint in the given situation that I do not mean my right neighbour’s son but my left neighbour’s dachshund – who would in this context both be more salient than any other Fritz – then, according to Wettstein, “Fritz” would have no reference in this context and my utterance no truth value; according to Kaplan, it would on the other hand refer to the dachshund and be true. In this case, Kaplan’s view seems more plausible to me.

Another type of example is problematic for the intentionalistic view: The speaker intends an object \( x \), but his behaviour can only be interpreted as him intending an object \( y \) that is different from \( x \) – which comes about by deception on the speaker’s side. Perhaps the best-known example for this goes back to Kaplan (1978, before he favoured the purely intentionalist view): Behind Kaplan’s desk there normally hangs a picture of Carnap. But, without Kaplan noticing, someone has replaced it by a portrait of Spiro Agnew. Now Kaplan stands behind his desk, talks to a visitor and says, pointing to the wall behind him: “That is a picture of one of the greatest philosophers of the twentieth century”. Kaplan means Carnap by “that”, whereas intuitively “that” seems to refer to Spiro Agnew.

Similar cases can be constructed for proper names too: I am sitting with people who are talking about Arthur Miller, but I am a bit distracted, not listening carefully, and thinking talk is about Henry Miller. If I then say “I saw a feature on Miller on TV last night”, then my audience will relate my utterance to Arthur Miller, although I meant Henry Miller. Here again intuition may incline towards Wettstein’s rather than Kaplan’s view.

But these two cases are more complex than they appear. For, strictly speaking, it is not that clear which reference the intentionalist view yields for “that” or “Miller”. After all, in both cases we are not just dealing with one intention – to refer to Carnap or to Henry Miller, respectively – but also with a second one that cannot be reconciled with the first one: the intention to refer to whoever one is pointing at – Agnew, that is – or whoever the others refer to – Arthur Miller, that is.

One may thus by all means hold the view that an intentionalist interpretation of salience predicts that “that” or “Miller” do not refer at all; if we interpret “that” as “the object that the speaker means and at which he is pointing” and “Miller” as “the object that the speaker means and that is called ‘Miller’”, then these definite descriptions may not be satisfied by any object at all in the contexts considered. I will leave it at this conjecture; the short discussion should suffice to elucidate the notion of salience a little.

Let us now go on with our comparison of the two variants of the indexical theory of names. We had said that our two indexical rules for proper names,

\[
\ll N \ll_{E} (c) \, (i) = \text{the object which in } w_{c} \text{ stands at the origin of the English usage of } N \text{ that is salient in } c
\]

and

\[43 \] See again Reimer (1992) on this, with her distinction of ‘primary’ vs. ‘secondary’ intentions.
Names as Overt Indexicals: Second Variant

\[ \| N \|_E (c) (i) = \text{the object that is salient in } c \text{ and in } w_c \text{ stands at the origin of an English usage of } N \text{ in } w_c \]

differ in that the first makes a connection to the ambiguity theory, whereas the second brings out the parallelism with other demonstrative noun phrases. This also involved a difference in substance: The first rule relates to the present state of English and its usages of \( N \) that are thereby given and fixed; in the second rule, on the other hand, the state of English and its usages of \( N \) vary with the context. This is the only serious substantial difference between the two rules.\(^{44}\) This raises the question of whether this difference has any new consequences for the problems that we are interested here, concerning the connection between subjective and objective meaning. Surprisingly the answer is a partial ‘yes’.

Taking the second rule to account for the open indexicality of proper names, there then seems to be no compelling reason left to regard names as hidden indexicals; at any rate, the reason given above no longer counts. The information value a sentence like (13) has in English is now represented by the diagonal (13c):

\[
(13) \quad \text{Hesperus is Phosphorus,}
\]

\[
(13c) \quad \{ c \mid \text{the object that is salient in } c \text{ and in } w_c \text{ stands at the origin of an English usage of "Hesperus" in } w_c \text{ is identical with the object that is salient in } c \text{ and in } w_c \text{ stands at the origin of an English usage of "Phosphorus" in } w_c \}.
\]

If we gave up hidden indexicality here and assumed that usages of \( N \) are individuated by their origins, then – unlike (13b) (cf. p. VVV 181) – (13c) would even be a contingent proposition if there could only be one usage of "Hesperus" and one usage of "Phosphorus" per context \( c \), or if it were known that only one sort of celestial bodies is salient in \( c \). For even under such assumptions would there be possible contexts where the object at the origin of the usage of "Hesperus" differed from that at the origin of the "Phosphoros"-usage, viz. just those that we used above to demonstrate the hidden indexicality of "Hesperus" and "Phosphorus". The reason is that such contexts admit other English usages of "Hesperus" and "Phosphorus" than does the actual state of English.

But this does not mean that all our reflections from section 4.1 have become pointless. For as we shall see, the second variant of the indexical approach also has its problems with accounting for objective meaning. To that end, the ambiguity theory still is to be preferred. If we then want to preserve the intuition that sentences like (13) are

\[\footnote{This means that the two rules would define the same function if we had the English usages of \( N \) vary with the context in the first rule or excluded that variation in the second rule. For the notions of salience of usages in the first rule and of objects in the second rule must, of course, be understood as ultimately making no difference whether one first considers the salient usage and then the object at its origin, or whether one immediately considers the salient object itself.}\]
objectively informative English, and not only for this or that subject, our framework then cannot do without the hidden indexicality of proper names.

Moreover, it appears as if we could solve the problem of internalism by the second indexical rule, without abstracting over the language in question and thus having to define a formal character in addition to the objective one. The sentences

(16) Londres est jolie,
(17) London is not pretty

now have the diagonals:

(16e) \{c \mid \text{that object that is salient in } c \text{ and in } w_c \text{ stands at the origin of a French usage of "Londres" in } w_c \text{ is pretty in } w_c\},
(17e) \{c \mid \text{that object that is salient in } c \text{ and in } w_c \text{ stands at the origin of an English usage of "London" in } w_c \text{ is not pretty in } w_c\}.

Just as (16c) and (17c) (cf. p. VVV 178) are compatible with each other, so are (16e) and (17e). But, in contrast to (16c) and (17c), we now also get a non-empty set of contexts, and hence a consistent belief content, if we intersect (16e) and (17e) with a belief set whose assumptions about salience are already so specific as to only admit for the intended senses of "Londres" and "London". The reason is, again, that the new indexicality rule admits contexts in which English and French contain respective usages of "London" and "Londres" whose origins differ from each other.

What is the more profound reason underlying this difference? Unlike the first one, the second variant of the indexical theory only takes the name predicates as lexemes of the linguistic state under consideration, whereas names used as individual terms are complex expressions whose causal-intentional history is recorded as little as that of other complex expressions. One can thus – as far as names go – ascribe competent speakers perfect command of their language without thereby demanding that they have perfect knowledge about the actual usages of \(N\) in their language. This explains why in cases like Pierre's we can do without abstraction over the language in question; knowledge of the relevant usages of \(N\) is now right away constructed as contextual knowledge.

However, an adequate account of the subjective meanings of names can actually still not do without formal character. Though the diagonal of the second indexical rule allows the English usages of \(N\) to vary with context, the English language itself, as a concrete spatio-temporally located object, must be present in every context. This means that the diagonal of a sentence with a name can only correctly represent the speaker's belief if in this sense he also fully knows the English language. But as our discussion about Twin Earth in section 3.8 has shown, we cannot assume competent speakers to have such knowledge. The problem can be made visible by constructing another doppelganger story: Let us take the variant in which Earth and Twin Earth co-exist in the same world but are not in contact with each other, and let us now assume that Peter
utters the English sentence (27) and his Twin Earth *doppelganger* utters the Twin English sentence (28):

(27) London$_E$ is pretty,
(28) London$_T$ is pretty.

Then the diagonals

(27a) \{c \mid the \ object \ that \ is \ salient \ in \ c \ and \ in \ w_c \ stands \ at \ the \ origin \ of \ an \ English \ usage \ of \ "London" \ in \ w_c \ is \ pretty \ in \ w_c\}.

(28a) \{c \mid the \ object \ that \ is \ salient \ in \ c \ and \ in \ w_c \ stands \ at \ the \ origin \ of \ a \ Twin \ English \ usage \ of \ "London" \ in \ w_c \ is \ pretty \ in \ w_c\}.

are different sets of contexts and thus cannot reflect the intuition that Peter and Twin Peter are internally identical; we would, of course have obtained the same result with the other rules for the objective character of proper names. Once again, the only remedy is to pass over to formal character, abstracting over language in the by now familiar way.

It must also be pointed out that some problems of the first variant of the indexical theory of proper names remain without change. E.g., even the second rule cannot cope with the translation argument: "that London" and "dieses London" again do not have the same character, because the two phrases do not have to refer to the same object in every context. So this problem remains a good argument for the ambiguity theory. Now, though, the double strategy deliberated at the end of section 4.6 – ambiguity theory for objective character, indexical theory for formal character – would no longer work because, unlike the first indexical rule, the second one cannot be obtained from the ambiguity rule by abstraction.

Here another problem recurs that has already been mentioned: the character defined by the second indexical rule, too, escapes counterfactual interpretation. The question which object *would* be salient if the name $N$ were uttered in a context $c$ in which it is not actually uttered, receives as little an unequivocal answer as the question of which usage of $N$ would be salient in such a case. Thus here again the character can only be defined for actual utterances of names – with the unwelcome consequences already described in section 3.3.

The final point I would like to make here, also concerns both indexical rules of names (as well as the general rule for demonstratives). Although it is crucial, I have been ignoring it up to now, because its adequate treatment would force us to make principal changes of the framework we have moved in so far. The point is due to the at first glance trivial fact that all demonstrative expressions and all proper names, too, may occur more than once in the same sentence and hence in the same context, and in a way that different occurrences also have different referents. So far only the ambiguity theory has no problems accounting for this phenomenon; it would, e.g., transpose sentence (29) into the logical form (29a):
(29) Wolfgang rings up Wolfgang,
(29a) Wolfgang\(U\) rings up Wolfgang\(U'\).

It is then clear that the two occurrences of "Wolfgang" may refer to different individuals. (They need not, for it is conceivable that, even though \(U\) and \(U'\) are distinct usages, they still have the same origin in some contexts.)

The indexical rules, on the other hand, cannot account for this phenomenon.\(^{45}\) For, on the indexical theory, (29) can only get the intuitively inadequate truth conditions (29b).

(29b) \(\parallel\) Wolfgang rings up Wolfgang \(\parallel (c) (i) = 1\) iff in \(w\) the individual that is salient in \(c\) and is called "Wolfgang" in \(w\) rings up the individual that is salient in \(c\) and is called "Wolfgang" in \(w\).

According to (29b), then, (29) comes out as true if the Wolfgang in question – provided there is exactly one – rings up himself; and if there is not exactly one, (29) becomes false, or lacks a truth value. Maybe this is a possible reading of the sentence, but it is certainly more straightforward to understand it as being about two different persons called Wolfgang.

The indexical theory also cannot give a satisfactory treatment to another frequently discussed example due to Kripke (1979, p. 265), viz. the Paderewski case:

Peter has heard about a famous pianist called Paderewski and he is inclined to seriously and honestly utter sentence (30):

(30) Paderewski had musical talent.

In another connection, Peter hears about a Polish politician called "Paderewski". Having the prejudice that politicians are in general not musical, it would naturally never enter Peter’s thoughts that this is the same person; he rather thinks this is someone else who happens to have the same name as the famous pianist. He is therefore inclined to seriously and honestly utter sentence (31):

(31) Paderewski had no musical talent.

Again – as in the London case – one would want to ascribe Peter a consistent belief content. However, the indexical rule has it that the two diagonals (30a) and (31a) must be supersets of Peter’s belief set:

(30a) \(\{c \mid \text{the object that is salient in } c \text{ and in } w \text{ stands at the origin of an English usage of "Paderewski" in } w \text{ has musical talent in } w\}\),

\(^{45}\) In what follows I will only refer to the second rule; but of course whatever I am going to say mutatis mutandis applies to the first variant too.
But (30a) and (31a) obviously have an empty intersection.\footnote{The ambiguity theory cannot solve the Paderewski case either. Since Peter – unknowingly – links up to the same usage $U$ of "Paderewski" with both of his uses, both in (30) and in (31) the name must be disambiguated by the same subscript $U$; the diagonals (30b) and (31b) are however again incompatible with each other:}

There are two ways of extending the indexical theory so as to cope with these problems. The first possibility is to define the character of names (and likewise of all other demonstrative noun phrases) only for such contexts in which the name is uttered just once, and then make reference to sufficiently many smaller context parts when it comes to defining the character of entire sentences – which may contain more than one occurrence of the name. However, this strategy leads to considerable technical complications and has therefore hardly ever been seriously suggested.\footnote{Von Stechow (1979), who took pains to elaborate this approach for the treatment of demonstrative pronouns, is an exception.}

The usual escape, rather, is possibility number two: different occurrences of the same name – or, in general, of the same demonstrative – are distinguished by different subscripts. These subscripts, however, must not be mistaken for the subscripts we have used within the ambiguity framework. For they are not supposed to stand for quite definite, once and for all given, uses of names but just indicate sameness or distinctness of reference. It does make a difference within the ambiguity framework whether one disambiguates sentence (29) as (29a) or as (29b), but within the indexical framework it is at first semantically irrelevant whether it is analyzed as (29c), (29d), or (29e)\footnote{It only becomes when we consider these sentences in larger contexts and in connection with further anaphoric references.}:

\begin{itemize}
\item (29) Wolfgang rings up Wolfgang,
\item (29a) $\text{Wolfgang}_{U}$ rings up $\text{Wolfgang}_{U'}$,
\item (29b) $\text{Wolfgang}_{U'}$ rings up $\text{Wolfgang}_{U}$,
\item (29c) $\text{Wolfgang}_{1}$ rings up $\text{Wolfgang}_{2}$,
\item (29d) $\text{Wolfgang}_{2}$ rings up $\text{Wolfgang}_{1}$,
\item (29e) $\text{Wolfgang}_{5}$ rings up $\text{Wolfgang}_{99}$.
\end{itemize}

This second kind of subscripts are usually taken as a kind of variables and thus interpreted by a variable assignment semantically. If one wants to regard this variable...
assignment as part of the context – which suggests itself because the variables in question are not bound but free – this does however have far-reaching consequences within the semantic framework adopted here. How would diagonalization have to be construed then? What is it supposed to mean that belief contents as sets of contexts are no longer represented by sets of triples \(<w,s,t>\), but as sets of quadruples containing, on top of world, speaker and time, also a variable assignment? These and a host of other questions are in need of more careful investigation.\(^\text{50}\)

One will thus arrive at a treatment of noun phrases that is closely related to what Heim (1982) and Kamp (1981) have developed as, respectively, "File-Change-Semantics" and Discourse Representation Theory. It should, however, be clear that these questions and these relations can only be analyzed in an all-embracing theory of noun phrases, which is not our present objective. Complete clarification of the semantics of proper names can, at any rate, only be given within that framework.

\(^\text{49}\) These questions are treated in Spohn (1995). There an attempt is made at justifying an extension of doxastic alternatives by variable assignments, or sequences of individuals, on epistemological grounds, not only by semantic reasoning.

\(^\text{50}\) Under the assumption that these questions can be given satisfactory answers, an adequate treatment of the Paderewski example becomes available too. Regarding contexts as quadruples \(<w_c,s_c,t_c,j_c>\), where \(j_c\) is a variable assignment, i.e., a function from the set of (individual) variables into the set of individuals, then the diagonal of sentence (32) – the conjunction of (30) and (31) – is the set (32a):

\[
(32) \text{Paderewski}_1 \text{ has musical talent and Paderewski}_2 \text{ has no musical talent,}
\]

\[
(32a) \{<w_c,s_c,t_c,j_c>| \text{ in } w_c,j_c(x_1) \text{ stands at the origin of an English usage of "Paderewski" in } w_c \text{ and has musical talent in } w_c, \text{ and in } w_c,j_c(x_2) \text{ stands at the origin of an English usage of "Paderewski" in } w_c \text{ and has no musical talent in } w_c. \}
\]

Now, if (32a) is a superset of Peter’s belief set – which now too is constructed as a set of quadruples \(<w,s,t,j>\) – and we also take it that this belief set contains the assumption that \(j_c(x_1) \neq j_c(x_2)\), then obviously Peter’s belief content has thereby been described as being consistent. It is simply said that Peter believes that there is a person \(x\) who is called "Paderewski" and has musical talent, and another person \(y\) different from \(x\), who is also called "Paderewski" and who has no musical talent.

Almog, Joseph (1981), „Dthis and Dthat: Indexicality Goes Beyond That“, Philosophical Studies 39, 347-381.


Austin, John L. (1962), How To Do Things with Words, Oxford.


Burge, Tyler (1979), „Individualism and the Mental“, Midwest Studies in Philosophy 4, 73-121.


Cook, Monte (1980), „If ‘Cat’ is a Rigid Designator, What Does it Designate?“, Philosophical Studies 37, 61-64.


Donnellan, Keith (1966), „Reference and Definite Descriptions“, Philosophical Review 75, 281-304.


Frege, Gottlob (1879), Begriffsschrift, eine der arithmetischen nachgebildete Formelsprache des reinen Denkens, Halle, wieder abgedruckt in Frege (1977).


Heim, Irene (1985), „Presupposition Projection and Anaphoric Relations in Modal and Propositional Attitude Contexts“, Manuskript, Austin.


Kaplan, David (1990), „Words“, *Proceedings of the Aristotelian Society, Supplementary Volume* 64, 93-119.


Lewis, David (1979b), „Attitudes de dicto and de se“, The Philosophical Review 88, 513-543.


Perry, John (1977), „Frege on Demonstratives“, The Philosophical Review 86, 474-497.

Perry, John (1979), „The Problem of the Essential Indexical“, Noûs 13, 3-21.


Quine, Willard Van Orman (1952), Methods of Logic, London.


Stalnaker, Robert C. (1987b), „Semantics for Belief“, *Philosophical Topics* 15, 177-190.


Wettstein, Howard (1984), „How to Bridge the Gap between Meaning and Reference“, *Synthese* 58, 63-84.

Index of Proper Names

Abbott, Barbara 121, 123
Almog, Joseph 142, 176
Austin, John L. 93
Bennett, Jonathan 93
Boër, Steven 72, 128-130
Burge, Tyler 10, 15, 41-43, 65, 97, 107, 131f., 134f., 140f., 144, 147f., 182f.
Carnap, Rudolf 9, 15, 22, 31, 110, 114
Chierchia, Gennaro 62, 70
Cook, Monte 128-130
Cresswell, Max 168
Davidson, Donald 120
Devitt, Michael 94, 148f., 152, 154, 157
Donnellan, Keith 29, 128-130
Evans, Gareth 149, 152-155, 159, 161, 174
Feyerabend, Paul 114
Fodor, Jerry 60, 139, 161
Frege, Gottlob 9, 11, 15-20, 22, 32, 54, 106, 109
Gärdenfors, Peter 59
Gazdar, Gerald 55
Grice, H. Paul 93
Heim, Irene 5, 29, 45, 52, 55, 61f., 194
Herweg, Michael 83
Kamp, Hans 27, 194
Kneale, William 161
Kornblith, Hilary 124
Kratzer, Angelika 44
Kripke, Saul 9, 11f., 15, 18f., 23, 28, 31, 33, 38, 43, 48, 65, 93, 97, 124f., 127-130, 140, 144, 147-149, 154, 159, 161, 164, 166, 168f., 174, 191
Kuhn, Thomas 114
Lerner, Jean-Yves 148f., 155, 167, 172, 176
Lewis, David 12, 26f., 38, 51, 60f., 64f., 67-71, 74f., 78, 93
Linsky, Bernhard 128-130
Loar, Brian 107, 161
Lycan, William 72
Nida-Rümelin, Martine 139
Perry, John 54, 67, 70f.
Putman, Daniel 124
Putnam, Hilary 10, 12, 15, 34, 39f., 42f., 51, 65, 87-95, 97, 99, 106-112, 114f., 118, 120-123, 126-128, 130, 132-135, 140-142, 144, 147
Quine, Willard Van Orman 15, 35-37, 102, 139f.
Recanati, François 161, 176
Reichenbach, Hans 73
Reimer, Marga 186f.
Russell, Bertrand 15, 17-19, 22, 103, 161
Salmon, Nathan 107
Schwartz, Stephen 122-124, 130
Searle, John R. 161, 167
Spohn, Wolfgang 5, 193
Stalnaker, Robert C. 11-16, 29, 35, 44-61, 71-77, 140, 149, 158
von Stechow, Arnim 5, 55, 70, 84, 168, 192
Stich, Stephen 141
Wettstein, Howard 186f.
White, Stephen 141f.
Wunderlich, Dieter 83
Zimmermann, Thomas Ede 5, 27, 61, 82, 127, 148f., 155, 167, 172, 176, 185
Index of Subjects

Evening Star/morning Star  16f., 160, see also Hesperus/Phosphorus
derived, see context
absolute, see expression
accommodation 51f.
analyticity 31, 97f., 106, 108, 119, 125f., 131, 143, 155
Aneignungskontext QQQ  141
incremental model 46, 59
anti-individualism 42, see also individualism
a posteriori, see a priori
a priori, aprioricity 10f., 31, 104f.
linguistic ~ 120, 160
division of labour, linguistic ~ 40
arbitrariness (of linguistic signs) 106, 108f.
arthritis/tarthritis 41, 140f.
attributive/referential reading 29, 71, 109, 159, 162f.
expression
  absolute ~ 28
  absolutely rigid ~ 28
  descriptive ~ 29
  directly referential ~ 21, 28, 127
  indexical ~ 20f., 25, 28, 54, 65, 127
token-reflexive ~ 73
utterance 27, 44, 78, 100f.
  place of ~ 81f.
utterance situation, see context
utterance time 78
circumstances of evaluation, see index

meaning
  objective ~ 9f., 13-16, 20, 23, 31, 39-43,
  71, 94, 132, 135, 181, 188
  subjective ~ 9-17, 20, 31-33, 39-46, 55,
  58-60, 64, 71, 75, 86f., 94, 132-135,
  142f., 167, 169, 174, 181, 188, 190
  theory of ~, see theory of reference,
  pragmatics, semantics
  metalinguistic ~ 113, 184, see also meta-
  language objection
semantic change 113f.

semantic state (of a language) 108f.
concept
  linguistic vs. non-linguistic ~ 138f.
  subjective ~ 138f.
  scientific ~ 114
assertion 44f.
acquaintance, direct 103

character 11, 27, 57-60, 70, 87-93, 135, 167
  change of ~ 113f.
  (functional) domain of ~ 99-102, 142,
  150, 173, 179, 190f.
  formal ~ 13, 136f., 140-143, 158,
  168f., 175-179, 181, 189f.
  objective ~ 136f., 140, 142, 149f., 168,
  172, 176, 185, 189
  partially defined ~ 97
  partial ~ 141f.
  propositional concept vs. ~ 47-51, 57, 60
character function 31f.
  domain of ~ 181f.
Charity, Principle of 120
Cicero/Tullius 166f.
Common Currency Conception 150f.

de dicto, de re, de nunc, de se, see belief, belief
ascription
deduction problem 35
designation chain XXX 148
designator, see rigid
descriptive, see expression, name
descriptive content, see name
diagonal 11f., 30, 59, 61, 71, 113
  the contradictory ~ 30
  the tautological ~ 30
  informative ~ 30, 54
diagonalisation 52-55, 57, 95, 167
diagonal operator 29-31, 52-54, 60f.
directly referential, see expression
disjunction problem 139
Discourse Representation Theory 194
Sachverzeichnis

disquotational principle 33, 41, 145
double indexing 26f.
doxastic alternative 12, 34, 64-67, 75, 193
structural identity of ~ and contexts 61f.
dth-at-operator 29

proper name, see name

property 23, 27
~ vs. concept 138
essential ~ 116f., 128
pertinence 181, 185-187
attitude, see belief
attitude ascription, see belief ascription
empiricism 114
cognitive value, see belief ascription
expert 151f., 156
extension 22, 28
extensional, see construction

fiction, see name, predicate
fictional object 126, 158
File-Change-Semantics 194
formal, see character
multiple grounding 152
function term 87, 123f.

thought, see belief content
thought, Fregean 9
time parameter 61, 76f.
generic, see name
belief
~ de nunc 60f., 77
~ de se 12, 60, 67-72
propositional ~ 68f.
belief content 34f., 42f., 60f., 76
belief set 12, 34f., 58, 61, 64, 138
belief revision 59
belief ascription 33f.
   de-dicto~ 13, 15, 35-38, 41-43, 144f., 169
   de-re~ 36f., 139
identity, internal 132-134, 144

haecceitism 69, 74, 77
calling 93, see also usage
Hesperus/Phosphorus 19, 35, 50, 165f., 180f., 188f., see also Evening Star/Morning Star
here, see place parameter
holism (as to subjective meaning) 11, 59
hearer/speaker-perspective 33, 62, 75
hyperintensional, see construction
H2O/XYZ 39, 88-92, 96-99, 133

I, see belief de se, subject parameter
identity statement 16f., 20f., 106, 129, 155
implicature 52
index 24-27, 47
   ~ parameter 26, 60
   structural identity of ~ and context 29f., 61
indexical, see expression
indexicality 11, 27, 149
   overt ~ 12, 89, 97
   hidden ~ 12, 40, 87, 89, 97, 112, 114, 119, 126f., 130, see also name
individualism 9f., 34, 135, see also anti-individualism
individual concept 23, 27
informativity 12, 45, 168, 177, see also diagonal, informative
content, see belief content, intension
intension 22f., 27
   partially defined ~ 23, 28
intensional, see isomorphism, construction, operator
internal, see identity, consistency
internalism, problem of ~ 10-13, 42f., 55, 64-66, 75, 87, 132, 140, 167-169, 174-178, 189
isomorphism
   character ~ 110
   intensional ~ 110

now, see time parameter, belief de nunc
causal, see theory of reference
description, definite ~, see attributive/referential
cognitive significance, see subjective meaning
chain of communication 148, 174
Compositionality, Principle of ~
at character level 30
   Extensional ~ 22
   Intensional ~ 24
consistency, internal 38
construction
   characterial ~ 30f.
   extensional ~ 22
   hyperintensional ~ 35
   intensional ~ 24
   local ~ 24, 82-84
   temporal ~ 24, 78-80
consument of a name usage 152
<table>
<thead>
<tr>
<th>context</th>
<th>25-27, 47, 99-102, see also construction derived ~ 47 structural identity of ~ and doxastic alternatives 61f. parts of ~ 192 context dependence, see indexicality context set, see conversational background context change potential 45f., 55 context change theory 15, 55 contingency, see necessity</th>
</tr>
</thead>
<tbody>
<tr>
<td>layman</td>
<td>151f., 156 empty, see name, predicate logically proper name, see name logical consequence 100 logical truth 31 local, see construction Londres/London 38, 167, 169, 175, 178f.</td>
</tr>
<tr>
<td>metasemantics</td>
<td>104 meta-language objection 105-109, 113, 161-165</td>
</tr>
<tr>
<td>monster</td>
<td>see construction, characterial</td>
</tr>
<tr>
<td>name</td>
<td>13, 18f., 38f., 48-50, 147-149, see also usage as ambiguous expression 13, 147, 170-174, 181, 192 as demonstrative noun phrase 182-185, 188-192 as overtly indexical expression 13, 147, 170, 176-182, 188, 191 descriptive ~ 159 descriptive content of a ~ 160f. fictional ~ 158f. generic ~ 183 ordinary ~ 105 homonymous ~ 170 identity conditions for homonymous ~ 172 empty ~ 157-159 logically proper ~ 103-105 multiple occurrence of ~ in a context 191-193 hidden indexicality of ~ 154, 172, 188 name form 172, 174 name predicate 183f. name tradition 148, see also usage noun phrase demonstrative noun phrase 185, see also name</td>
</tr>
</tbody>
</table>
language 136f.
   inseparability of ~ and theory 140
language acquisition 138f.
speaker/hearer-perspective 33, 62, 75
speaker presupposition 44
rigid designator 19, 21, 23, 28, 87, 95, 126-130, 147, 150
    strictly ~ 23
stereotype 40, 142, 160
structured proposition, see proposition
subjective, see meaning, concept
subject parameter 60-62
natural kind term 12, 19, 39f., 112-115, 120
substitutivity of identicals 17

temporal, see construction
token-reflexive, see expression
tradition 148, see also usage

translation 109-112, 171, 179f.
translation argument 107, 109-112, 162f., 180, 190
    individual variant of ~ 107, 113, 144
translation principle 33
re-interpretation 51f.

variable assignment 61, 193
    ~ as contextual parameter 62, 193

truth 28
    logical ~ 31
truth condition, see objective meaning
essential, see property
essentiality convention 115-120, 153
knowledge
    semantic ~ 12f., 40, 58, 87, 139-142
    incomplete ~ of utterance context 57, 135
    incomplete ~ of one’s language 131-137
    ~ of oneself 66f.
scientific concept 114
word
    identity criterion for ~ 91f.

centered world 74