



Shelf Mark Theses Series: 2

MANOLAKAKI, Ph.D. 2000



**TRUTH-EVALUABILITY in
RADICAL INTERPRETATION**

Eleni Manolakaki

A thesis submitted in fulfilment of the requirements
for the degree of Doctor of Philosophy
to the
University of Edinburgh

July 2000

ABSTRACT

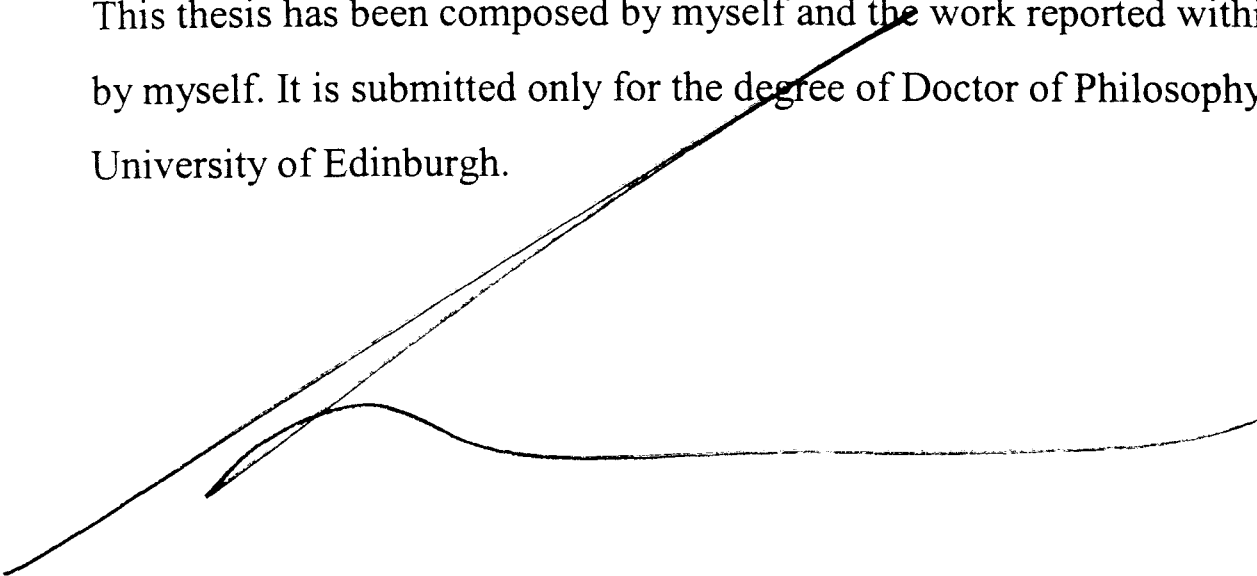
The central problem of the dissertation concerns the possibility of a distinction between truth-evaluable and non-truth-evaluable utterances of a natural language. The class of truth-evaluable utterances includes assertions, conjectures and other kinds of speech act susceptible of truth evaluation. The class of non-truth-evaluable utterances includes commands, exhortations, wishes i.e. utterances not evaluated as being true or false. The problem is placed in the context of radical interpretation theory and it is shown that it is a substantial problem of Davidson's early theory of radical interpretation.

I consider the possibility of distinguishing between locutionary and illocutionary act in uttering a sentence and its significance in the present project. I discuss the suggestion that the mood of the verb of the sentence signifies the required distinction between truth-evaluable utterances and non-truth-evaluable ones. I argue that no criterion for the distinction based on the mood of the verb is adequate.

The solution that I propose to the problem of classifying truth-evaluable utterances appeals to mental states. The view that grounds this line of inquiry is that the truth-evaluability of an utterance is a characteristic of it exclusively relevant to the doxastic dimension of the speaker's mind. I discuss the constraints that the nature of radical interpretation puts upon the way we construe the notion of belief. I propose that a possible classification of mental states into doxastic and non-doxastic that would result in a classification of utterances into truth-evaluable and non-truth-evaluable ones can be given by an elaborated version of a decision theoretic scheme. I suggest that a decision theoretic scheme based on a decision theory that, like Savage's theory, grants independence axioms is a better candidate to offer a solution to the central problem of the dissertation than a scheme based on a non-standard decision theory such as Richard Jeffrey's. I conclude by showing that the proposal I make satisfies the constraints I have considered and that it can be accommodated by a radical interpretation theory.

Declaration

This thesis has been composed by myself and the work reported within was executed by myself. It is submitted only for the degree of Doctor of Philosophy to the University of Edinburgh.



Acknowledgements

I am greatly indebted to Dr Peter Milne, my principal supervisor, for his genuine interest in this project, his constant encouragement and support, his constructive comments and bright suggestions.

I am deeply indebted to Dr Richard Holton who generously offered me help in revising the thesis. I am grateful to him for his insightful suggestions for improvement, the constructive long discussions he was always willing to have with me, his interest in the thesis and many more that is difficult to remember.

I am thankful to Professor Jane Heal and Professor Tim Williamson for the very constructive comments they offered to me both in my oral examination and in their recommendation.

I would like to thank Mr William Charlton for supervising me at the beginning of the project, Dr Paul Schweitzer who gave me useful comments on an earlier draft of the thesis and Dr Denis Walsh who gave me constructive comments on parts of the thesis. Special thanks to Professor Dionysios Anapolitanos for his insightful questions on the dissertation and for his direction and support.

I thank the Greek National Scholarship Foundation (I.K.Y) for financing this project.

Contents

Abstract	page i
Declaration	page ii
Acknowledgements	page iii
Introduction	page 1
Chapter One:	
Radical interpretation theory	
1.1 Introduction to interpretation theory	page 10
1.2 Davidson's theory of radical interpretation	page 18
1.2.1 Davidson's early theory of interpretation	page 21
1.2.2 Davidson's later theory of interpretation	page 33
Chapter Two:	
Speech act analysis	
2.1 Introduction	page 39
2.2 Criteria that do not appeal to the notion of truth	page 41
2.3 Locutionary meaning as the truth-value bearer	page 54
Chapter three:	
The mood criterion and the phenomenon of mood	
3.1 Introduction	page 63
3.2 The criterion of mood	page 66
3.3 The phenomenon of mood and truth-conditional semantics	page 74
3.3.1 Non-indicatives as explicit performatives	page 74
3.3.2 An account of mood appealing to two notions of truth	page 77
3.3.3 The paratactic analysis of mood	page 82

Chapter four:

Appealing to belief

- 4.1 Utterances and attitudes page 88
- 4.2 Truth-evaluable utterances and belief page 91
- 4.3 Belief a relation to a sentence page 94
- 4.4 Belief as a disposition to act page 103
- 4.5 Belief as a functional state page 104

Chapter five:

Identity of content across attitudes

- 5.1 Introduction page 109
- 5.2 Identity conditions of content of functionally individuated states page 111
- 5.3 Relations to sentences page 114
- 5.4 A logical analysis of intentional state page 117
 - 5.4.1 Identity of content across state in Searle's analysis of intentional state page 118
 - 5.4.2 Direction of fit page 123

Chapter six:

The decision theory approach

- 6.1 Introduction page 133
- 6.2 Bruno de Finetti's and Frank Ramsey's theories of subjective probability page 135
- 6.3 Savage's axiomatised decision theory page 140
- 6.4 Jeffrey's decision theory page 164
- 6.5 A comparison between the theories page 168

Chapter seven:

Decision theory and radical interpretation

- | | | |
|-----|--|----------|
| 7.1 | Integration of decision with interpretation | page 171 |
| 7.2 | Jeffrey's radical decision theory | page 180 |
| 7.3 | A Savage-style radical decision theory | page 187 |
| 7.4 | A comparison between Jeffrey's
and Savage's radical decision theories | page 191 |
| 7.5 | Some closing remarks | page 195 |

- | | |
|-------------------|----------|
| References | page 197 |
|-------------------|----------|

INTRODUCTION

This dissertation is about truth-value susceptibility and the doxastic dimension of the mind. These subjects are discussed from the point of view of a certain question. The central question in the dissertation concerns the possibility of delimiting a certain range of utterances of a natural language, the utterances that are evaluated as true or false, and the possibility of classifying a certain range of states/attitudes of a speaker or agent, the doxastic states/attitudes. I consider the possibilities of these two classifications to be related. The utterances that are susceptible of truth-evaluation are related to a certain category of attitudes/states that includes beliefs and partial beliefs. Utterances are related to attitudes/states by a way of expressing an attitude/state or by a way of representing one to have an attitude/state. Classifying the doxastic states of an agent would contribute to a classification of the possible or actual utterances of his that are susceptible of truth-evaluation. I consider the problem of a classification of truth-evaluable utterances of a speaker and the problem of a classification of the doxastic states of a speaker/agent within the context of a radical interpretation theory.

The possibility of a classification of truth-evaluable utterances should be kept apart from relative and connected problems. It should be kept apart from problems concerning the attribution of truth-value to sentences in which occurring terms lack reference. And it should be kept apart from problems concerning cases in which it is in practice or in principle impossible to determine the truth-value of a sentence. The possibility of the classification considered here is a different inquiry from a classification of utterances into true and false resulting by the determination of their truth-values. The question asked in this dissertation is how it is possible to delimit the range of utterances for which a question about their truth-evaluation arises at all.

The assumption that there is a class of utterances characteristically susceptible of truth-evaluation is not unusual in contemporary philosophy. Assertive, fact-stating, descriptive utterances are taken to be evaluated as true or false. Imperative, directive, prescriptive utterances are not susceptible of truth-evaluation. Up to a certain extent and in a way that will be clarified the dissertation is about the possibility of delimiting the class of utterances that includes assertive, descriptive, fact-stating utterances in

contrast to the class of utterances that includes imperative, directive, prescriptive ones.

The assumption that not all utterances of a natural language are truth-evaluable is a basic assumption I make in the dissertation. Two kinds of ground for holding this assumption are based on common sense applicability of the values true and false. One ground of this kind relates to physical features of the utterance like syntactical or grammatical features of the sentence uttered. Some other physical features of an utterance, like the intonation of the voice of the speaker in making the utterance, the way the sentence uttered is stressed might be considered. The common sense way in which the values true and false apply suggests that in many cases an utterance of a sentence in the indicative mood or a sentence uttered with a certain intonation or stressed in a certain manner are cases of utterances susceptible of a truth-value. The utterance of the sentence "The door is shut." is eligible for eliciting the evaluating response 'this is true' or 'this is false'. Utterances of sentences in another grammatical mood are usually not eligible for this kind of evaluating response. The utterance of the sentence "Shut the door!" is neither eligible for eliciting the response 'this is true' nor eligible for eliciting the response 'this is not true'. Common sense considerations such as these indicate that an utterance of the sentence "Shut the door!" is beyond the domain of utterances to which the value 'true' is applicable.

There is another ground for the assumption based again on common sense considerations about the applicability of values true and false. This ground concerns the kind of attitude/state attributed to the speaker in relation to an actual or possible utterance of his. We consider the premise that the truth-evaluable utterances are related to a distinctive category of attitudes/states, namely the doxastic category of attitudes that includes belief and partial belief. There is a common intuition that beliefs are attitudes/states susceptible of truth-evaluation. Attitudes/states that are not included in this category like desires, intentions, preferences are not evaluated as true or false. So, common intuitions about the applicability of the values true and false suggest that if the attitude expressed by a speaker in making an utterance or attributed to a speaker in understanding his utterance is of the doxastic kind, then the utterance is a truth-evaluable one. Utterances that are interpreted as related to attitudes labelled here affective and which include desires, intentions, preferences are not truth-

evaluable. Attitudes from this category are also called conative, evaluative, directive attitudes.

These considerations tend to justify the basic assumption I make, that is that some utterances are truth-evaluable while others are not. The assumption allows me to raise the question about the possibility of classifying the truth-evaluable utterances. But despite the observation that truth-value susceptibility is relevant only to a certain class of utterances the suggestion that not all utterances are susceptible of truth-evaluation has been challenged. The arguments on the basis of which the assumption is challenged intend to show that utterances seeming to be non-truth-evaluable can be reduced to truth-evaluable ones. The reduction would be of a certain kind relevant to some aspects of the utterance. The assumption is challenged is by challenging the two grounds on which we could hold the assumption.

Concerning the ground for the assumption which is based on the syntactic or other physical properties of utterances, the arguments consist of a specification of ways in which it might be possible to establish a semantic equivalence between the alleged non-truth-evaluable syntactic category and the truth-evaluable one. For instance, it has been proposed that a sentence in the imperative mood is semantically equivalent to another sentence in the indicative mood.

Concerning the ground for holding the assumption based on the kind of attitude related to an utterance the arguments against it run again by considering an equivalence between belief states and desire states. The kind of the equivalence depends on the particular theory of mental terms adopted. The arguments in this case intend to show that desiring such-and such to be the case is equivalent to believing that so-and-so is the case. The arguments purport to show that desires are beliefs, from which it follows that the corresponding attitude attributions are equivalent in the particular respect.

In most of the following I will take the assumption of a distinction between truth-evaluable and not truth-evaluable utterances for granted. Like a big group of philosophers who work under this assumption and in accordance with common sense considerations I will leave aside worries about a justification of the assumption.

The common sense applicability of the values true and false hints at two classes

of utterances, the truth-evaluable ones and the non-truth-evaluable ones. In addition to common sense use of the terms true and false we could appeal to further background intuitions that give a shape to the perspective taken in this dissertation. In the background of the present perspective there is the suggestion that the distinction a truth-evaluable/non-truth-evaluable classification introduces reflects a substantial distinction within intentional concepts. The suggestion is that it is not an accident of natural languages that they permit evaluation by truth and falsity only to a certain range of utterances. It is not simply a matter of grammatical generosity that there are more than one grammatical moods, and that one of them, namely the indicative, is systematically related to utterances susceptible of truth-evaluation. It is not a matter of linguistic perversity that reports of some utterances by indirect speech are constructed with the subordinate clause in the indicative mood, while reports of some other utterances are constructed with a subordinate clause containing an infinitive. The suggestion is that the linguistic discriminations to which truth-value susceptibility is sensitive are the linguistic symptoms of a distinction within the intentional, a distinction in the intentional relevant ways humans interact with their environment. The distinction is intuitively grasped as one between doxastic and affective intentional states and it is sometimes labelled as a difference in the direction of fit within intentional states. Beliefs, suspicions, assumptions, conjectures, that is states related to truth-evaluable utterances, represent certain reactions to one's environment. They characterise the way individuals take the world to be. Desires, intentions, preferences represent different kind reactions to one's environment. They characterise the ways individuals consider their intervention in the world. It is ultimately the difference between two kinds of interaction between individuals and the world to which the truth-value susceptibility is here taken to be sensitive.

The linguistic discriminations to which truth-value susceptibility is, from a common sense point of view, sensitive trigger the investigation for a characterisation of the deeper distinctions they represent in human interaction with the environment. The characterisation of the doxastic intentional states and its distinction from the affective intentional states is the ultimate aim of the project. Within this perspective, a good answer to the possibility of a classification of truth-evaluable/non-truth-evaluable distinction would be one that enables a grasp of the distinction in

intentionally described interactions between individuals and their environment.

The problem of the classification of truth-evaluable utterances and of doxastic and affective states finds its proper place in a theory of understanding. Intentional concepts like beliefs and desires are central in a theory of understanding. In this dissertation it is pointed out that the characterisation of truth-evaluable/non-truth-evaluable utterances and doxastic and affective attitudes is significant for a theory of understanding speech and other actions of individuals. An identification of an utterance as a truth-evaluable or non-truth-evaluable one and an identification of the attitude related to an utterance as a doxastic or affective one is necessary for understanding speech and other action of individuals. This is a reason for which the problem of the classification of truth-evaluable utterances and attitudes/states has a place in the context of a radical interpretation theory.

The interpretative perspective adopted in the dissertation is basically along Davidson's theory of radical interpretation. There are central assumptions of Davidson's radical interpretation theory that in the dissertation are held back assent. Concerning the issue of the place of the holding true attitude in Davidson's early theory of interpretation, the issue of the common content assumption as a presupposition for a paratactic analysis of indirect speech and propositional attitude reports, and Davidson's suggestion for a Jeffrey-style radical decision theory there is divergence from Davidson's views in the dissertation. But those issues as well as most of the rest considered and discussed in the dissertation are envisaged with reference to Davidson's radical interpretation theory.

The problem of a classification of truth-evaluable utterances has an additional significance for theories of interpretation such as Davidson's theory that takes the notion of truth to be the key notion for semantic interpretation of a language and for theories such as Davidson's early theory that takes the attitude of holding a sentence true or the concept of believing a sentence to be true to provide the bridge for the construction and choice of a correct theory of interpretation. Davidson's early theory of radical interpretation might be criticised for taking for granted notions such as the notion of holding a sentence true or the notion of belief, of which a theory of radical interpretation should account. The attempt we make here for a classification of truth-evaluable utterances and doxastic states could then be seen as an attempt to free a

radical interpretation theory of this charge. The proposal that we will make in this dissertation is that a classification between the doxastic and the affective attitudes would be given by the attitude of preference between options.

The attempt we make here does not amount to an attempt to reduce beliefs to behaviour or to reduce the intentional to the non-intentional. It is not a reduction of this kind that is attempted here. The problem to which the proposal is relevant concerns a distinction among the intentional concepts employed by a theory of interpretation that would suffice as a theory of understanding. Identifying the doxastic intentional concepts has a significant role in interpretation theory. The attempt we make here concerns a distinction to be drawn within the intentional.

Some comments about the place of the notion of proposition in the present project of a classification of truth-evaluable utterances and doxastic states are due here. Propositions are usually considered to be the bearers of truth and falsity and the objects of attitudes, the contents of mental states, the references of that-clauses, or what sentences express.

The assumption that propositions construed as the objects of attitudes and what utterances express are also the primary bearers of truth and falsity does not have a place in the present project. The question in this dissertation is the possibility of a characterisation of those utterances that are susceptible of truth evaluation and the possibility of a classification of the related attitudes/states, the doxastic states/attitudes. This question arises only under the assumption that not all utterances of individuals are susceptible of truth evaluation. Taking propositions as the objects of attitudes/what is expressed by utterances to be also the primary bearers of truth or falsity brings us to the following. Either a) not all utterances express propositions and not all attitudes have propositional contents or b) all utterances express propositions and all attitudes have propositional contents but the truth-evaluability of utterances/attitudes is not grounded on the truth-evaluability of propositions. In case a) the assumption that propositions as what is expressed by utterances or as the objects of attitudes are also the primary bearers of truth and falsity is question begging. The problem becomes how it is possible to characterise the utterances which express propositions or how to characterise the attitudes that have propositional contents. In case b) the assumption that propositions are the primary bearers of truth

and falsity becomes irrelevant to the problem of characterising the truth-evaluable utterances and attitudes of individuals. There would be utterances/attitudes that though they might express or determine propositions are not truth-evaluable.

The considerations made in the dissertation in relation to the notion of proposition have as subjects the notions of locutionary meaning of an utterance and the content of an attitude/state. The issue of the truth-value susceptibility of these two notions will be envisaged from the point of view of a certain question. That is, the question of whether common locutionary meaning and common content can be identified across truth-evaluable/non-truth-evaluable utterances and attitudes. The whole discussion touches upon the issue of whether the kind of utterance/attitude as a truth-evaluable or non-truth-evaluable one has a bearing on intentional content.

In the first chapter of the dissertation I present the framework of a theory of radical interpretation. The questions I discuss concern the aim of a theory of interpretation and the constraints of the theory. In particular, I discuss Donald Davidson's early and late theories of radical interpretation. I place the problem of the possibility of a classification of truth-evaluable utterances and of doxastic states in the context of those theories. I argue that the problem of identifying the truth-evaluable utterances of speakers and of identifying the doxastic states of speakers/agents is a substantial problem for radical interpretation theories.

In the second chapter of the dissertation I discuss issues related to the enterprise of linguistic interpretation. On the basis of an analysis of the notion of speech act offered by J. L. Austin I inquire into identification of the interpretation relevant aspects of a speech act. I suggest that within the present project of inquiry there are difficulties with identifying locutions across two certain categories of speech act. I also suggest that the interpretation relevant aspects of a speech act are locutions identified as of a certain kind.

In the third chapter of the dissertation I discuss a criterion for the required distinction between truth-evaluable and non-truth-evaluable utterances based on the phenomenon of variety of grammatical mood. I suggest that the required classification cannot be based on a criterion of mood. In the same chapter I also discuss the possibility of reducing the variety of grammatical mood to one category, namely the

indicative, and I discuss the possibility of accounting for the variety of grammatical mood within the context of truth-conditional semantics. I point out that there are problems with the attempts to account for the variety of grammatical mood within truth-conditional semantics.

In the fourth chapter of the dissertation I introduce the notion of belief and of a state/attitude to the present inquiry. In this chapter I make the suggestion that the required classification between truth-evaluable/non-truth-evaluable utterances will be given by a classification of attitudes/states. I consider various ways we could construe the notion of belief and of an attitude/state. I discuss these suggestions in the light of adopting them in the present project of classifying truth-evaluable utterances and doxastic states in the framework of an interpretation theory.

In the fifth chapter of the dissertation I consider the assumption that doxastic and affective attitudes can share a common content. I consider conditions for identifying content across doxastic and affective attitudes. The conditions for identification of content considered are i) content identified by functional role, ii) content identified as a mental sentence and iii) content identified by conditions of satisfaction. I point out that the criteria i), ii) for identification of content do not grant identity of content across doxastic and affective states. Criterion iii) requires an account of the difference in direction of fit. I critically examine some accounts of the difference in direction of fit and I point out that they do not satisfy required conditions. I suggest a) that the common content assumption might not be properly established and thus it might not be indisputable and b) that an account of the difference in direction of fit needs to exploit a difference in the psychological dimension of doxastic and affective states/attitudes.

In the sixth chapter of the dissertation I consider and propose a possible approach to the notion of doxastic state as introduced by decision theoretic considerations. Doxastic states are ascribed a characteristic role in explaining deliberate behaviour of agents accounted by decision theoretic considerations. Decision theoretic considerations suggest an account of agent rationality based on constraints on the range of preferences of agents. I consider some theories of decision and I critically examine them under the light of their possible contribution to the present project. I explain how decision theoretic considerations result in the required

classification between doxastic and affective attitudes of speakers/agents.

In the seventh and final chapter of the dissertation I make a positive proposal concerning a classification of doxastic states of an agent/speaker in the context of radical interpretation. I suggest the way decision theoretic considerations incorporated into a radical interpretation theory would answer the central question of the dissertation. The proposal is that an integration of decision theory with interpretation theory would result in a classification of truth-evaluable utterances and states/attitudes. In particular I suggest that that the integration of decision theory with radical interpretation theory takes the form of a Savage-style radical decision theory. I present the alternative suggestion of a Jeffrey-style radical decision theory and I present the reasons for which a Savage-style radical decision theory is preferable to a Jeffrey-style radical decision theory.

CHAPTER ONE

RADICAL INTERPRETATION THEORY

In this chapter I introduce interpretation theory. In the first section I present the aim of interpretation, I consider the suggestion that interpretation is given by certain redescriptions and I discuss the relation between interpretation of language and interpretation of behaviour. I continue by considering certain constraints on the redescriptions required by a theory of interpretation. In the second section of this chapter I introduce Davidson's theory of radical interpretation. The discussion of Davidson's theory is divided into two parts. In the first part I present Davidson's early theory of interpretation where the empirical primitive of the theory is suggested to be the attitude of *holding true* a sentence. I point out that identification of the *holding true* attitude is a substantial problem for Davidson's theory of interpretation. In the second part of my discussion of Davidson's theory I present Davidson's later theory of interpretation. In this theory the empirical primitive is the attitude of preference between the truth of sentences. I discuss the importance of the shift between the two theories and I consider the problem of the possibility of a classification of doxastic/affective attitudes in the later theory.

1.1 Introduction to interpretation theory

A theory of interpretation is a theory of understanding the actions of speakers and agents. Interpreting speech and actions of agents' amounts to making sense of linguistic and other behaviour. A speaker emits the noise "Gib mir wasser!", an agent waves his hand in the presence of someone else. We interpret the noise the speaker is making and the behaviour the agent is showing by making sense of what the speaker and agent are doing. Considering the context of the utterance "Gib mir wasser!" and the circumstances in which the agent waves his hand we might understand that the speaker is asking for water and that the agent is greeting a friend. A theory that enables understanding behaviour would be a theory of interpretation.

There is a general suggestion about how to go about in making sense of speech and other behaviour. The suggestion is that the key to understanding behaviour is given by certain redescriptions and, accordingly, the project of

interpretation involves an enterprise of redescription. Davidson says the following:

“We know that the words ‘Es schneit’ have been uttered on a particular occasion and we want to redescribe this uttering as an act of saying that it is snowing” (1974, p. 141).

“A theory of interpretation, like a theory of action, allows us to redescribe certain events in a revealing way” (1975, p. 161).

Understanding behaviour of agents requires identifying the actions they perform in the way they behave. On a certain occasion an agent waves his hand. Seeing his waving as an action of greeting a friend would enable understanding of what the agent is doing. Seeing the waving of the hand as the action of greeting a friend amounts to attributing to the agent the desire to greet a friend and the belief that by waving his hand he greets his friend. Interpreting the behaviour of an agent is redescribing it as action that can be explained and understood by the intentions, wants and beliefs of the agent. A theory about agents’ behaviour that results in redescrptions of pieces of behaviour as intentional actions that are intelligible in the light of the agents’ propositional attitudes is a theory of interpretation of agents’ behaviour. Attributions of propositional attitudes to an agent show the intelligibility of the agent’s behaviour by showing it to be rational action. We understand the behaviour of individuals by seeing them to be rational agents. Identifying behaviour of agents as action intelligible in the light of agents’ propositional attitudes would enable a distinction between intentional action and non-intentional behaviour like the twinkling of the eye or the growing of one’s hair.

Interpretation of speech deals with utterances of sounds. Interpretation of utterances of sounds requires identification of sounds as utterances of sentences of a language that are performances of speech acts. The suggestion concerning interpretation of speech is that a theory that results in redescrptions of uninterpreted utterances as certain kinds of speech act is eligible for a theory of understanding the language of speakers. Identifying that a speaker in uttering “Gib mir wasser!” asked for water and that a speaker in uttering “Es schneit.” said that it is snowing would enable understanding the speech behaviour of the speakers. The input of a theory of interpretation of language consists in uninterpreted utterances that agents, speakers of the language, make. The output of the theory consists in redescrptions of every uninterpreted utterance of the language of speakers; every utterance is redescribed as a certain kind of speech act. A theory of this kind, namely a theory that results in

redescriptions of utterances as speech acts, and that enables understanding of the speech behaviour of speakers would be a theory of interpretation of the language of speakers. Such a theory would enable a distinction between meaningful utterances of speakers of a language and utterances of meaningless sounds like a yawning.

One constraint on a theory of speech interpretation is that it should deliver the required redescriptions of utterances in a systematic manner. A redescription of sounds as an utterance of a sentence of a language employed for the performance of a speech act must be given in a systematic way. Natural languages contain an infinite number of actual or potential utterances that are composed of a finite vocabulary. A theory that would suffice for interpretation of language must enable understanding any utterance of speakers, actual or potential, heard or unheard. Finding structure in an utterance amounts to identifying its component parts and the way they compose the utterance. A theory of language interpretation that distinguishes between the component parts of an utterance and way of composition of the utterance would deliver the required redescriptions in a systematic way.

Another constraint on a theory of language interpretation is that the theory should be such that it can be employed for understanding speech behaviour. As with interpretation of behaviour in general, understanding speech behaviour of an agent/speaker is enabled by redescriptions of utterances as certain sorts of speech act that are seen to be intelligible in the light of what the speaker believes, wants and intents. Attributions of beliefs, wants and intentions to the speaker show the intelligibility of his speech acts by showing the speaker to be rational. As with the enterprise of behaviour understanding, making sense of speech behaviour requires attribution of rationality to the speaker/agent.

The considerations made above show a relation between interpretation of language and intelligibility of the lives of speakers/agents. For a theory resulting in the required redescriptions to be acceptable as a theory of language understanding the following condition should hold. The theory, as applied to speakers, should show their lives intelligible by showing them to be rational agents. Davidson puts it in the following way:

“The interlocking of the theory of action with interpretation will emerge in another way if we ask how a method of interpretation is tested. In the end, the answer must be that it helps bring order into

our understanding of behaviour.” (1975, pp.161).

John McDowell (1977) glossing on theories of linguistic understanding suggests the following:

“The adequacy of the total theory would turn on its acceptably imposing descriptions, reporting behaviour as performance of speech-acts of specified kinds with specified contents, on a range of potential actions – those which would constitute speech in the language- describable, antecedently, only as so much patterned emission of noise. For that systematic imposing of descriptions to be acceptable, it would have to be the case that speaker’s performances of the actions thus ascribed to them were, for the most part, intelligible under those descriptions, in the light of propositional attitudes: their possession of which, in turn, would have to be intelligible in the light of their behaviour-including of course their linguistic behaviour- and their environment”. (J. McDowell, 1977)

In accordance with the considerations above, we adopt here the perspective according to which language interpretation and non-linguistic behaviour interpretation are interrelated enterprises. Seeing linguistic behaviour as a certain kind of behaviour might suggest that the project of language interpretation is a subproject of the wider enterprise of behaviour interpretation. Redescribing an utterance of a speaker, redescribing a mere phoneme a speaker utters, as a certain kind of speech act amounts to redescribing (linguistic) behaviour of an agent as intentional action. Understanding the language of speakers would then amount to identifying their linguistic behaviour as intentional action explained and understood in the light of propositional attitudes of speakers. Described like this the enterprise of language interpretation becomes a special case of interpretation of behaviour in general, a redescription of an event as a certain kind of intentional action. However, it has been suggested (Davidson, 1975) that the two projects are not independent of each other. Interpretation of language and behaviour requires attribution of a system of propositional attitudes, attribution of a system of beliefs and desires in the light of which the behaviour of the agent is seen as action and it is understood. A fine discrimination between propositional attitudes is required for identification of propositional attitudes that would enable identification of behaviour as action. A fine discrimination between propositional attitudes requires language¹. Following this line of thought leads us to view the project of language interpretation not merely as a subproject of the project of behaviour interpretation but as an interrelated component of the enterprise of interpretation of language and action. According to the present perspective a fine discrimination between

¹ This line of thought, Davidsonian in character, is not uncontroversial. See Peacocke (1986, chapter 8) for objections to the view that any account of thought requires language and also see Stalnaker (1984) for an

propositional attitudes requires interpretation of language. Understanding linguistic and non-linguistic behaviour of speakers/agents requires attribution of propositional attitudes.

In introducing interpretation theory we need to consider a certain constraint imposed on the theory in virtue of its aim. In a theory of interpretation no prior understanding of speech and other behaviour is presupposed. Understanding the language and behaviour of agents is the aim of interpretation theory. The theory would deal with its aim in a question-begging manner if it relied on a prior-to-theory understanding of speech or other behaviour. Aspects of speech and other behaviour of agents that require a theory of understanding in order to be identified are, then, not taken to be known in advance of a theory of interpretation. Semantic aspects of sentences speakers of the language use, meanings of words of the language they speak, social and cultural features of the community of which they are members that determine the agents' goals and purposes are not among the data of the theory. Identification of those features is part of the aim of the theory. Identifying the meaning of an expression of the language under interpretation, the proposition a sentence expresses, the propositional attitude attributed to the speaker uttering a sentence, the propositional attitudes that explain the utterance or the behaviour of speaker/agent, constitute conditions for understanding that have to be introduced within a theory of understanding language and behaviour. In a nutshell, the constraint is that that no prior-to-theory understanding is presupposed. A theory of interpretation on which this constraint is imposed is a radical theory of interpretation.

A radical theory of interpretation is required to result in certain redescriptions of utterances and behaviour of agents that would enable understanding the agents' behaviour. The theory is required to result in such redescriptions in a radical way. The propositional attitudes of speakers/agents in the light of which the actions of speakers as identified of their behaviour can be seen to be intelligible are not known in advance of the theory. However the attitude attributions to the agent/speaker are constrained in a certain way. The attitudes attributed to the agent/speaker must be such that the agent/speaker is rational. This constraint is constitutive of interpretation and we call it

the principle of rationalisation². The principle is constitutive of interpretation because seeing individuals as rational agents is required for understanding their behaviour.

We can consider here some further principles of a theory of radical interpretation. One principle at which we have already hinted is that the theory of language interpretation should deliver the required redescriptions in a systematic way. Interpretation of speech requires identification of a semantic structure of language. The utterances of speakers are interpreted by assigning meanings to their component parts and by assigning a structure in the way they compose the utterance. Call this principle the principle of generativity.

Another principle of methodological importance for a radical theory of interpretation is the triangle principle. This is the principle according to which the attitudes attributed to the speaker/agent should be the same either described in the speaker's language or in the interpreter's language. A speaker might use a sentence token "Es schneit." to make a statement. The speaker is expressing a belief by making this statement. We can consider the utterance "Es schneit." to be a first person belief attribution. On the other hand, interpreting the utterance "Es schneit." might result in its redescription as a statement that it is snowing. We can consider this case as one of third person belief attribution described in the interpreter's language. The principle of triangle establishing a connection between the speaker's description of his attitude and the interpreter's description of the speaker's attitude enables identification of first person and third person attributions. The triangle principle will be shown to have significant methodological importance in the present project. Utterances are related to the attitudes of the speaker/agent by a relation of attitude expressing (speaker's language) or attitude ascribing (interpreter's language). The triangle principle would enable identification of an agent's attitude by two kinds of description.

Another methodological principle the manifestation principle. This principle would constrain the interpretative enterprise by assuming a kind of sincerity condition to be fulfilled. A principle of truthfulness constrains the attitudes attributed to the speaker.

Another principle that might be taken to be constitutive of interpretation is the

² The jargon used here to name the constraints of interpretation theory is due to D. Lewis (1974).

principle of charity. This is the principle that the beliefs attributed to speaker/agent must be true according to the interpreter's standards. Discussion of the charity principle will be given in 1.2.1. The interrelation of the two principles, charity and rationality, will be shown to be a significant issue in the present context.

In a general discussion of a theory of language interpretation it is advisable to clarify some points. One point is that theories of language interpretation, like theories of translation, are not dealing with the problem of language acquisition. A theory of interpretation for a language does not imply the claim that in learning the mother tongue we embody any such theory. And it is independent of the claim that in understanding each other in everyday life we embody a particular theory that might be proposed as a theory of interpretation for the language. A theory of interpretation does not answer the question of what do we know in understanding language. The question to which it primarily attempts to offer an answer is what could we know that would enable us to understand a language. The primary claim of a proposal of a theory of interpretation is that a theory of interpretation of a language would suffice for an understanding of the language.

Interpretationist philosophers often do make the stronger claim that all understanding of language and mind involves interpretation. (Davidson, 1973, Dennett, 1981). I suggest we distinguish between the two claims. We should consider the claim that the aim of a theory of interpretation is a theory sufficient for understanding as the primary aim of interpretation theories. The stronger claim that all understanding involves interpretation is a more controversial claim. To accept it we need to show that what we actually know that enables us to understand language and mind has the theoretical representation described by an interpretation theory. Since a presentation of conclusive reasons for acceptance of this claim would be the topic of a different dissertation, let us just reserve our judgements on the claim here.

Another point to be clarified in discussing interpretation of language concerns its relation to language translation. The project of interpretation involves an enterprise of redescription. So, one might suggest that a translation manual would offer the required redescriptions. Similarly to the case of radical interpretation we can consider the case of a radical translation theory. The aim of this theory would be, by taking a radical perspective, to construct a translation manual, that is, a synonymy list between

expressions of two languages. Taking a radical perspective would constrain the theory in a way such that no relation of synonymy between expressions of the two languages would be taken to be known in advance of the theory³.

However, there are objections to the suggestion that a theory of radical interpretation of language should aim at a translation manual. Davidson (1973) has pointed out that the translation method would not, in all cases, yield a theory that would suffice for understanding an unknown language. A translation manual of a language constructed with the restrictions of a radical translation theory (Quine, 1960) is not adequate to serve the purposes of interpretation theory. It can be shown to be inadequate by considering the case in which the metalanguage, that is the language in which the theory of translation is stated, is unknown to the interpreter. In this case there is an additional requirement for a translation manual of this language to a language the interpreter understands. A translation method that would generally suffice for interpretation would, then, involve three languages. These would be the unknown language under interpretation, the language in which the translation is given and a language known to the interpreter. So, not any translation manual of the language under interpretation would suffice for understanding of the language.

An additional reason for which a translation manual would not suffice for interpretation is that it might not explain facts concerning the semantic structure of a language. Translation maps expressions of one language to expressions of another but it may not provide an account of how the meaning of a sentence depends on the meanings of its constituent parts. While for the case of an infinite language it might be difficult to imagine how a translation manual could be constructed in a way insensitive to semantic structure, in the case of a finite language constructing a translation manual does not need to be sensitive to semantic structure. In contrast, a theory that

³ Introducing a radical perspective in a theory of translation and in a theory of interpretation is significant for both theories. For Quine (1960), the main theorist of radical translation, considering the perspective of a radical translator provides us with an insight about the nature of the notion of meaning. Taking a radical perspective would clarify which semantic features of an expression or sentence we can identify on the basis of empirical evidence and which features of it we have grounds to question. In interpretation theory taking a radical perspective would shed light on the notion of understanding a language. But the two theories were introduced by Quine and Davidson respectively in order to illustrate different accounts of the nature of linguistic meaning. While Quine uses the case of radical translation in order to undermine the notion of meaning, Davidson uses the case of radical interpretation in order to illustrate the interdependence of meaning, belief, desire and the holistic nature of understanding language and action of speakers/agents.

explains semantic structure of a finite or infinite language satisfies a requirement for the possibility of understanding sentences of the language. For the case of a finite language the theory that respects semantic structure is apt to account for the understanding of unheard sentences if the axioms that govern their constituent parts are known (Evans, 1981).

The above suggest that a translation manual would not suffice for the purposes of interpretation theory.

1.2 Davidson's theory of radical interpretation

In this section we introduce the main features of Davidson's theory of radical interpretation. It has already been said that a theory that in systematic way results in redescriptions of utterances as speech acts is a candidate theory of interpretation of language. A candidate theory of interpretation of German language will result in a redescription of the utterance of "Es regnet.", made on a particular occasion, as a saying that it is raining. Davidson makes more specific suggestions about a theory of interpretation of a language. The first suggestion is that a theory that would give necessary and sufficient conditions for the truth of every sentence of the language under interpretation could serve as a theory of semantic interpretation for the language.

"To know the semantic concept of truth for a language is to know what it is for a sentence –any sentence- to be true, and this amounts, in one good sense we can give to the phrase, to understanding the language." (1967, p. 24)

The suggestion is that a theory that structurally derives an infinite set of biconditionals of a certain sort could serve as a theory of meaning for the language. Having a theory that for every sentence *s* of the language under interpretation yields as theorem a T-sentence: '*s is true if and only if p*', where *p* is a sentence of the interpreter's language would suffice as a theory of understanding the language under interpretation. That theory would suggest that uttering *s* be redescribed as saying that *p*. The way the theorems are derived reveals the logical structure of the language.

Davidson's second suggestion is that a theory having the form of Tarski's (1956) theory of truth for a language, appropriately modified, would satisfy the condition stated above. (Davidson, 1973, 1974, 1976). It would yield for every

sentence of the language under interpretation a T-sentence. A Tarski-style theory of truth would then be a theory having the required form so that it would suffice as a theory of understanding a language under interpretation. The claim is a bold one and not without complications.

Tarski's theory of truth is given for a formal language. The theory does not derive equivalencies in truth-value between sentences. It fixes the extension of the predicate 'truth-for-L', where L is the formal language under consideration. In Tarski's theory the extension of the predicate 'truth-for-L' is fixed in a structural way and on the basis of finite resources. So, it seems that Davidson's proposal is neither straightforward nor uncontroversial. There are two kinds of complexities concerning Davidson's suggestion.

Of the first kind are problems such as structural complexities (i.e. the problem of whether Tarski's recursive structure can be identified of the surface or deep structure of a sentence.) There are complexities due to the phenomenon of ambiguity and due to the phenomenon of referential opacity. These are some of the problems that make it hard, if not impossible, to contrive structural recursive rules for deriving a T-sentence for every sentence of the language under interpretation. The universality of natural languages that allows the appearance of semantical paradoxes was for Tarski a reason for rejecting the possibility of devising a theory that would fix the extension of a truth-predicate for a natural language. Davidson decided to be optimistic about the complications that arise by using a modification of Tarski's theory of truth for a natural language. (Davidson, 1967, 1968, 1973b). In the modified version of Tarski's theory truth is a property not of sentences of a language but of utterances, that is sentences of the language in a context. This modification suggests a way of dealing with sentences containing demonstratives. It also paves the way to a paratactic treatment of referentially opaque contexts.

The second kind of complexities arises from the modifications that have to be made to the conceptual mechanism employed by Tarski's theory of truth. Tarski's formulation of convention T presupposes the relation of synonymy in the guise of a requirement present in convention T. In the T-schemes, '*s is true if and only if p*', derived by the theory, p should be a translation of s in the metalanguage. Since Davidson's aim is a semantic theory of a language the relation of synonymy is a

relation that the theory should explicate and not one it presupposes. Using Davidson's own words the proposal is

"...to reverse the direction of explanation: assuming translation. Tarski was able to define truth; the present idea is to take truth as basic and to extract an account of translation or interpretation" (1974, p. 134).

The modification of convention T he proposes is the following:

"...An acceptable theory of truth must entail, for every sentence *s* of the object language, a sentence of the form: *s* is true if and only if *p*, where '*p*' is replaced by any sentence that is true if and only if *s* is" (1973, p. 134).

The correctness, then, of the T-schemes derived by the theory is not a matter of synonymy, but a matter of equivalence in truth-value. A T-scheme derived by the theory is correct if the object language sentence and the metalanguage sentence are equivalent in truth-value. The adequacy of the whole theory is decided on the basis of the totality of the T-sentences; an acceptable interpretation theory yields for every sentence of the language under interpretation a correct T-scheme.

The suggestion, as it has been presented up to now, is that a theory that has the form of a Tarski-style definition of truth for a language *L* can serve as a theory of interpretation for *L*. For every sentence *s* of the language under interpretation the theory gives as a theorem a T-sentence: '*s* is true if and only if *p*'. A T-sentence resulting as a theorem of a theory of interpretation suggests that an utterance of a sentence *s* by a speaker of language *L* be redescribed as a 'saying' that *p*.

It is not obvious that a theory of truth as envisaged by Davidson would suffice for interpretation. Translation and interpretation seem to be far richer concepts than truth. Understanding an utterance seems to involve more than or to require different conditions to knowledge of the truth conditions of the utterance. Davidson suggests that holistic considerations imposed on the adequacy criterion of the theory would to a certain degree restrict the acceptable T-schemes. Holistic considerations would prevent the acceptability of theories that yield truth-schemes like ' 'Snow is white' is true if and only if grass is green', or other undesirable equivalences in truth-value. (1973a, p.139, 1976, pp.171-179). The holistic considerations that aspire to narrow down the acceptable T-sentences are based on the suggestion that the totality of the acceptable T-sentences should fit evidence about the speaker's speech behaviour. In the subsections 1.2.1, 1.2.2 this suggestion is further explained and elaborated.

1.2.1 Davidson's early theory of interpretation

What has been stated up to this point concerns the aim of a theory of interpretation. The suggestion presented in the previous section is that a theory that for every sentence of the language under interpretation derives a correct T-scheme would suffice as a theory of interpretation of the language. Here we are going to discuss some issues concerning the adequacy of the suggestion. Let us assume that we agree on the form of the theory we want; we agree that it has the form of Tarski's recursive definition of truth-for-L. Such a theory would satisfy the generativity constraint. The question that arises here is how a theory having the form of Tarski's recursive definition of truth for L can be employed for making sense of the linguistic and other behaviour of agents/speakers. In other words, the question is what are the conditions for a theory having the suggested form to be adequate as a theory of interpretation. Davidson's answer is the following. The T-sentences the theory generates should fit evidence about speaker's behaviour.

Observations of the linguistic and non-linguistic behaviour of agents/speakers of a language in observable environmental situations constitute data relevant to the verification of the theory. But accommodating of a radical perspective in a theory of interpretation deprives access to beliefs and other propositional attitudes of the agents/speakers independently and in advance of a theory of interpretation.

"The evidence cannot consist in detailed descriptions of the speaker's beliefs and intentions, since attributions of attitudes, at least where subtlety is required, demand a theory that must rest on much the same evidence as interpretation." (1973, p. 134)

In Davidson's early theory of interpretation we find the suggestion that the empirical primitive on the basis of which the adequacy of the theory is judged is the attitude of *holding true* a sentence, relativised to some contextual parameters¹. Ascription of the *holding true* attitude toward a sentence in observable environmental circumstances provides the basis for theory verification.

The suggestion is along the following lines. Since we have a theory of

¹ Without intending to disregard other distinctions in the development of Davidson's thought, we take the mark for a distinction between Davidson's early and late theory of interpretation to be the empirical primitive suggested in each case. In Davidson's early writings this is the attitude of *holding true* a sentence. In Davidson's later writings this is the attitude of preference. The issue concerning the empirical primitive is of our own interests here.

interpretation we can test it in the following way: say that the theory results in the T-scheme:

(T) '*Es regnet*' is true in German when spoken by *x* at time *t* if and only if it is raining near *x* at *t*'.

Evidence for the correctness of the above T-scheme could be given by observations along the following line:

(O) *Kurt belongs to the German speech community and Kurt holds true 'Es regnet' on Saturday at noon and it is raining near Kurt on Saturday at noon.*

The suggestion is that observations like (O) of members of the German speech community are evidence for the truth of (T). The suggestion conjures up an assumption known as the principle of charity. The assumption is that most of the time sentences are true when *held true*⁵. The attitude of *holding true* a sentence is the result of belief of the speaker and of the meaning of the sentence. A version of the principle is that most of the time speakers should be viewed as being correct in their beliefs. A methodological version of the principle is that interpretation should be conducted so that agreement in the beliefs of the interpreter and speaker is maximised⁶. The principle introduces a holistic character on the adequacy condition. The T-sentences are judged to be adequate as a system. On the system of the totality of the T-sentences generated by the theory agreement on truth between speaker and interpreter should be maximised.

The particular aspect of the principle of charity that we discuss here concerns the central role the *holding true* attitude has in the principle. The attitude of *holding true* a sentence here appears to be the key attitude to interpretation. It is the attitude on which the principle of charity applies and in this way suggested T-sentences judged to be adequate.

⁵ In the literature there is a variety of formulations of principles labelled as principles of charity. In Quine (1960) we find it as the principle of maximising agreement between the translatee and ourselves. In Grandy (1973) we find the principle of humanity. Namely, it is the principle according to which the pattern of relations between beliefs, desires and the world attributed be as similar to our own as possible. Here I restrict myself to a conception of the principle as it appears and is used in the particular interpretation theories I consider.

⁶ There are many questions concerning the principle of charity that we leave out of discussion here. For instance, there is the problem that the proportion of truth to error is indeterminate if we consider all possible utterances. There is the problem of the grounds for the possibility of massive error to be a priori excluded. Here

The attitude of holding a sentence true

In the remaining part of this subsection I discuss the place of the *holding true* attitude in Davidson's theory of radical interpretation as it has been presented up to now. The importance of the possibility of identifying the *holding true* attitude in the context of this dissertation is that it relates in a certain way to the possibility of identification of truth-evaluable utterances and of doxastic attitudes. Though the truth-evaluable utterances and the utterance of sentences *held true* do not coincide, the two classes are related. A speaker who, in uttering a sentence, *holds true* the sentence makes a truth-evaluable utterance. The attitude of wanting a sentence to be true ascribed to a speaker in uttering the sentence "Close the door!" deprives the utterance of truth-evaluability⁷. On the basis of the attitude a speaker has towards a sentence in uttering the sentence it might be possible to distinguish between the truth-evaluable and the non-truth-evaluable utterances. And this is the line that we will finally suggest here. But first we have to look more closely to the place of the attitude of *holding true* and the possibility of its identification.

The attitude of *holding true* a sentence is suggested to be the empirical primitive in Davidson's early theory of interpretation. One thing we have mentioned about the attitude of *holding true* a sentence is that it provides the bridge for meaning and belief. The *holding true* attitude is the product of meaning and belief. A speaker *holds true* a sentence on a particular occasion because of what he believes and because of what the sentence means. A speaker, who *holds true* the sentence "Es regnet." on a certain occasion, does so because of what he believes about the weather and because of what the German sentence "Es regnet." means.

The other thing to be mentioned about the attitude of *holding true* a sentence is that it is central to the principle of charity. The principle is suggested to be an adequacy condition on a theory that satisfies the generativity constraint. A theory that

we focus on problems related to the *holding true* attitude as it appears in the principle.

⁷In 'Thought and Talk' Davidson (1975) comments on the attitude of *holding true* in relation to commands: "If he (the speaker) utters a command, we may usually take this as showing that he *holds* a certain sentence (closely related to the sentence uttered) to be *false*...". (italics mine). The issue deserves extensive treatment, but suffice it to mention a case of issuing a command in which the ascription of the attitude of *holding true* or

has the form of Tarski's definition of truth is judged to be adequate as a theory of interpretation on the basis of the charity principle. The principle of charity might be given a methodological status. That is, that it is a principle about how to go about in the construction and verification of the theory. However in Davidson's writings the principle appears to be one constitutive of interpretation. Using his own words the case is the following.

"...If we cannot find a way to interpret the utterances and other behaviour of a creature as revealing a set of beliefs largely consistent and correct by our own standards, we have no reason to count that creature as rational, as having beliefs, or as saying anything." (Davidson, 1973).

The point here is that massive error or inconsistency in beliefs threatens the possibility of understanding. Unless we can see the speaker to be correct and consistent in most of his beliefs it is not possible to make sense of his speech and other behaviour. The charity principle, then, becomes a close relative of the principle of rationalisation. Taking most of a speaker's beliefs to be true is suggested to be a condition for seeing the speaker to be rational. The relation between the two principles will be discussed further in the following parts of the dissertation.

Identifying the attitude of holding a sentence true

Within the perspective presented above, the claim we intend to make here is that identification of the attitude of *holding true* a sentence is a substantial problem for Davidson's theory of interpretation. Let us notice that the problem of the identification of the *holding true* attitude is a certain version of the problem of a classification of truth-evaluable utterances. In Davidson's theory, the latter problem is projected on the problem of the identification of *the holding true* attitude. In this context, the classification of truth-evaluable utterances is a requirement for determining the domain of applicability of the principle of charity.

Let us then see the reasons for the claim we make here. That is, the claim that the possibility of identification of the *holding true* attitude is a substantial problem of Davidson's theory of interpretation. The principle of charity applies to the domain of sentences *held true* on certain occasions. A common sense assumption we have taken

false is inappropriate. In issuing the command "In five minutes close the door!", the *holding true* or *holding false* attitudes towards the sentence "The door is closed in 5 minutes." would not be successful ascriptions.

in this dissertation is that the attitude of *holding true* a sentence is not the only kind of attitude speakers have toward sentences. A theory of radical interpretation that grounds its adequacy on the principle of charity needs to provide the resources for a distinction between those utterances in the making of which sentences are *held true* and utterances not associated with the attitude of *holding true* a sentence.

Identification of the *holding true* attitude or identification of an attitude as a doxastic attitude is required for application of the principle of charity on the suggested use of the observational scheme:

'Agent A holds true sentence s at time t and place p, and x at t and p'

providing evidence for the correctness of the T-sentence

's is true if and only if x'.

In other words the point we raise here is that identification of the *holding true* attitude is required for a determination of those utterances of speakers on which the suggested evidential scheme constituted by the principle of charity is applicable. The suggested evidential scheme tests the correctness of T-sentences. The theory is adequate if, for every truth-evaluable sentence of the language, it results in a T sentence shown to be correct by the adequacy condition constituted by the principle of charity. Implicit in the adequacy claim is the assumption that interpretation of the truth-evaluable utterances of a language suffices for interpretation of the whole language. The assumption is not uncontroversial but it is beside the point discussed here to challenge it. A discussion of the acceptability of the assumption will be given in later parts of the dissertation⁸.

Within the framework of the theory as it has been elaborated up to now we can say the following concerning the significance of an identification of the *holding true* attitude. One problem a theory without resources for the identification of the *holding true* attitude toward sentences encounters is the following. Given that there are other attitudes speakers have toward sentences, we would not have any grounds on which we could decide that the attitude the speaker has in a situation of an utterance is that of *holding true* the sentence he utters or that of believing the

⁸ In the second part of chapter 3 are presented difficulties with accommodating non-truth-evaluable utterances to truth-conditional semantics.

sentence rather than, say, desiring it. If the attitude the speaker actually has in uttering a sentence is, say that of a desire, use of an observational scheme like the one suggested for the case of the *holding true* attitude would be inappropriate. The utterance the speaker makes, in this case, would be misinterpreted if use of the above mentioned scheme is made. A theory that does not provide resources for a distinction between the attitude of *holding true* a sentence and attitudes related to utterances not susceptible of truth-evaluation would result in a systematic misapplication of the principle of charity. A theory like this would systematically attribute truth to utterances not susceptible of truth-evaluation or would systematically apply an incorrect evidential scheme for checking the correctness of T-sentences. A systematic misapplication of the principle of charity, resulting from lack of resources of the theory for identification of the *holding true* attitude, would threaten the possibility of correct interpretation.

In addition, a theory that lacks resources for determining the domain of applicability of the charity principle, the domain of the *held true* sentences, would result in attributions of a truth-value invariably to all utterances. This would be the case in virtue of the constitutive character of the charity principle. If we insist on the assumption that not all utterances are susceptible of truth-evaluation, the domain of application of the principle needs to be identified. We must clarify that the claim is grounded on the feature of the charity principle that it applies to utterances. Even if a truth-evaluable core can be identified in all utterances, truth-evaluable and non-truth-evaluable ones, we need to separate the truth-evaluable utterances from the non-truth-evaluable ones for application of the charity principle.

Let us clarify further this point. The problem presented in the paragraphs above does not concern only a less than ideal empirical approximation of the attitudes ascribed to speakers in the process of applying an interpretation theory for a certain language community. The problem we might encounter in the process of interpretation is that we need distinctions between the *holding true* and other attitudes that we lack at least at the early stages. However, these distinctions could be made available in further stages of the interpretation process by considering holistic relations between the attributions. The issue we raise here concerns the kind of holistic considerations, which would result in the required distinction between the

holding true and other attitudes. In Davidson's early theory of interpretation the adequacy condition that determines the kind of holistic considerations we could invoke in interpretation is the principle of charity. Now, charity suggests holistic consideration on the holding true attitude or on the beliefs. As such, it would not suffice to introduce the required distinction between the holding true and other attitudes or the distinction between the truth-evaluable utterances and non-truth-evaluable ones.

We said that a theory that lacks resources for the required distinction would result in a systematic misapplication of the charity principle. This is a substantial problem for a theory of interpretation for two reasons. One the one hand, there is a problem based on the fact that there are no grounds on which we could assume that the number of cases of *holding true* a sentence in making an utterance outweighs the number of cases of having a different attitude towards a sentence in making an utterance. If so, the resulting incorrect T-schemes might not diminish in the process of interpretation. One the other hand there is a more important problem than this one. A theory that does not provide the resources for the distinction in question leaves us blind as to how the systematic misapplications of the principle of charity can be amended by holistic considerations. In this case the systematicity of error resulting by the lack of resources of the theory for identifying the domain of utterances to which the principle of charity is applicable cannot be counterbalanced by holistic considerations. The reason for this is the following. Without expanding the resources of the theory concerning the possibility of a distinction between the *holding true* attitude and other attitudes the kind of holistic considerations we can make will be based again on the principle of charity. Without expanding the resources of the theory, the holistic considerations that could be invoked in order to constrain the acceptable T-sentences will be again of the kind of observations like (O) which, by use of the principle of charity, would be taken to lend support to the correctness of certain T-sentences⁹. The upshot of the discussion here is that holistic considerations based on the principle of charity solely do not seem to be sufficient. There does not

⁹ In later parts of the dissertation it will be suggested that the resources of the theory can be expanded in the required way by an elaboration of the other constitutive principle of interpretation, the principle of rationality. It will be suggested that on the basis of a certain way of construing the rationality principle the domain of applicability of the charity principle can be determined.

seem to be a straightforward answer to the question about *how* from this sort of holistic considerations we can draw the distinction.

In Davidson's early discussions of theory of interpretation the possibility of an identification of the *holding true* attitude towards sentences belongs to the presuppositions of the theory.

"It is an attitude an interpreter may plausibly be taken to be able to identify before he can interpret, since he may know that a person intends to express a truth in uttering a sentence without having any idea what truth" (1973, p. 135).

Ramberg (1989), commenting on the role of the holding true attitude in Davidson's theory of interpretation says:

"...we have to assume that we are observing creatures who assert, and, crucially, that we are reasonably adept at telling when those observed are engaging in this particular linguistic activity—even if we have no clue as to what is being asserted...If we cannot identify the utterances of L over which the truth-predicate of L ranges, there is no hope of our capturing its extension in T-sentences. And in this case, there would be no chance of bringing out the semantic structure via a Tarskian theory" (1989, p. 68).

The *holding true* attitude is one among many kinds of attitude toward sentences an interpreter might ascribe to a speaker. And the sentences over which the truth predicate varies constitute only a subclass of the sentences of a language the theory is required to interpret. Taking into consideration Davidson's suggestion that a identification of propositional attitudes requires a semantic interpretation of language, it follows that the possibility of identifying the attitude of *holding true* a sentence cannot be considered to be an innocent methodological presupposition of the theory. Identification of the attitude of *holding true* a sentence requires attributions of propositional attitudes to speakers/agents. Propositional attitudes cannot be identified prior to a theory of radical interpretation. Making the assumption that identification of the *holding true* attitude can be made prior to the theory, that is, that it can be considered to be a presupposition of the theory would have the effect that our theory of interpretation would turn out to be less radical.

In other contexts Davidson acknowledges the weight of the assumption of the possibility of identification of the *holding true* attitude.

"...Of course it should not be thought that a theory of interpretation will stand alone, for as we noticed, there is no chance of telling when a sentence is held true without being able to attribute desires and being able to describe actions as having complex intentions. This observation does not deprive the theory of interpretation of interest, but assigns it a place within a more comprehensive theory of action and thought." (Davidson, 1975).

The above considerations show the significance of the possibility of identifying the *holding true* attitude in a theory of interpretation. Davidson seems to acknowledge the significance of the other attitudes for an identification of the attitude of *holding true* a sentence. But, still, we need to show how the attitude of *holding true* a sentence is identified on the basis of other attitudes. We need to show what kind of holistic considerations would result in the identification in question. The principle of charity as it applies narrowly to the *held true* attitude and to belief does not provide access to different attitudes.

The sufficiency of one attitude worry

At this point we might consider some issues concerning the choice of one key concept as the primary assessment of utterances. There is an extensive treatment by D. Wiggins (1980, pp.189-221) of the issue. Wiggins considers a property Φ of sentences such that for any sentence s ,

" { s means that p in L (or s has a content in L appropriate for the assertion that p)}

if and only if

{some θ canonically implies that [(s has Φ) iff p], where θ is a theory of Φ , and θ combines with an anthropology α to make better total sense of the shared life and conduct of L -speakers than any rival pair, $\langle \theta', \alpha' \rangle$ consisting of rival theory θ' plus rival anthropology α' }. (1980, p. 204).

Wiggins' purpose is to point to the marks of property Φ which would satisfy the above condition. What is of particular interest in his strategy is that he does not assume the property Φ to be the property of being true or the property of being *held true*. The marks of this property would be characterised by anthropological constraints. If it turns out that the property of sentences characterised by the above condition is interchangeable with a pre-theoretic notion of truth then we are justified in assuming truth to be the key concept in a theory of radical interpretation.

One mark of the property Φ that he points to is that Φ , or assertibility as he calls it, is the primary dimension of assessment for sentences, or that property which sentences have normally to be construed as aiming to enjoy.

A question that arises here is whether one single property would be sufficient for interpretation. The question here concerns the grounds for a doctrine called *semantic monism* (Platts, 1980, pp.1-19). How justified are we in claiming that there

is one single property that can be attributed in a uniform way to sentences, which they have to be construed as aiming to enjoy, or considering it as the primary dimension of assessment for all sentences to be interpreted? The question of the possibility of identifying such a property is discussed in chapter 2., chapter 3, 3.4, 3.5, 3.6. The upshot of the discussion there will be that identifying a single property of utterances uniformly had by all kinds of utterance is not a straightforward enterprise. Here let us quote another passage from Wiggins' paper, in which he states the expectations we can have of a theory combined with an anthropology for an interpretation of a language.

"The anthropology that adopts the theory θ along with the best theories of context and force that it can find, and then deploys all of these *in concert* for purposes of the interpretation of utterances that are appropriate to be considered as expressive of beliefs, concerns, needs, wants, or whatever, (i) describes and explains, in the light of what can impinge on L-speakers or is presented to their experience, the doxastic and the affective reactions of L-speakers to their environment and explains such reactions better than an anthropology with different theories of sense, force and context and (ii) makes intelligible, in the light of the doxastic and the affective attitudes attributed to L-speakers by the total theory, the motivation and conduct of L-speakers - more intelligible than would an anthropology with different theories of sense, force and context. Finally, explanations are judged here, not so much by the successful predictions they generate (if these are plentiful and uniformly accurate, that is a bonus), but by the extent to which they enable us to grasp, or identify imaginatively with a norm of rationality to which the thoughts and actions of L-speakers are seen as answerable" (1980, pp. 199-200).

The question I raise concerning the uniqueness of the property in question is the following. What guarantees that the anthropological constraints that would characterise the marks of a property Φ that for an utterance u our best theory would result in

[u means that p] if and only if

[u has Φ if and only if p],

will suggest a unique property Φ attributable to all utterances in a uniform way? I raise the question here but it cannot be answered without a close consideration of the anthropological constraints relevant to a theory of understanding. The particular anthropological constraint that will be considered in this dissertation is the principle of rationality¹⁰.

Does it have to be the holding true attitude?

I finish this section with a few more comments related to the place of the

¹⁰ A detailed discussion of the rationality constraint will have to wait up to chapter 6 of the dissertation.

holding true attitude in a theory of radical interpretation. My comments concern the choice of the *holding true* attitude as the empirical primitive of the theory.

As it has already been mentioned Davidson suggests that the choice of the *holding true* attitude as the central attitude toward sentences provide the bridge between meaning and belief. A speaker *holds true* a sentence (on a particular occasion) because of two factors: what the sentence means and what the speaker believes. But these two factors are not independent of each other. A fine discrimination of beliefs requires a theory of interpretation and a semantic interpretation of sentences requires attributions of beliefs.

“A speaker who holds a sentence to be true on a particular occasion does so in part because of what he means, or would mean by an utterance of that sentence, and in part of what he believes. If all we have to go on is the act of honest utterance, we cannot infer the belief without knowing the meaning, and have no chance of inferring the meaning without the belief” (1974, p. 142).

The attitude of *holding true* a sentence is the result of two interrelated but unknown factors; the meaning of the sentence and the belief of the speaker. If we knew either of the two factors entering in uttering a sentence, meaning or belief, it would be possible on the assumption that the speaker *holds true* the sentence on the particular occasion to infer the other. That is, if we knew that the speaker believes to be the case that it is raining in uttering the uninterpreted *held true* by him sentence “It is raining.”, we could infer that “It is raining.” means it is raining. And if we knew the meaning of the sentence “It is raining.”, and that the sentence is *held true* on a particular occasion we could infer that the individual believes that it is raining.

These considerations, though plausible enough, do not suffice to justify the choice of the *holding true* attitude as the empirical primitive of the theory. They show how the *holding true* attitude depends on meaning and belief and how it is possible by a simultaneous use of the principle of charity (that speakers have true beliefs by the interpreter’s standards or that speaker and interpreter share their beliefs) to infer meanings. The *holding true* attitude might prove itself not to be indispensable as the empirical primitive of the theory by considering how another attitude can play the role of a bridge to meaning. We can put forward parallel considerations to the interrelation of belief and meaning in issuing a *held true* sentence to the interrelation of desire and meaning in issuing a sentence that the speaker *holds good*. The attitude of *holding good* a sentence or wanting a sentence to be true can play an equally central role in a

theory of radical interpretation. The explication of its role would be as follows: an individual wants a sentence to be true or *holds good* a sentence because of two interrelated factors: the meaning of the sentence and the desire of the individual. If we knew what the sentence means, we could, on the assumption that the individual *holds good* the sentence or wants the sentence to be true to infer his desire. And if we knew the desire of the agent, we could, on the assumption again that he *holds good* the sentence or he wants the sentence to be true understand what the sentence means¹¹.

Adequacy conditions of a theory on the basis of the attitude of *holding good* a sentence suggest a modified principle of charity. The modified principle of charity will have to be that speaker and interpreter share their desires or that the sentence a speaker *holds good* is good by the interpreter's standards.

The above considerations are not meant to be anything more than a schematic presentation of an alternative proposal. The point of presenting these considerations here is not to undermine the importance of the *holding true* attitude as the bridge to meaning and belief. It is not the replacement of the *holding true* attitude by the *holding good* attitude that is suggested here. The suggestion here, which will be elaborated in the development of the discussion in the dissertation, allows for the view that both attitudes have an important role to play in a theory of interpretation and that the one cannot replace or be reduced to the other. In the context of this framework it is the possibility of identifying each attitude and distinguishing between the two that becomes a significant issue. The importance of both attitudes for a theory of interpretation is acknowledged in Davidson's later theory of interpretation where the empirical primitive of the theory is suggested to be the attitude of preference between two sentences. Davidson's late theory of interpretation is presented in 1.2.2.

I finish my discussion on the issue of placing the possibility of an identification of the truth-evaluable utterances and of the doxastic states of a speaker/agent in a theory of interpretation like Davidson's with the following comments. The general issue I raise is that a distinction between the doxastic and the affective attitudes of a

¹¹ I should comment here that the suggestion that the attitude of wanting a sentence to be true as an alternative to the attitude of *holding true* a sentence presupposes an assumption that I will question later. The assumption is that sentences related to desires or wants are susceptible of truth evaluation. The suggestion that the alternative attitude be the attitude of *holding good* a sentence accords better with the direction of this dissertation.

speaker/agent is required by a theory of interpretation for the identification of the *holding true* attitude or the identification of the *holding good* (wanting to be true) attitude. Identification of the *holding true* attitude or the *holding good* attitude towards sentences is required for application of the principle of charity. Without a way of identifying at least one of the two a systematic misapplication of the principle of charity will result.

1.2.2 Davidson's later theory of interpretation

In this section I introduce some aspects of Davidson's late theory of interpretation. In particular, I discuss the significance of replacing the *holding true* attitude as the empirical primitive of Davidson's early theory by the attitude of preferring the truth of one sentence to the truth of another. The purpose of presenting these considerations here is mainly to point to the problems that the suggested replacement intends to answer. In the development of the dissertation it will be argued that alternative solutions to these problems are not sufficient. An elaboration of the considerations put forward in this section will be given in chapters 6 and 7 of this dissertation.

The replacement of the attitude of *holding true* a sentence by the attitude of preferring the truth of a sentence to the truth of another is suggested in order to cope with two sorts of problem of interpretation theory. The first problem is the possibility of interpretation of utterances that contain theoretical terms and are less closely connected to observation than utterances of observational sentences. The second problem is the possibility of identifying the doxastic and the affective attitudes of speakers towards sentences.

The second problem for which a replacement of the primitive attitude of *holding true* attempts to offer a solution is related to some issues already discussed in the previous section. It has been pointed out that identification of the attitude of *holding true* a sentence, or identification of the attitude of believing a sentence true is a substantial problem in interpretation theory. Identification of this attitude is required

for identification of the domain of application of the principle of charity, that is the principle according to which the utterances/beliefs of a speaker are taken to be true according to the interpreter's standards. An additional reason for which identification of the attitude of a speaker towards a sentence is important is that identification of the attitude of a speaker towards the sentence he utters is necessary for interpretation of the utterance. Understanding an utterance requires identification of the attitude of the speaker¹².

"...Or, more generally and more precisely, an interpreter must be able to tell, often enough, when a speaker holds a sentence he speaks to be true or false, or wants it to be true, or intends to make it true. This is, however, a fact about the nature of radical interpretation, and places no constraints on the illocutionary or ulterior intentions of a speaker." (1982c, p. 6)

The point that identification of attitude toward a sentence is required for understanding the utterance of the sentence can be reinforced by the following considerations. To understand speech and other behaviour of agents is the principle we need to see them as rational agents. Identification of the attitude of the speaker/agent towards a sentence, that is identification of the attitude related to an utterance (by being expressed in the utterance or by being ascribed by the redescription of the utterance) is required for the rational embedding of the utterance. Let us assume that the negation operator has been identified for the language under interpretation. If we insist that one kind of attitude is sufficient for showing the utterance to be rationally embeddable, (e.g. the *holding true* attitude as related to the charity principle) then redescrptions of two utterances as sayings *s* and not *s* will violate the charity principle. Redescrptions of two utterances as *holding true s* and *holding true not s* will attribute inconsistent beliefs to the speaker. Under the constraint that one kind of attitude would be sufficient for a rational embedding of ascriptions, the ascriptions of *s* and not *s* do not have a rational accommodation by interpretation theory. On the other hand it seems straightforward that attributions of *s* and not *s* might have a rational embedding if they are related to different attitudes. For instance, redescrptions of two utterances as an affective saying *s* and a doxastic saying not *s* i.e. as wanting *s* and believing not *s* (the speaker wants to have his dinner soon but he believes that he will not have his dinner soon) would not show the speaker to be irrational. Identification of the kind of attitude toward a sentence is

¹² A detailed discussion about identification of the aspects of an utterance that are involved in interpretation will be given in chapter 2.

required for an adequate rational embedding of the utterance.

It is with respect to this point also that the attitude of preference taken as the primitive attitude of the theory is more promising for the possibility of identification of each kind of attitude. The attitude of preference held between two sentences embodies belief and desire. It might be the case that by construing an adequacy condition of the theory on the basis of the notion of preference the required distinction between doxastic/affective attributions will result. Taking an adequacy condition on the attitude of preference and analysing the components of a notion of preference it might be possible to detect the beliefs and desires of the agent¹³. The suggested replacement of the primitive attitude promises a solution to the problem of identification of the attitude of the speaker (as a doxastic or an affective/evaluative/conative/directive attitude)¹⁴. Having a solution to this problem would enable identification of the domain of applicability of the principle of charity (that sentences *held true* are true by the interpreter's standards) and it would enable identification of the attitude of the speaker towards a sentence required for understanding the utterance of the sentence. The suggestion is that we expand the adequacy conditions of the theory so that roughly speaking a) certain constraints on the range of preferences of speakers are imposed b) by a) the doxastic and affective attitudes are detected and c) charity constraints apply to the doxastic attitudes.

The first problem of interpretation theory for which a replacement of the attitude of *holding true* a sentence by the attitude of preferring the truth of a sentence to the truth of another offers a solution concerns the interpretation of sentences that

¹³ A detailed account of the relation between preference, belief and desire will be given in chapter 6.

¹⁴ In the version of radical interpretation problem as considered by D. Lewis (1974) the possibility of a distinction between the affective and the doxastic attitudes of a speaker/agent constitute a substantial problem of the theory. Davidson's and Lewis's theories deal with a different problem because they differ in the kind of data of the theory. In Davidson's theory the data are utterances and behaviour of speakers in observable, publicly available, environmental circumstances. In Lewis' theory the data consist of the totality of physical facts about an agent which includes by far more than what is publicly available. But both Lewis' theory and Davidson's later theory have similar aims. The desiderata of Lewis' theory consist of the agent's attitudes, his beliefs and his desires, described both in the agent's and the interpreter's (our) language and the meanings of the sentences he utters (which might be specified by giving truth-conditions for the sentences). In this dissertation it is the Davidsonian framework that is adopted. That is, the project of interpretation is envisaged as an attempt to account for the possibility of understanding language and mind in everyday life situations. The justification for making this assumption is, as Davidson puts it, the following. "The requirement that the evidence be publicly accessible is not due to an atavistic yearning for behaviouristic or verificationist foundations, but to the fact that what is to be explained is a social phenomenon." (1990, p. 314). Lewis' problem of accounting of the possibility of interpretation on the basis of the totality of physical

contain theoretical terms. The interpretation of those sentences depends by large on the relations of evidential support between sentences. Relations of evidential support between sentences can be accounted by probabilistic relations. Evidential relations between sentences envisaged as probability relations are accounted by *Bayes's theorem*. According to *Bayes's theorem*, evidence *e* is relevant to the truth of a sentence, a hypothesis *h*, according to the following rule:

$$\text{pr}(h \text{ given } e)/\text{pr}(h)=\text{pr}(e \text{ given } h)/\text{pr}(e)$$

Relations of evidential support between the sentences require determination of a subjective probability function for agents. Let us take degree of probability of a sentence in the present context to be degree of belief in the sentence. The relation of evidential support between sentences is related to a network of beliefs and partial beliefs of an individual.

“The interpretation of terms less directly keyed to untutored observation must also depend in large measure on conditional probabilities, which show what the agent counts as evidence for the application of his more theoretical predicates. If we want to identify and so interpret the role of a theoretical concept or its linguistic expression, we must know how it relates to other concepts and words. These relations are in general holistic and probabilistic. We can therefore spot them only if we can detect the degree to which an agent holds a sentence true, his subjective probabilities. Simple assent and dissent are at extreme and opposite ends of a scale; we need to be able to locate attitudes intermediate in strength. Degree of belief, however, cannot be directly diagnosed by an interpreter; as we saw in discussing decision theory, degree of belief is a construction based on more elementary attitudes.” (Davidson, 1990)

It is a good coincidence that the attitude of preference between two sentences, which by embodying belief and desire is promising an account for detecting the doxastic and affective attitudes of a speaker/agent, can be envisaged as a decision theoretic notion. On the basis of a decision theoretic notion of preference between two sentences a subjective probability function for the agent can be determined¹⁵. So, the replacement of the *holding true* attitude by the attitude of preferring one sentence to another promises to cover the inadequacy of Davidson's early theory of interpretation that concerns the possibility of interpretation of utterances that express partial beliefs of the speaker and of sentences containing theoretical terms. The suggestion is that on the basis of the preferences of an agent between sentences and by use of decision theoretic considerations it would be possible to determine a subjective probability function for the agent that would represent the strengths of his

facts about an agent is not the issue in most of the discussion in this dissertation.

¹⁵ Full exposition of the suggestion will be given in chapters 6, 7.

beliefs and would account for the relations of evidential support between sentences he believes.

I finish my introduction of Davidson's later theory of interpretation with the following remarks. On the one hand, there are the subjectivist's assumptions about probabilities, which let us not challenge here. These are the assumptions that a subjective probability function for an individual determines the degrees of beliefs of the individual, and that the strength of the beliefs of the individual are represented by probabilities. On the other hand, there is the interpretative suggestion that sentences containing theoretical terms are understood on the basis of their evidential relations to other sentences, with the evidential relations being mainly probabilistic. These considerations suggest that a classification of the doxastic states of an individual, which would include the beliefs and the partial beliefs of the individual, is related to the determination of the domain of the evidential relations. Evidential relations are required for interpretation of sentences containing theoretical terms. The importance of the possibility of a classification of doxastic states in the present inquiry shows itself by considering the deprivation of desires, intentions and other affective states of exhibiting the required sort of relations. It is the beliefs and partial beliefs of an individual, what is labelled here the doxastic states of the individual that are characteristically susceptible to evidential relations¹⁶.

I recapitulate the main points I introduced and discussed in this chapter. A theory of interpretation is a theory of understanding. A theory that results in redescriptions of utterances and behaviour as acts is a candidate theory of interpretation. The theory is adequate if the redescriptions as acts it results are made intelligible by ascriptions of attitudes to the speaker/agent. Interpretation of language and behaviour are interrelated enterprises. A theory of interpretation is radical if no prior to theory understanding is presupposed. We considered Davidson's suggestion that a theory, like a modified theory of Tarski's theory of truth has the form of resulting for every sentence of the language in theorems of the kind:

's is true iff p'

is a candidate theory of interpretation. An acceptable theory of this kind is verified by

¹⁶ The consideration presented in the paragraph contain in a nutshell the proposal made in the dissertation. Elaboration of these considerations is given in chapters 6, 7.

use of the principle of charity which is constitutive of interpretation.

I commented on the significance of the problem of identification of the holding true attitude for the principle of charity. The identification is required for determining the domain of applicability of the charity principle. The place of the holding true attitude in a theory of interpretation was discussed. I put forward considerations concerning the sufficiency of one key attitude for interpretation and concerning the indispensable of the holding true attitude. It was suggested that the possibility of identifying the attitude of holding a sentence true is a substantial problem in Davidson's early theory of interpretation. The problem of identification of the holding true attitude is promised a solution in Davidson's later theory of interpretation where the central attitude is considered to be the attitude of preference between the truth of two sentences.

CHAPTER TWO

SPEECH ACT ANALYSIS

2.1 Introduction

In chapter 1 it was suggested that a theory of interpretation is given by a system of redescriptions. In this chapter we discuss an issue concerning the redescriptions of speakers as certain sorts of speech act. The discussion here is about the aspects of a speech act which as identified of an utterance would enable understanding of the utterance. The issue that is central to this chapter concerns the kind of redescription of an utterance minimally sufficient for suggesting the redescription to be eligible for a theory of understanding. Here I will consider an analysis of speech acts taken from speech act theory. The analysis of speech act presented here is indebted to the father of speech act theory, Austin (1962). The aspects of Austin's analysis of a speech act will be presented and the relevance of each of those aspects to interpretation will be examined. Borrowing a jargon from Austin's theory, the main issue discussed here is whether it is locutions or illocutions that are relevant to interpretation. The terms will be explained shortly. There is a suggestion that redescriptions of utterances as certain kind of locutions are the aim of interpretation or it is a step towards interpretation. If identification of the locutionary meaning of an utterance is sufficient for understanding the utterance, or it can be individuated as a first or separate condition towards understanding the utterance, then the interpretation theory will include redescriptions of utterances as certain sort of locutionary acts. This suggestion is the target of the present chapter.

The suggestion requires the possibility of a distinction between locutionary and illocutionary act performed in the performance of a speech act or identification of the locutionary meaning of an utterance. Application of these considerations to a theory of interpretation like Davidson's would suggest the following. A specification of the truth-conditions of the aspect of an utterance, which might be identified as its locutionary meaning, suffices for or it is a step toward interpretation of the utterance. For this to be possible, the locutionary meaning of a speech act should be identifiable as a truth-bearer.

I will discuss the suggestion and argue for a modification of it. The

modification of the suggestion I put forward is that though it is not illocutions that the required redescriptions intend to capture they are locutions identified as of a certain kind. And I will question whether all locutions of the relevant kinds can be identified as truth-bearers.

I start, then, by discussing the possibility of the distinction between locutions and illocutions. Placing the issue in its proper context requires a presentation of an analysis of speech acts. I consider then Austin's analysis of speech acts as presented in *How to do things with words* (1962). Austin writes that in saying something the following acts are performed: a) a phonetic act, which is the uttering of noises; b) a phatic act, which is the uttering of series of words as belonging to a certain vocabulary and as conforming to a certain grammar; c) a rhetic act, which is the uttering of words in a certain construction with a certain more or less definite 'sense' and a more or less definite 'reference'. The last kind of act is the act of saying something in the full sense of saying and the performance of this act is the performance of a locutionary act. Austin's analysis of speech acts includes the notions of illocutionary and perlocutionary act.

"To perform a locutionary act is in general, we may say, also *eo ipso* to perform an illocutionary act, as I propose to call it. To determine what illocutionary act is so performed we must determine in what way we are using locution: asking or answering a question, giving some information or an assurance or a warning,...." (1962)

In addition, saying something might produce certain effects upon the feelings, thoughts or actions of the hearer. Doing this is performing a perlocutionary act. I will not discuss further points of Austin's theory of speech acts, since the points of speech act theory that concern my inquiry are more specific than those Austin's theory discusses.

Austin's attempt to identify the aspects of a speech act in virtue of which the performance of a locutionary and an illocutionary act are distinguished involves the consideration of various criteria. I will present the criteria that have been taken to ground the distinction in question. The criteria will be criticised on the basis of whether they provide sufficient conditions for the distinction in question (2.2). My presentation will be supplemented by arguments intending to show that, in the context of the present inquiry, the distinction that is founded on the criteria is not one that identifies the truth-value relevant aspects of an utterance (2.3).

2.2 Criteria that do not appeal to the notion of truth

According to Austin we perform a locutionary act when we utter certain noises, certain words in a certain construction with a certain 'meaning' in the favourite philosophical sense of that word, that is with a certain sense and with a certain reference (1962, p. 94). An illocutionary act is an act we perform in performing a locutionary one. In performing a locutionary act we use speech. But the way we use speech on a particular occasion is the performance of an illocutionary act (p. 99). Here it could be suggested that the criterion for sharpening the contrast between locutionary and illocutionary act is the 'in saying x, I was doing y' criterion. The criterion is supposed to distinguish between the locutionary act in terms of 'saying' and the illocutionary act in terms of what was being done in 'saying'. In uttering the sentence "The bull is raging." a speaker might perform various illocutionary acts; the utterance can be issued with various kinds of force. The utterance might be a warning, a report, an assertion. But in all these cases of a speech act it is the same saying that has been performed.

Let us now consider whether the criterion provides sufficient conditions for the distinction in question. A closer consideration shows that the 'in saying x, I was doing y' criterion cannot be used as a criterion for distinguishing the locutionary from the illocutionary act performed in the performance of a speech act. If saying something with a certain sense and reference is a mark of a locutionary act, the construction 'in saying x, I was doing y' can be used to specify a locutionary act. Consider the following sentence: "In saying "Your brother sent me a postcard." I was referring to John." The criterion as applied here specifies a feature of the utterance that according to Austin belongs to the locutionary act performed. So it does not demarcate the illocutionary act performed in the speech act. In addition, the criterion does not demarcate the performance of a locutionary act. The construction 'in saying x, I was doing y' applies to the case of a phatic act as well. "In saying "Es regnet.", I uttered a German sentence." would be an example of the construction suggested by the criterion specifying the phatic act performed. So, application of the construction 'in saying x, I was doing y' is not a sufficient condition for classifying 'saying x' and 'doing y' as locutionary and illocutionary acts respectively and, thus, it cannot be used as a criterion for distinguishing between two kinds of act.

Some comments on the relation of the discussion here to the general issue are in due course. What is required for a substantiation of the distinction between locutionary and illocutionary act in the performance of a speech act is a criterion that identifies the locutionary act in contrast to the illocutionary act; a criterion that would point to that sort of use of a sentence token that would mark the performance of the locutionary act. A criterion of the required sort is given for instance for the case of a phatic act as contrasted to the phonetic act. The uttering of a string of noises which are of certain type, uttered in a certain order, that is noises standing for words belonging to the vocabulary of a language and uttered in a certain order that conforms to the grammar of that language marks the performance of a phatic act. A phatic act is contrasted to the phonetic act, a mere uttering of noise, and it is contrasted to the locutionary act. Uttering a grammatical sentence of a language the speaker does not understand amounts to the performance of a phatic act. Both acts are performed in the performance of any speech act. The question that is asked here is whether there is a similar criterion for the performance of a locutionary act as contrasted to the performance of a phatic act and to the performance of an illocutionary act. We seek a criterion of a similar sort for the performance of an act which requires more than the performance of a phatic act and that is contrasted to the performance of an illocutionary act.

Another possible criterion for the distinction in question might be given by the difference in the way the two acts are reported. Since reporting an utterance is a way of redescribing it, this kind of criterion might provide us with an account of the locutionary redescriptions. A case of reporting that is ruled out here as irrelevant is the one that uses inverted commas, that is, reporting by direct quotation. The use of inverted commas marks the phatic act performed by the speaker which is a kind of act that is essentially mimicable, reproducible. A speaker who does not understand German and who utters the German sentence "Es regnet." performs a phatic act. By using inverted commas to report speech we report the phatic act the speaker performed. Thus the 'inverted commas' criterion provides a sufficient condition for demarcating the phatic act performed in the performance of a speech act; it does not identify a kind of act that we want to call locutionary.

A more promising criterion based on the ways of reporting a speech act would

be one that uses the indirect speech construction. Let us examine then how application of the indirect speech construction could help in identifying a kind of act, distinct from the phatic act and distinguishable from the illocutionary act, performed in the performance of a speech act. The suggestion is that a reporting sentence like "He said that the cat is on the mat." is a report of that aspect of a speech act that is to be identified as the performance of a locutionary act. Indirect speech reports are made with the reporting phrase 'tell that' as well. Sometimes in a construction in which the 'tell that' phrase is used the recipient-hearer of the speech act are mentioned. Examples are: "He told me that the cat is on the mat.", "He told Mary that the cat is on the mat.". Of course the recipient-hearer might be specified in a 'says that' construction, as in the case "He said to me that the cat is on the mat.".

Report of a speech act by use of indirect speech is a report of a meaningful 'saying' of the speaker. It is, then, contrasted to direct quotation of the phatic act. A correct report by an indirect speech construction of the speech act of a German speaker uttering the sentence "Es regnet." would be, in English, "He said that it is raining.". Granted the indirect speech criterion the suggestion is that the relation of redescription between uninterpreted utterances and locutionary acts would be one holding between the utterance and an indirect speech report of the utterance. The indirect speech report redescribes the utterance as a certain saying. The utterance 'Es regnet.' is redescribed by the report "He said that it is raining.".

The indirect speech report of an utterance redescribes the utterance not as the making of a phatic act. The question here is whether the redescription of the utterance offered by the indirect speech construction identifies an aspect of the utterance contrasted to the illocutionary aspect of the utterance. Proponents of the indirect speech criterion answer affirmatively to this question. Their reason is that an indirect speech report exhibits the meaning of the utterance but not the illocutionary force of the utterance; it leaves unspecified the kind of illocutionary act performed in the utterance of the sentence under interpretation. Someone who understands the indirect speech report sentence "He said that it is raining." understands an aspect of the utterance of which it is a report, different from the phatic act performed, but he need not have a clue about the illocutionary act the speaker performed. Thus, on the basis of an indirect speech report an aspect of the utterance different from the phatic act

performed in making it but distinguishable from the illocutionary act is identified. That aspect of the utterance redescribed by the indirect speech report identifies the locutionary meaning of the utterance. An indirect speech report exhibits the locutionary meaning of an utterance and it reports the performance of a locutionary act.

The indirect speech report is contrasted, on the one hand, to reports by direct quotation that demarcate the performance of a phatic act, and on the other hand, to reports of utterances in which the kind of illocutionary act performed is explicitly mentioned. “He said that the cat is on the mat.” as a kind of report that identifies the locutionary act performed by the speaker is contrasted to reports like “He warned me that the cat is on the mat.” or “He asserted that the cat is on the mat.”. The last ones are cases of reporting the illocutionary act the speaker performed in making the utterance under interpretation. The construction of indirect speech report by means of the reporting phrase ‘says that’ provides, then, a criterion for distinguishing between the locutionary and the illocutionary act performed in speech.

At this point I intend to discuss a certain application of the criterion. I think that this discussion will show that, in a certain sense to be clarified shortly, what is identified by the criterion is not independent of the category to which the speech act belongs. The indirect speech criterion identifies the ‘saying’ of the speaker as belonging to a certain category of speech acts.

Let us, then, consider the application of the criterion in the two following cases of utterances made by a certain male individual.

1. “John goes to the movies.”
2. “John, go to the movies!”

An indirect speech report providing a criterion for identifying the locutionary act performed by the speaker would exhibit the locutionary aspect of the two cases of speech acts. A report of speech act 1 using indirect speech is “He said that John goes to the movies.”. The report does not specify the kind of illocutionary act the speaker performed in uttering the sentence. It does not specify whether the speaker made an assertion, gave a report, made a guess in uttering 1. Proponents of the criterion suggest that the report specifies the locutionary meaning of the utterance.

Looking at case 2 we notice that there is no direct application of the criterion. A natural way to report the speech act 2 is by the construction “told ...to...” instead of “...said that...”. An indirect speech report of speech act 2 would be “He told John to go to the movies.”. The two constructions are different. The “...told ...to...” construction in the case of reporting speech act 2 is followed by a subordinate clause constructed with the verb in the infinitive. In contrast, in a report of speech act 1 made either with the “...said that...” reporting phrase or with the “...told...that...” reporting phrase the verb of the subordinate clause is in the indicative mood. I discuss the difference between indirect speech report of case 1 and indirect speech report of case 2 immediately below. But, notice that, despite the difference, there is a similarity between the indirect speech report of case 2 and an indirect speech report of case 1. The similarity is that in a report of case 2 by use of “...told...to...” as in a report of case 1 by use of “...said that...” there is no specification of the particular illocutionary act the speaker performed. The report “He told John to go to the movies.” does not specify whether the speaker gave an order, made a request, or gave advice in uttering the sentence. There is no specification of the particular illocutionary force with which the sentence was uttered.

Now let us see whether there is a difference between the two cases. It seems that there is. The difference in the construction used in the two cases of indirect speech report signifies different aspects of the two utterances. Someone, who understands the two reports, “He said that John goes to the movies.” and “He told John to go to the movies.”, understands that two different kinds of speech act were made in the two cases. The indirect speech reports of 1 and 2 redescribe the two utterances by, simultaneously, exhibiting a difference between the utterances. The use of the infinitive for the verb in the subordinate clause of the reporting sentence of utterance 2 signifies a difference between the two utterances. The question here is where this difference should be located.

If we want to insist that the indirect speech report construction provides a criterion for identifying the locutionary meaning of an utterance, then we will have to locate the difference between the two reports in the locutionary aspect of the speech act reported. In this case speech act 1 and speech act 2 do not have the same locutionary meaning. This is an undesirable consequence for those who want to

identify the locutionary meaning of a sentence with a proposition that has the properties of being a truth bearer and being something that is identified in sentences 1 and 2. Nevertheless the suggestion is not unattractive. The criterion of indirect speech, according to this suggestion, identifies the locutionary meaning of an utterance independently from the particular illocutionary force of the utterance but it identifies it as being a locution of a certain kind. The use of the infinitive or indicative mood for the verb in the subordinate clause of the reporting sentence signifies that the utterance under interpretation is redescribed as a locution of a certain kind. This distinction will be further elaborated later in the context of presenting other criteria. It will be suggested that the distinction is significant for the truth-evaluable/non-truth-evaluable distinction.

The alternative suggestion to the one made above would be to abandon the idea that the criterion of indirect speech report identifies the locutionary meaning of the sentence. In this case we are still in need for a criterion for a distinction between a locutionary act and an illocutionary act performed in the performance of a speech act and for a distinction between locutionary meaning and illocutionary force. I think that the criterion is not pointless and that we should keep it. In certain cases of utterances, the criterion does identify the locutionary act performed. So, application of the criterion in the case of performing the speech act of asserting in uttering the sentence "John goes to the movies." and in the case of performing the speech act of reporting by uttering a token of the same sentence does identify the same locutionary meaning. In some cases the criterion is a useful one because it enables identification of the locutionary meaning of a speech act that belongs to a certain general category. What we should keep in mind in applying the criterion is that the criterion does not grant the desirable identification of a locutionary act/locutionary meaning of the speech act independently of any other aspect of the speech act. It grants an identification of the locutionary act/locutionary meaning in a way that classifies it as being an aspect of a speech act of a certain general category. It does not exclude the possibility of having to distinguish between two kinds of locutionary meaning. The following attempt to clarify further the claim.

Observing certain features of grammatical constructions allows us to make the following conjecture. The observations concern constructions in indirect speech of

verbs that express an affective attitude, that is verbs with an object that relates to the performance of an action.

The term 'affective' here should not be confused with the term 'affect' as it appears in theories of emotion (Clair Armon-Jones 1991, Michael Stocker 1996) or as the common use of term might suggest¹⁷. So according to the use of the terms 'doxastic' and 'affective' suggested here, both doxastic and affective states can be related to emotions. For example, someone might be pleased by the nice weather, or be pleased to have a walk with a friend. (The point is clarified further in a following paragraph where contentful emotional states are discussed.) The attitudes that are here labelled affective are related to action. So, according to the use of the term affective suggested here, someone is attributed an affective attitude towards the contents specified by the that-clause that it is raining as long as he can see rain to be a consequence of an action. Normal cases would not allow for this, but we can think of ritual ceremonies taking place in certain tribes where rain is seen to be a consequence of their performance. So, an individual has an affective attitude about a certain case, if he can see the case as an action to be performed or as a consequence of his actions. Here the conjecture we make is that indirect speech reports of this category of verbs are constructed with the verb of the subordinate clause in the infinitive or, in some cases by use of modal and infinitive. In contrast, indirect speech reports of verbs that express a doxastic attitude are constructed with the verb of the subordinate clause in the indicative mood.

In grammar texts we find the following suggestion concerning the issue. The category of verbs which in the active voice can be followed by an infinitive or a that-clause containing a modal and infinitive includes the following verbs: advise, agree (when it concerns an action), ask, beg, command, decree, demand, direct, insist (to do something), intend, order, plead, pray, prefer, propose, recommend, request, rule, suggest, urge. In addition, there is a range of verbs that in the active voice they are constructed with a that clause in the indicative. This range includes the following verbs: accept, acknowledge, add, admit, agree, allege, announce, answer, argue, assert, assume, assure, believe, boast, claim, comment, complain, concede, conclude,

¹⁷ The doxastic/affective distinction relates to the 'direction of fit' of the attitude. The affective attitudes are contrasted to the doxastic attitudes in the sense that they have different direction of fit. The notion of direction

confirm, consider, contend, convince, decide (in the sense of concluding or coming to believe), deny, determine, discover, dispute, doubt, dream, elicit, estimate, expect (in the sense of believing), explain, fear, feel (in the sense of having an opinion), figure, find, foresee, forget, gather, guarantee, guess, hear, hold, hope, imagine, imply, inform, insist (in the sense of believing-claiming, maintaining a view on a matter), judge, know, learn, maintain, mean, mention, note, notice, notify, object, observe, perceive, persuade, pledge, pray, predict, promise, prophesy, read, realise, reason, reassure, recall, reckon, record, reflect, remark, remember, repeat, reply, report, resolve, reveal, be rumoured, say, see, sense, state, suppose, swear, teach, tell, think, threaten, understand, vow, warn, wish, worry, write. (Collins Cobuild, 1990, p. 325).

An interesting phenomenon to observe is that verbs that turn up in both lists, for instance the verb to agree, are constructed with a that-clause in the indicative when they are used doxastically; "He agreed that the weather is nice today.". In cases in which the same verb is used to express an affective attitude, an attitude about how one should act, like "He agreed that he should give the lecture." the construction is one with a modal plus infinitive occurring in the that-clause. Of course I do not take this sort of empirical conjecture to be a justification of anything that I suggest. It only points to a phenomenon of linguistic usage that is in accordance with the suggestion I make here; it does not justify it.

At this point it would be advisable to make some comments on a special case of attributions, the case of contentful emotional states. The range of contentful emotional states includes hope, fear, amazement, astonishment, embarrassment and many other. A way to analyse attributions of contentful emotional states that is in accordance with the line I consider here is the following. In an attribution of an emotional state we distinguish between the content of the state, which can be of a doxastic or an affective kind –the locutionary meaning of the report identified as belonging to a doxastic or an affective category- and an emotion. So, for instance, in the attribution "John was amazed that it suddenly started snowing." we distinguish between the doxastic content specified by the that-clause, 'that it suddenly started snowing' and the emotion of amazement¹⁸. There are cases of emotions that can

of fit and an adequate account of it will be elaborated in chapters 5.6.7.

¹⁸ Robert Gordon (1969) claims that in the case of amazement the propositional content is not just of the

accompany both doxastic and affective states. Examples are cases of the emotion of fear accompanying a doxastic or an affective state. “John fears that the punishment will be harsh.” and “Mary fears to disobey the regulation despite her judgement that the regulation is unfair.” are two cases in which the attribution of an emotional state of fear has a content of a different kind. In the case of the first attribution the content is of the doxastic kind while in the case of the second attribution the content is of the affective kind. In the first case the subject considers the possibility of the occurrence of an event with an uneasy sense of fear. In the second case the subject envisages the taking of a course of action from which she refrains.

Similar, in the relevant respect, to the case of fear is the case of hope. “John hopes that the punishment will be harsh.” and “Mary hopes to disobey the regulation.” could be treated in a similar manner with the cases of fear. In the first example the subject, John, considers the possibility of the occurrence of the event that the punishment is harsh with happiness. In the second example the subject, Mary, considers taking the option of disobeying the regulation with a feeling of pleasure.

The claim that emotional states like fear or hope do not accompany one kind of content is made plausible by further considerations. The claim is that emotional states like fear or hope do not implicitly carry a certain direction of fit. These states inherit their direction of fit from the direction of fit of the content they have. A clumsy criterion for distinguishing between direction of fit is based on what is at fault in cases that things go wrong¹⁹. According to this criterion a state is of the doxastic kind if, in case things go wrong, it is the state that is at fault. A state is of the affective kind if, in case things go wrong, it is the world that is at fault.

Let us apply then this criterion on the two cases of fear attribution discussed above. Considering that things go wrong in the case of fear attribution by the sentence “John fears that the punishment will be harsh.” we could say the following. If it turns up that the punishment in question is not harsh the subject should give up both the belief and the fear. The subject had a false belief or a wrong anticipation that was the

doxastic kind. It is not sufficient that the subject believes that p in the cases he is embarrassed that p. It is required that the subject knows that p. The same is suggested for the cases of amusement, embarrassment, being annoyed, being delighted and other.

¹⁹ In chapter 5 I criticise this criterion for being inadequate for grounding the distinction between doxastic and affective states in the context of this dissertation. Here I use it only in order to exhibit the possibilities of fear

basis of his fear. Similar is the case of things going wrong with the attribution “John hopes that the punishment will be harsh.”. In case the punishment is not harsh the subject should give up both the belief and the hope.

Let us now consider the case that things go wrong with the attribution “Mary fears to disobey.”. In case that Mary disobeys there is no ground on which it is judged to be wrong that she had the intention to refrain from disobeying. What is wrong is that she acted against her intention. It is not the intention that was flawed. Similarly, in case things go wrong with the attribution “Mary hopes to disobey.”. In case she obeys there is no ground on which her intention to disobey is judged to be flawed. It is acting against her intention that is at fault; it is not the intention at fault.

I return to the main issue of this section which is about conditions for identifying the locutionary meaning or locutionary act in a way that shows it to be separable from illocution. The criterion of indirect speech that was presented previously distinguishes between locutionary and illocutionary act in the performance of a speech act. But the criterion identifies locutionary meaning as being of a doxastic or affective kind.

Another possible criterion for the distinction in question based on Austin’s writings is one that identifies the locutionary act as the act of uttering a sentence with a certain ‘meaning’, which in his own words is accounted

“in the favourite philosophical sense of that word, as having a certain sense and a certain reference”.

We find a development of this suggestion in Searle (1969). Searle’s theory includes an attempt to expand and clarify Austin’s ideas on the notion of a speech act. He takes the marks of the locutionary act, -in his theory a locutionary act is called propositional and a locution is called proposition-, to be reference and predication. Two distinct illocutionary acts which involve the performance of the same type of locutionary act in virtue of identity conditions applying to reference and predication. The following are examples of utterances made in speech acts which, according to Searle’s suggestion, involve the same locutionary act but are classified as distinct illocutionary acts.

1. Sam smokes habitually.
2. Does Sam smoke habitually?
3. Sam, smoke habitually!
4. I wish Sam smoked habitually.

Searle comments that in all cases of speech act stated above we can distinguish three kinds of act: a) Uttering words which is the performance of an utterance act. The utterance act is equivalent to Austin's phatic act. b) Referring and predicating which is the performing of a propositional act and c) stating, questioning, commanding, promising which are performances of illocutionary acts. So, for Searle, the identity of the locutionary act is determined in terms of the propositional acts of reference and predication.

"Whenever two illocutionary acts contain the same reference and predication, provided that the meaning of the referring expression is the same, I shall say the same proposition is expressed". (1969).

In addition he makes the following comments:

"...a proposition is to be sharply distinguished from an assertion or statement of it, in utterances 1-4 the same proposition occurs, but only in 1 it is asserted. Stating and asserting are acts, but propositions are not acts. A proposition is what is asserted in the act of asserting, what is stated in the act of stating. The same point in a different way: an assertion is a (very special kind of) commitment to the truth of a proposition." (1969. p. 29).

Let us discuss some points related to these considerations. The crucial question here is the question of what is this that admits of truth. It seems that there are two possible answers to this question. One possibility would be to take the propositional content of a certain kind of illocutionary speech act, namely of the kind of assertive illocutionary acts –assertive here has to be read as a generic term- to be what admits of truth. This answer is motivated by the fact that we call assertions true and not questions, commands etc. In this case the truth-apt propositional contents would be the contents of the assertive utterances. of a classification of truth-evaluable utterances. A second possibility would be to take the propositional content identified independently of illocutionary force to be a truth bearer. But in the context of the present discussion this would require identity condition for proposition or propositional act before they are assigned truth-conditions.

Both Searle and Austin seem to oscillate between the two possibilities. Austin oscillates between the two views about truth-bearers. On the one hand there is the view that truth bearers are locutions identifiable independently of any illocutionary

mode and, on the other hand, there is the view that truth-bearers are certain illocutions. This oscillation of his between the two views has given the right to some commentators to interpret Austin as accepting that locutions are truth bearers. Richard Garner (1972) points to inconsistencies in the way Austin expresses his views, which have caused confusion about the relations between locutionary act, truth, and illocutionary force in his writings. There are some contexts in which Austin takes truth and falsity to attach to assertions, and, thus, to mark the performance of an illocutionary act, and other contexts where it seems that he takes truth and falsity to mark the performance of a locutionary act. But he argues that an overall evaluation of Austin's writings suggests that locutions of assertive speech acts are the truth-bearers. Garner's interpretation of Austin is very close to where my arguments will lead. His interpretation suggests that locutions of a special kind, the ones issued with what he calls constative force and belonging to the assertive kind of acts are truth-bearers. This interpretation suggests that the truth bearers are the locutions of a certain category of speech acts. The point at which I will diverge from this interpretation is that I suggest we expand the category of truth-bearing locutions so that it includes, in addition to assertions, locutions of suspicions, conjectures and of other truth-evaluable utterances.

Searle seems to oscillate between the two views as well. So, while he says that in the examples 1-4 stated above the same proposition occurs and propositions are identified in terms of reference and predication, his analysis of predication gives rise to a different interpretation. While he admits that reference is an abstraction separable from the speech act, he states that "predication never comes neutrally but always in one illocutionary mode or another". (1969, p. 123).

Now, since the speech acts of the examples 1-4 are each made with a different illocutionary force and predication does not come neutrally of illocutionary force, the claim that there might be a different propositional act performed in each case is not refuted. The suggestion does not show predication to be an aspect of a speech act that can be identified independently from any illocutionary mode. If this were the case then a common locution across 1-4 could be identified. The identity of propositions across 1-4 would require an account of the nature of predication such that it is independent of whether it is employed in assertions, conjectures, hypothesis or it is

employed in commands. As long as such an account is not at hand, the proposal that truth-evaluability of propositional content is relevant to contents of, say, assertive kind of speech acts and not of directive kind is not eliminated.

In Searle's later writings (1985) there is a less explicit account of the notion of a propositional act. He grants a distinction between illocutionary force and propositional content on the basis of different identity conditions:

"the same propositional content can occur with different illocutionary forces and the same force can occur with different propositional contents". (Searle, 1985)

But this suggestion, in the context of the present discussion, would be question begging. It begs the question concerning the issue of whether the same kind of locution is identifiable across different kinds of speech act.

The issue we have been discussing up to here concerns the possibility of a criterion that would enable us to identify an utterance as an act of saying in a sense of doing something more than or different from performing a phatic act. Identification of the phatic act performed in making an utterance is not sufficient for interpreting the utterance. A question that is examined here is how it is possible to identify an utterance as an act of saying distinguishable from the illocutionary act performed in the performance of a speech act. Given that further conditions are satisfied, the identification of the utterance of a sentence as an act of saying would enable understanding of the utterance. A theory of interpretation of language would include redescriptions of utterances as sayings. The possibility of an identification of an act of saying that would enable understanding brings in the question of whether there is a genuine act of forceless 'saying' distinct from the phatic act. We saw that a criterion based on indirect speech report identifies an aspect of the speech act independently of the particular illocutionary force with which it is uttered but only as belonging to a certain category of speech acts. In understanding an indirect speech report we understand something more than a forceless saying of the speaker; we understand it as a saying of a certain kind.

There are certain suggestions implied by general facts about linguistic behaviour. A linguistic act is not the mere uttering of noises that happen to be words belonging to a certain vocabulary. Linguistic acts are performed intentionally; a linguistic act is an intentional act. In performing a speech act, the speaker

communicates thoughts in order to achieve purposes, to carry out intentions. The conception of a locutionary act as a forceless ‘saying’ would not be sufficient for identifying it as linguistic action. While these considerations about linguistic behaviour are more or less straightforward the additional issue discussed here is whether a forceless ‘saying’ can be identified as a separate condition for understanding an utterance.

We saw that indirect speech reports present the ‘saying’ of the speaker as belonging to a certain category; an indirect speech report of the utterance “Sam smokes habitually.” and an indirect speech report of the utterance “Sam, smoke habitually!” present the two ‘sayings’ of the speaker in different ways. In the first case, the reporting sentence “He said that Sam smokes habitually.” reports the ‘saying’ of the speaker as being of the kind sometimes called assertive speech acts. In the second case, the reporting sentence “He told Sam to smoke habitually.” reports the ‘saying’ of the speaker as belonging to another category of speech acts named by Searle ‘directives’. A report of a forceless ‘saying’ of the speaker, that is a report that would not specify even the general category of speech acts to which the ‘saying’ belongs, is not given by the indirect speech report. The criteria we considered by Austin and Searle for identifying the locutionary act identify it as being of a certain kind. They do not grant common locutions across general categories of speech act. For an identification of propositional content/locutionary meaning across assertive/truth-evaluable and directive/non-truth-evaluable speech acts to be possible, their account needs, at least, to be supplemented. These considerations lend support to the idea that it might be required to stipulate more than one general category of locutionary meaning; the assertive/doxastic and the directive/affective.

2.3 Locutionary meaning as the truth-value bearer

Up to now, the criteria I considered for a distinction between locutionary and illocutionary act did not employ the notions of truth and falsity. I will consider now the role the notions of truth and falsity can play for the distinction in question. First I will discuss the way the notions of truth and falsity have been employed by speech act theorists concerning the distinction in question.

The suggestion is along the following lines. A locutionary speech act is the act of saying something true or false, while an illocutionary act is the act of saying something other or more than saying something true or false. The criteria we considered up to now were suggested as part of an attempt to identify the locutionary meaning independently of the notion of truth so that we could take the extra step of assigning truth conditions to the locutionary meaning. The suggestion here is to identify the locutionary meaning as what is truth-evaluable in the utterance, and the locutionary act as the saying of something truth evaluable.

A distinction between locutionary and illocutionary act in the way suggested here is similar to the performative/constative distinction as a distinction made between utterances. The notion of truth is employed for making the distinction. It has been suggested that a possible identification of a class of utterances, called performatives, be given by contrasting them to utterances that are the mere saying of something true or false. So, Max Black (1970), glossing Austin's notion of performatives, suggests: an utterance is said to be performative when used in specified circumstances, if and only if, its being used so counts as a case of the speaker's doing something other than or more than saying something true or false. Otherwise it is called constative. Call this definition A. This definition of performatives is contrasted to a definition of the notion of performative utterance considered by Austin. According to this definition (say definition B) an utterance of the form 'I X (such and such)' is said to be performative if and only if its being used counts as a case of the speaker's thereby X-ing. Black claims that it is the first definition considered which will serve the purpose of the inquiry and that difficulties existing in Austin's theory of performatives can be overcome by this choice.

In order to test this definition for its adequacy we have to make clear the cases of utterances that are the mere sayings of something true or false. Under what circumstances does the utterance of the sentence "The bull is loose." count as a mere saying of something true or false? In uttering the sentence I say something that actually is true or false but we can consider cases in which the uttering of the sentence is a warning. According to the definition in these cases the utterance of this sentence should be classified as a performative. More problematic classifications on the basis of the definition are the cases of assertions. In uttering the sentence "The cat is on the

mat.” I say something which is true or false but at the same time I might assert that the cat is on the mat, which is something more than merely uttering of a sentence which is true or false. Is then the utterance of a token of the sentence "The cat is on the mat.” a constative? But is it not the case that the making of an assertion in uttering a sentence should count as doing something more than saying something true or false? If this is not the case, and we want to exclude assertions from performatives, then assertions will have to be considered as performed as part of other speech acts. This will have undesirable consequences. It is counterintuitive to suggest that someone who expresses a wish by uttering “Good luck!” asserts that the hearer has good luck. It seems that assertions should be classified on the basis of definition A as performatives and treated in a similar manner to other speech acts.

The distinction between performatives and constatives given by definition A introduces a distinction among the class of utterances parallel to the intended illocutionary/locutionary act distinction in the performance of a speech act. We saw that there are some difficulties in trying to identify that aspect of a speech act that theorists want to call the locutionary act independently of any illocutionary feature of the speech act. The difficulties are similar, I think, if we try to substantiate a constative/performative distinction similar to the one of illocutionary/locutionary aspect of a speech act among the class of utterances.

Generally it is assumed to be the content of a statement, of an assertion, a declaration of which truth and falsity are predicated. The utterance of a declarative sentence or of an assertive sentence, that is the issuing of an assertion or a declaration is not anything that can be true or false itself. These considerations have led philosophers to the distinction between the content and force of an utterance. According to this line of thought, it is the content of an assertion, the proposition expressed by it, or the propositional content of the utterance that is a truth/falsity bearer. This line of thought, though dominant in contemporary philosophy, is not uncontroversial. A consideration that might be relevant here is that there does not seem to be any genuine linguistic act which is the mere expressing of a proposition; the performance of an utterance which counts as a case of the speaker saying something true or false and not at the same time doing something other or more than this seems at least to be unmotivated.

At this point there is a clarification that should be made. The line of inquiry that we follow is to examine every syntactically well formed sentence as a potential utterance, that is, what a potential speaker means or does in issuing the sentence. Mathematical sentences or analytic sentences are held to have a constant truth-value in the sense that whoever utters them or whenever they are uttered the speaker is supposed to have said something which has a constant truth-value. But what the speaker does in issuing the utterance is something more than saying something true or false; he asserts, he declares, he states. Even utterances from the class of mathematical sentences are not the mere sayings of something true or false; they are uttered with the force of an assertion, as hypotheses, as conjectures.

Clarifying this issue is important to our inquiry. What needs to be clarified are the conditions in virtue of which relations like 's' is redescribed as 'p' are established, where s and p are sentences belonging to the language and the metalanguage respectively. The discussion in this chapter concerns what is involved in the enterprise of the interpretative redescription. It is about conditions on the 'the sayings' that as identified of utterances enable interpretation of language.

I return to the possibility of a distinction between locutionary and illocutionary acts in the performance of a speech act. The suggestion that locutionary meaning of an utterance as the truth evaluable aspect of the utterance is in accordance with Davidson's proposal. In Davidson's theory truth is suggested to be the key notion to meaning. Additional problems to the ones presented above -that a characterisation of a genuine saying is not exhausted in its truth-evaluability- concern the identification of the truth-evaluable aspect of sentences in non-indicative mood. The truth-evaluable aspect of an utterance, according to the present suggestion identifies the locutionary meaning of the utterance. I will offer a more thorough analysis of mood in the following chapter but let us make here a brief introduction to some related problems. These are problems that the suggestion that locutionary meaning is identified as the truth-evaluable aspect of the utterance encounters, and they are related to the phenomenon of mood.

In uttering the sentence "Does John leave?" we ask a question. In uttering the sentence "John, leave!" we may give an order. It is not straightforward that in uttering these sentences we are saying something true or false. There is a suggestion that with

a proper analysis of these sentences the hidden common locutionary meaning would be made explicit. So the sentence "Does John leave?" should be analysed as "I ask you whether it is the case that John leaves.". The sentence "John leave!" should be analysed as "I order John to make it the case that John leave."²⁰. A suggestion to be rejected here is that the locutionary act we perform in uttering the sentence is the saying of the expression 'John leaves' which is something true or false. This suggestion is rejected on the ground that the meaning of the expression 'John leaves' in the context of an order or of a question is the same as the meaning of "John leaves." uttered alone. As I suggested above, the utterance of "John leaves." belongs to the assertive category of speech acts.

I proceed with presenting further considerations on the possibility of an identification of the locutionary meaning of an utterance. Austin wanted to keep the distinction between locutions and illocutions, insisting that the locutionary meaning of an utterance has to do with its sense and reference. The first thing that comes to mind when we talk about meaning as sense and reference is Frege's proposal. Probably the Fregean theory of sense and reference of component parts of a sentence does not serve Austin's purpose. Austin distinguishes between locutionary meaning and illocutionary force of complete utterances; he does not touch upon the issue of the meanings of the components of a sentence. Frege acknowledges an element in the meaning of a sentence, the force of a sentence distinct from the sense of the sentence, the thought expressed by it. But a thought for Frege is something that is true or false and the distinction between the sense and force of a sentence is introduced by him in order to keep distinct the apprehension of a thought from the recognition of its truth-value. He states clearly that although he does not deny sense to a command, this sense is not such that the question of truth arises (1892, 1918). A command and a request are not thoughts. The meaning of the subordinate clause depending upon a command or a request is a command and a request. Michael Dummett commenting on this part of Frege's philosophy attributes to earlier Frege the view that a difference between assertoric, interrogative, imperative is a difference in their sense rather in the force.

"...In "Ueber Sinn und Bedeutung" he explicitly repudiates the view that any sentences other than assertoric express thoughts; rather, he regards the difference between assertoric, interrogative,

²⁰ In chapter 3 a general discussion concerning the truth-evaluability of non-indicatives is presented. Concerning an attempt to extract a truth-evaluable locution out of non-indicatives see in particular section 3.6.

imperative and optative sentences as a difference in their sense rather than in the force attaching to them. Thus he says that, just as assertoric sentences express thoughts, so interrogatives express questions, imperatives commands and optatives wishes." (M. Dummett, 1981, p. 307)

In "Der Gedanke" Frege modifies his view on the subject. He maintains the view that the sense of a command is not something about which a question of its truth arises. But he changes his view concerning questions. As we can assert the truth of a thought we can question the truth of the thought. Frege's distinction between sense and force is relevant to his theory of truth and assertion.

So it seems that Frege's doctrine of sense and force is not on a par with views that identify locutionary meaning as the truth-relevant aspect of an utterance, independently of the kind of utterance. Austin concludes that

'in general the locutionary act as much as the illocutionary is an abstraction only; every genuine speech act is both (p. 147).

But it seems that we are still in need of a criterion for identifying the locutionary act that is to be abstracted. We saw that use of indirect speech, the notions of truth and falsity, the Fregean sense of a sentence do not provide a clear way of isolating the locutionary meaning of a sentence from other features of the sentence.

The next step is to examine how Austin's doctrine of infelicities is relevant to the point. Austin claims that

'typically we distinguish different abstracted 'acts' by means of the possible slips between cup and lip, that is, in this case the different types of nonsense that might be engendered in performing them' (1962, p. 147).

If the person who utters "I announce that the president has decided to withdraw the statement." is not the authorised person to make such an announcement, the utterance is an unhappy one. Someone may wish to distinguish the locutionary from the illocutionary act performed in issuing the utterance by pointing to the different ways in which an utterance can be an unhappy one. For example, the ways in which the utterance can be wrong when the speaker is not the appropriate person to make such an announcement are different from the way the utterance can be wrong when it is simply not the case that the president has decided to withdraw the statement. The first one is a case of fault in the locution while the second one is a case of fault in the illocution. If it is the case that the president has not decided to withdraw the statement, the utterance is an unhappy one in respect of the locutionary act performed. The utterance is unhappy with respect to the illocutionary act performed

when the speaker is not the authorised person to make such an announcement.

We clarify here the following: The doctrine of infelicities supplies the means for drawing the required distinction between locutionary and illocutionary act in relation to some utterances. The doctrine of infelicities, -that is that things ‘going wrong’ in different respects- does introduce a distinction between locutions and illocutions for the case of speech acts whose performance is governed by a constitutive convention. In the cases of such utterances if the convention is abused the performance misses its point as a performance of an illocutionary act. In constitutions in which a certain person is authorised to report the president's deeds to the public the utterance “The President has decided to withdraw the statement.” is a successful one if the conjunction of the following holds: a) the announcement was made by the authorised person b) the utterance is true. In the case in which, in this kind of constitution, the speaker who made the announcement is not the one assigned such an authority the utterance is an unhappy one with respect to the illocutionary act performed. In the case in which the person was authorised but the utterance is not true, then the utterance is an unhappy one with respect to the locutionary act performed. We can conclude that for the case of those speech acts which are governed by a constitutive conventions the doctrine of infelicities distinguishes locutions from illocutions: in the case the convention is abused the speech act is an unhappy one with respect to the illocutionary act performed and in case the failure of the speech act is not because the constitutive convention is abused then the utterance is an unhappy one as a locution

But what happens in cases in which there is no constitutive convention for the performance of a speech act? We can think of cases in which the context of the utterance "The President has decided to withdraw the statement." is not such that the utterance is to be taken as an announcement, that is, in the particular context of the certain constitution there is no convention according to which a certain person is authorised to state the President's deeds, or the context of the utterance is an informal conversation between friends.

The distinction between constitutive and non-constitutive convention is an important one but admittedly not easily clarified. There is a sense in which language is generally conventional. The English word ‘book’ is not a consequence of the mere noise ‘buk’. There is a rule and a convention in accordance with this rule with which

English speakers' use agrees that the word 'book' refers to a certain kind of object. This is roughly the sense in which language is in general conventional. The difference between this kind of convention and a constitutive kind of convention is that in the former case the convention links words to things or to forms of behaviour which exist independently of the convention, while in the latter case the convention establishes or introduces new forms of behaviour. Searle (1969) speaks about regulative and constitutive rules. Regulative rules regulate antecedently or independently existing forms of behaviour, while constitutive rules do not merely regulate, they create or define new forms of behaviour. His hypotheses is that

'The semantic structure of a language may be regarded as a conventional realisation of a series of sets of underlying constitutive rules and that speech acts are acts characteristically performed by uttering expressions in accordance with these sets of constitutive rules' (1969, p. 37).

Searle's distinction is here accepted but this does not show that all speech acts are performed with a constitutive convention. Which is, for instance, the constitutive convention according to which the utterance of the sentence "The bull is raging." is a declaration? If the performance of the act of declaring were governed by a constitutive convention, it would be explained and understood only by appeal to a convention which is constitutive of that certain form of behaviour, that is, declaring. Declaring is not a new form of behaviour introduced by any kind of constitutive rule; we declare such and such when we believe such and such and we intend to express such a belief. Similar is the case of exhorting. We exhort someone to do a certain thing when to desire him to do such a thing and we intend to express such desire. On the contrary I consider acts like announcing, betting, ordering to be institutional acts. We can imagine societies in which there is no form of behaviour like ordering because there is no convention according to which someone has the *right* to punish someone else when he does not act in accordance with the commander's exhortation. But that a speaker has an intention to exhort the hearer to do something does not depend on any constitutive convention.

How could we then apply the doctrine of infelicities for cases of acts like declarations or exhortations? The suggestion in this case would be along the following lines. The act of declaring by use of the sentence "John will be at the party tonight." Can go wrong the belief that it expresses is wrong. There does not seem to be another way in which the utterance could be wrong. In this case, the doctrine of

infelicities does not result in a distinction between locutions and illocutions

However there is a condition for the happiness of the performance of such a speech act; namely a sincerity condition. We can declare that something is the case while we do not actually believe so and we can exhort someone to act in a certain way while we do not actually want. The condition of sincerity is a condition without which interpretation would not be possible. It has been classified as one of the methodological constraints of a theory of interpretation. But it is not a condition that would provide the required distinction here. So, it seems that the required distinction between locutionary and illocutionary act or locutionary and illocutionary meaning in the general case cannot be based on the doctrine of infelicities.

Here is a review of the discussion in this chapter. The issue discussed here was about a condition of interpretation theory, the condition that utterances are redescribed as locutions. The subject of the discussion was on whether redescriptions of utterances as locutions could be identified as a separate condition in interpretation theory. The general discussion here suggests that utterances can be identified as locutions of a certain kind, as doxastic or affective locutions. We considered some criteria for distinguishing locutions from illocutions. The indirect speech criterion identifies locutions as of a certain kind. The criterion based on truth-aptness of locutions has some difficulties in identifying locutions in the non-assertive category of speech acts. The doctrine of infelicities, while it applies clearly in the cases of speech acts performed with a constitutive convention, faces some problems for the cases of other speech acts.

CHAPTER THREE

THE MOOD CRITERION AND THE PHENOMENON OF MOOD

3.1 Introduction

The results of the previous chapter lend support to the assumption that locutionary meaning might be of a doxastic or of an affective kind. The speech act performed by uttering the sentence "The bull is raging." has content of a doxastic kind, exemplified by the indirect speech report redescription of the utterance "He said that the bull is raging.". In addition the speech act might be a warning that the bull is raging. The speech act performed by uttering the sentence "Let loose the bull!" has a locutionary meaning of an affective kind, exemplified by the indirect speech report redescription of it "He told me to let loose the bull.". In addition the speech act might be a command to let loose the bull. If we accept these results, we are invited to answer a question that, as a consequence, arises. The question is whether there are identity conditions for the general categories of speech act. Identifying the assertive or doxastic category of speech acts in a way that would contrast it to other categories might contribute to the identification of a range of utterances that are truth evaluable. Common sense intuitions suggest that assertions, conjectures, reports are truth-evaluable while commands, exhortations, and requests are not. Identification of the assertive category of speech acts- in the extended sense that does not include only assertions- amounts to the identification of truth-evaluable utterances. In the first chapter of this essay we discussed the significance of the holding true attitude in a theory of radical interpretation and the significance of an identification of the doxastic states of agents. Because the central problem of the dissertation is the possibility of a criterion for distinguishing between truth-evaluable and non-truth-evaluable utterances or between doxastic and affective states of agents/speakers, I am not concerned here with identity conditions for each kind of act from each category, i.e. conditions for assertions, conditions for reports, conditions for conjectures. I am mainly concerned with the possibility of identifying the category of speech acts that have as a common feature that they are susceptible of truth-evaluation. We could label this category of speech acts assertive without letting the term mislead us. Assertions as well as reports and conjectures belong to the assertive category of speech acts.

In this chapter I examine the possibility of a distinction between the assertive category of speech acts, which as we saw includes a range of speech acts, and other categories suggested by the mood of the verb of the sentence uttered in the performance of a speech act. On the assumption that identifying the category of assertive speech acts amounts to delimiting the range of utterances that are susceptible of truth-evaluation, the criterion I examine here is that the verb of the sentence being in a certain mood, namely the indicative mood, is the mark of the truth-evaluable utterances.

The contribution of the first part of this chapter to the general project discussed in the dissertation is to eliminate a possibility of drawing the required distinction between truth-evaluable/doxastic and non-truth-evaluable/non-doxastic utterances/attitudes of speakers/agents. The possibility that is eliminated here is one of answering the question of a distinction between truth-evaluable and non-truth-evaluable utterances in the context of a radical interpretation theory by use of grammatical features of the sentence uttered. The considerations I put forward at this part of the section intend to show that a criterion for the required distinction based on the mood of the sentence is inadequate. One kind of problem that the suggestion that mood signifies the required distinction faces is that application of this criterion misclassifies utterances. This sort of problem, among others, has led Davidson to reject the view that mood signifies force.

A second objection that I will present against an adoption of the criterion is a more specific one. It is based on the observation that the criterion is not sufficient for the required classification in a theory of understanding. There is a question of what pre-theoretic distinctions the variety of mood exemplifies. An answer to this question, that would point to the significance of mood, is required in order to test whether the criterion of mood is adequate for the distinction in question between truth-evaluable and non-truth-evaluable utterances. The suggestion of the use of a criterion of mood in radical interpretation theory requires a justification of the choice of this criterion; that is, we have to give some kind of reasons for why, for instance, we think that use of the indicative mood could result in an attribution of a holding true attitude or in a doxastic attitude. The mere observation of a systematic connection between a certain mood and a certain category of speech act requires an independent of mood

classification of those speech acts. Unless the criterion of mood is supplemented by an account of the classifications it is used to signify, it leaves us blind as to what are the distinctions to which it corresponds.

In the second part of the chapter I will present a general discussion about the phenomenon of mood and about problems accompanying a semantic treatment of non-indicatives. I will critically present and examine some attempts to account for the phenomenon of mood in the context of truth-conditional semantics and I will show the problems that these attempts face. An account of the phenomenon of mood in natural languages has some significance in the context of this dissertation. The variety of mood observed in natural languages and the corresponding modifications in the evaluation of sentences that, from a common sense point of view, it accompanies, - indicatives are true or false, imperatives are obeyed or disobeyed- gives us *prima facie* reasons for denying that all utterances of a natural language are susceptible of truth-evaluation. In addition, though mood, as it will be argued here, is not a criterion of force or of the distinction between truth-evaluable and non-truth-evaluable utterances, cases of use of an indicative sentence typically exemplify susceptibility of a truth-value, while cases of use of an imperative typically exemplify lack of this kind of susceptibility. I will try to show that there are problems with attempts to accommodate non-indicative mood to truth-conditional semantics. The discussion will involve the consideration that indicatives and non-indicatives share a truth-evaluable core. The assumption that indicatives and non-indicatives can have a common truth-evaluable core will be challenged here on the ground that it requires the distinction between truth-evaluable and non-truth-evaluable utterances.

The conclusion of this chapter will be that though mood is not an adequate criterion for a distinction between truth-evaluable and non truth-evaluable utterances, and as such the possibility of drawing the required distinction on the basis of the variety of mood should be eliminated, it is nevertheless a non-accidental feature of natural languages. The distinction between indicative and non-indicative mood is a linguistic symptom rather than a criterion on which a deeper and substantial distinction in the ways humans relate to the world could be based.

3.2 The Criterion of mood

I start, then, by examining the criterion of mood, that is a criterion for a classification of utterances into types based on a grammatical feature of the sentence uttered. The grammatical feature of sentences considered here is the mood of the verb of the sentence or simply the mood of the sentence. The criterion would be the following; a classification of utterances into truth-evaluable and non-truth-evaluable ones is made according to whether the verb of the sentence is in the indicative mood. Utterances of sentences in the indicative mood, the criterion suggests, are utterances susceptible of truth-evaluation. Utterances of sentences in a non-indicative mood are utterances that are not truth-evaluable. The moods other than the indicative are many. Instances are the imperative mood, the optative mood, and the subjunctive mood. In some grammar texts the interrogative construction is classified as a mood, the interrogative mood, though the verb of an interrogative belongs grammatically to the indicative mood.

Use of the indicative mood, according to the criterion, is an indicator that the utterance should be identified as a speech act, the particular kind of which is left unspecified by the criterion, but as one that belongs to the assertive category of speech acts. In addition the utterance should be classified as one susceptible of truth-evaluation. Identification of the assertive category of speech acts is based on the mood of the sentence uttered in the performance of the act. So, utterances of the sentences “It is raining.”, “The bull is raging.”, “The temperature in Edinburgh today will vary between 5 and 11.” are identified as belonging to the category of assertive speech acts, though the first of them might be uttered with the force of an assertion, the second one with the force of a warning and the third one with the force of a report. The indicative mood of a sentence identifies its utterances as belonging to the assertive category of speech acts and, in addition, it classifies them as utterances susceptible of truth-evaluation. If, then, the criterion is judged to be a good criterion, it will be a criterion for drawing the right sort of distinction.

A criterion similar to this one is suggested by Dummett (1973). Dummett’s view is similar to the one suggested above but I think more vulnerable to criticism. He considers a narrower application of the criterion of mood, according to which mood is a conventional indicator of particular forces and not an indicator of force-families.

This view is more vulnerable to criticism than the one suggested by the criterion above because it suggests that there are conventional connections between various moods and particular forces; it takes indicative mood to be conventionally indicating assertive force, imperative mood to be conventionally indicating that a command is issued, interrogative mood to be conventionally indicating that a question is asked.

One point of criticism of Dummett's view concerns the suggestion that mood relates to particular forces. Another point concerns the kind of the connection between mood and force, i.e. it is conventional. Concerning the first point, the weakness of the view is shown by considering that the indicative mood, as a matter of fact, does not indicate a particular force; indicative mood is not employed only for assertions. What P. Geach (1965) calls the *Frege point* might be relevant here. The *Frege point* as presented by Geach attributes to Frege the view that no feature of a proposition (in the sense of a grammatical sentence) assigns to the proposition assertoric force. A proposition, - in this context a proposition may be taken to be an indicative sentence-, might occur in one context asserted and in another context unasserted and be identified as the same proposition. Occurrence of an indicative sentence might carry assertoric force or it might not, as in the case of indicative conditionals. The occurrence of the sentence "It is raining." in the context of the indicative conditional "If it is raining, then the ground is wet." does not indicate that the utterance of that sentence, in this context, carries assertoric force²¹.

We could revise Dummett's criterion so that it might escape the objection presented by the *Frege point*. This could be done by locating the indicative mood in the sentence as an indicative construction, and not in the mood of the verb. However, the criterion we proposed above for a classification of truth-evaluable utterances, does not link the mood of the verb of the sentence with a particular force of utterance of the sentence. It links mood with a whole category of speech acts. Use of the indicative mood is not a sign that an assertion is made; it is a sign that a speech act belonging to a certain category is performed. The suggestion is that utterances of

²¹ The *Frege point* is, in the literature, unfairly attributed to Frege. Frege denied that assertoric force, the assertion sign, can be located in the sentence and he did distinguish between the sense of a sentence, the thought expressed by it, and the recognition of its truth-value. But from Frege's writings it does not follow that a proposition, a Fregean thought, can occur asserted or commanded and be identified as the same thought. If the *Frege point* is taken to imply this, it should not be attributed to Frege. The discussion about Frege's view on sense of a sentence made in chapter 2 (2.3) is relevant here.

sentences in the indicative mood are utterances susceptible of truth-evaluation. A criticism of this criterion would be fair if it points to cases in which the indicative mood is used in speech acts not belonging to the category we might decide to call the assertive category of speech acts and which includes the truth-evaluable utterances. Here we will argue that both Dummett's revised criterion and the criterion for the truth-evaluable distinction are deficient. The argument will be based on Davidson's considerations. Davidson's objections have as a target that the kind of link in a criterion like Dummett's is conventional. But here we will consider a generalised version of Davidson's objections. We will argue that there is a deficiency in both Dummett's and the truth-evaluable distinction criteria independently of whether the link is conventional or not. We will point out that even if we consider the link not to be conventional the suggested criteria are deficient.

Let us present Davidson's criticism of Dummett's view. Part of his criticism applies to a criterion of mood for classifying truth-evaluable utterances. We consider here Davidson's criticism of the mood/force connection criterion as well as a possible similar criticism of the criterion of mood for a classification of truth-evaluable utterances.

Davidson considers some counterexamples to mood as a criterion of force. His counterexamples to a criterion that classifies force on the basis of mood include cases of utterances of indicative sentences with an interrogative force, cases of interrogative sentences uttered as assertions, cases of imperatives used as requests. We can think of many cases in which application of the criterion of mood would misclassify the force of the utterance. Davidson's examples are the following: "Did you notice that Joan is wearing her purple hat again?" and "Notice that Joan is wearing her purple hat again." which might be issued with assertoric force. These are occasions of utterance of these sentences that are misclassified by the mood criterion. They are occasions on which the utterances carry assertoric force and thus, they should be classified as assertions; the criterion of mood results in a different classification of them. It classifies the first one as a question and the second one as an order.

This category of Davidson's objections do not apply only to a criterion along Dummett's line according to which there is a conventional linkage between mood and particular forces. They apply to the criterion suggested above according to which

mood is linked (by a convention or not) to a category of speech acts. We could leave the link unspecified in the case of criterion for the truth-evaluable distinction. All that is required from a criterion of mood for a classification of truth-evaluable utterances is that it enables identification of truth-evaluable class of utterances. The linkage between mood and truth-evaluability does not have to be conventional. A criterion that would exploit some sort of systematic connections existing between mood and truth-value susceptibility would be acceptable as long as it is efficient for the required classification. The examples used by Davidson and discussed in the previous paragraph suggest an inefficiency of the criterion of mood for identifying the truth-evaluable utterances. A sentence in the indicative uttered with the force of a command should be classified as non-truth-evaluable. An occasion of utterance of the indicative sentence “You will not go to the party tonight (!)” uttered in an austere tone by a mother to her teenager son is not meant to elicit the evaluative response ‘this is true’ or ‘this is false’. It is meant to elicit a response of obedience. If, then, we want to accept the common sense intuitions according to which an order, for instance, is not evaluated as being true or false, Davidson’s counterexamples undercut the application of a criterion of mood for classifying truth-evaluable utterances as well.

Another category of counterexamples of a criterion that links mood to force is the examples of utterances uttered by an actor on the stage. The utterance of an indicative sentence used by an actor on the stage does not carry assertoric force. Davidson’s motive for the discussion of the deviant cases is to show that mood does not conventionally signify force. His primary concern is to show that the link is not conventional. If there was a convention according to which, for instance, indicative mood signifies assertive force, that is, that the utterance is issued with the intention to say something true, that the speaker represents himself as believing the sentence, then every potential liar would exploit the convention (Davidson, 1982a). These considerations relate to the second sort of criticism of Dummett’s view. Given that there is a grammatical classification of sentences into categories resulting from the variety of mood and a classification of speech acts into assertions, commands, requests, Davidson’s view is that there is no convention connecting uses of sentences from a grammatical category to performances of speech acts.

There is a point it is worth hinting here. Davidson’s rejection of the mood/force

connection by use of counterexamples is justified only under the assumption that there are, independent of mood, grounds for classifying utterances into assertions, commands, and requests. That is, we should assume that there are independent grounds on which, on a certain occasion, the utterance “Did you notice that Joan is wearing her purple hat again?” should be taken to be an assertion instead of a question as indicated by the mood criterion. Without an assumption like this we cannot reject the criterion as one that misclassifies utterances. This point will ground an additional objection to the ones Davidson presents against a criterion of mood. The point will be discussed more explicitly in the last part of this section. It will be suggested that we need to have some idea of the category of speech acts the grammatical sign of mood demarcates if the criterion is to be of any use. In later chapters of the dissertation I discuss the possibility of identifying the assertive category of speech acts on independent grounds. The ‘independent grounds’ will be the doxastic states of a speaker. At the moment I continue with examining the criterion of mood.

I return to Davidson’s argument against the view that mood conventionally classifies force. It has been objected that the counterexamples Davidson considers do not present strong reasons for rejecting the view that there is a conventional linkage between mood and force. It is suggested that in the cases of the counterexamples stated above another convention overrides the convention that links mood and force (Seumas Miller, 1991, William K. Blackburn, 1987). The convention that links mood to force is cancelled because on certain occasions another convention is in force. For instance the intonation of the voice of the speaker classifies the utterance as a question and it overrides the grammatical convention according to which use of indicative carries assertoric force. In the case of an actor uttering sentences on the stage another convention overrides the conventional linkage between sentences and force of their utterance. The utterances of an actor on the stage are made in a context in which other conventions override the conventional linkage between mood and force. The counterexamples do not show that there is no convention. They show that on certain occasions the convention is overridden by another convention.

But it seems that Davidson’s criticism against the view that there are conventional linkages between grammatical features of sentences and potential uses of

them is more general than the counterexamples show. Davidson's criticism is not restricted only to the view that mood conventionally signifies force; his point of view is that nothing conventionally signifies force.

"There is no known, agreed upon, publicly recognizable convention for making assertions. Or, for that matter, giving orders, asking questions, or making promises." (1982a, p. 270).

Davidson labels his view 'the autonomy of meaning' (1979, 1982a). This is the view that the linguistic meaning of a sentence cannot be derived from the ulterior purpose of its utterance.

Davidson in his discussion of 'the autonomy of meaning' view seems to waver between two views. In certain contexts (1982a, p. 274) he seems to blur a distinction between illocutionary and perlocutionary aspects of an utterance. He rejects the criterion of mood, or a criterion that uses other surface features of the utterance, because of 'the autonomy of meaning'. His view is that the ulterior purpose of a speaker in using a sentence cannot constitute an aspect of the utterance that can be grasped conventionally by the linguistic meaning of the utterance. I take this to be a fair and straightforward suggestion if 'ulterior purpose', as it turns up in his account of 'the autonomy of meaning' view, is treated as a perlocutionary aspect of an utterance. It might be the case that the ulterior purpose of my utterance "It is a nice day today." is my intention to convince my hearer to accompany me in a walk in the park. In this case it is almost straightforward that not only no convention could link the linguistic meaning of my utterance with the ulterior intention of mine in uttering it but hardly any systematic connection exists between the sentence "It is a nice day today." and the hearer walking with me in the park. But could we, in a similar manner, deny that there is any systematic connection between the utterance "It is a nice day today." and a classification of it as saying something held true and not wished to be true, or saying something that is truth-evaluable? It seems that there is a difference in the features of the utterance identified in the two cases. It might be that the distinction between illocutionary and perlocutionary aspects of a speech act, suggested by Austin, is introduced to distinguish between features of an utterance that can be conventionally or systematically associated with the utterance type and features of the utterance that are not susceptible of any systematic or conventional classification. A similar suggestion is made by Dummett who calls the aspect of an utterance like the one above that relates to the ulterior purpose of the speaker, of

making the hearer accompany the speaker in a walk in the park, the 'point' of the utterance and he suggests that it should be distinguished from the force of the utterance which is assertive (1993, pp. 202-223).

In other contexts Davidson seems to admit that mood relates to force. It seems that Davidson's objection is against the view according to which conventional links are established between mood and force. He does not deny that there are any systematic connections between mood and force; but he argues that the connections cannot be conventional.

"It is easy to confuse two quite different theses : on the one hand, the (correct) thesis that every utterance in the imperative labels itself (truly or falsely) an order, and the thesis that there is a convention that under 'standard' conditions the utterance of an imperative is an order. The first thesis does, while the second thesis can't, explain the difference in meaning between an imperative and a declarative sentence, a difference which exists quite independently of illocutionary force." (1982a, p. 275)

The criterion of mood suggests that there is a variety of grammatical categories that correspond to a variety of attitudes or categories of speech acts. We have some pre-theoretic intuitions about the kind of attitudes speakers might have toward sentences, the kind of attitudes speakers express in uttering sentences or about the variety in categories of speech acts. For instance, we have the following pre-theoretic intuitions about assertions: assertions are truth-evaluable speech acts, they are issued with an intention to say something true, the speaker usually believes what he or she asserts. Our pre-theoretic intuitions about exhortations are different; exhortations are not truth-evaluable, they are satisfied or obeyed, they concern the performance of an action and usually they express the speaker's desire for the performance of an action. It is on the basis of pre-theoretic intuitions like those that we judge whether a criterion for demarcating the distinctions is adequate or not. If we accept this claim, an interpretation of the criterion of mood that is straightforwardly rejected is one that would identify assertions with the uttering of an indicative sentence or exhortations with the uttering of an imperative sentence²². An assertion or an exhortation does not consist in the uttering of an indicative or an imperative sentence. Evidence for this is that an assertion is possible by use of a sentence in mood other than the indicative.

What these considerations suggest is that the criterion of mood alone is not

adequate to identify the holding true attitude or the truth-evaluable attitudes. For a criterion of mood -either one that holds the connection between mood and kind of force to be conventional or one that loosely appeals to systematic connections between the two- to be in force it is required an independent of mood classification of speech acts as of a certain kind. The kind of speech act that a certain mood might be used to demarcate. Since it is not use of a certain mood that constitutes the performance of a certain kind of speech act, i.e. an assertion or the uttering of a truth-evaluable utterance does not consist in the employment of indicative mood, an independent characterisation of the class of utterances that mood might be used to demarcate is required.

The point discussed here is not just against a conventional connection between mood and force. Even if we are presented with good reasons for the assumption that the connection between mood and force is conventional, as long as the convention is not constitutive for the performance of a certain kind of speech act, an identification of the kind of act independently of mood is required. The possibility of misclassifications of utterances by use of the criterion of mood eliminates the possibility of interpreting the link between mood and force as constitutive. If use of a certain mood were constitutive of the performance of a speech act of a certain kind, every use of a token of an indicative sentence would be the performance of a truth-evaluable speech act. Exceptions to this rule -the cases presented by Davidson- suggest that the link between mood and force is not constitutive; uttering a truth-evaluable sentence *is not* employing an indicative sentence.

The criticism of a criterion of mood shows it to be inadequate with respect to two points. One point of the criterion's insufficiency was shown in the discussion of the misclassifications a use of the criterion effects. Another point with respect to which the criterion shows itself insufficient, a point different and additional to Davidson's criticism, is that the criterion needs to be supplemented by an account of the classifications it is taken to demarcate.

What has been argued in this part of this chapter is that a criterion for

²² An interpretation like this would treat utterances of all sentences more or less as performatives. According to this interpretation uttering an indicative is making an assertion as the utterance of "I promise such-and-such" *is* to promise such-and-such.

distinguishing between truth-evaluable and non-truth-evaluable utterances based on the mood of the verb of the sentence is inadequate. It misclassifies utterances. In addition the criterion, without being supplemented by an account of what the criterion is supposed to demarcate, is not useful.

3.3 The Phenomenon of mood and Truth-conditional semantics

Our primary aim is a classification of utterances into truth-evaluable and non-truth-evaluable ones in the context of a theory of interpretation. In the above I considered the possibility of making that classification on the basis of the mood of the verb of the sentence uttered or other surface features of it. The arguments presented against the possibility of a criterion for the required classification based on the mood of the verb were of two kinds. The first pointed to the inadequacy of a criterion of mood by presenting counterexamples to its application. The second argument pointed to the insufficiency of the criterion for a classification of utterances in a radical interpretation theory.

Despite these considerations the phenomenon of mood, with the modification of the evaluation of sentences that it often introduces, provides a *prima facie* reason for granting one of the basic assumptions of this dissertation; the assumption is that not all utterances are susceptible of truth-evaluation. One of the *prima facie* reasons for holding this assumption is that the common use of language suggests that indicatives are evaluated as being true or false while, for instance, imperatives are not susceptible of truth-evaluations.

In the following part of the chapter I intend to discuss attempts to deal with the phenomenon of mood in truth-conditional semantics. The phenomenon of mood, and in particular sentences in non-indicative mood, presents a puzzle to truth-theoretic semantics. I will show that the attempt to deal with mood in a way that deprives non-indicatives of their feature of not being susceptible of truth evaluation faces problems. This is the attempt made by the explicit performative treatment of mood. In addition, I will discuss two attempts to embed mood in truth-conditional semantics by distinguishing the mood aspect of a sentence and a mood-element of the sentence. These are Davidson's paratactic analysis of mood and Stenius's proposal for an analysis of mood. Both of these views will be critically examined.

3.3.1 Non-indicatives as explicit performatives

I start by considering an attempt to embody the phenomenon of mood within truth-conditional semantics. The theory I consider first, called the explicit performative account of mood, suggests a kind of reduction of non-indicatives to indicatives. This theory opposes the common sense view that non-indicatives lack truth-value. A non-indicative sentence, according to this proposal, is taken to be equivalent in meaning to a sentence in the indicative mood. Non-indicatives are taken to be equivalent to indicatives of a certain kind, namely the explicit performatives. A semantic analysis in terms of truth-conditions of the explicit performative, which is a sentence in the indicative mood and which is taken to paraphrase the non-indicative, amounts to a semantic account of the non-indicative.

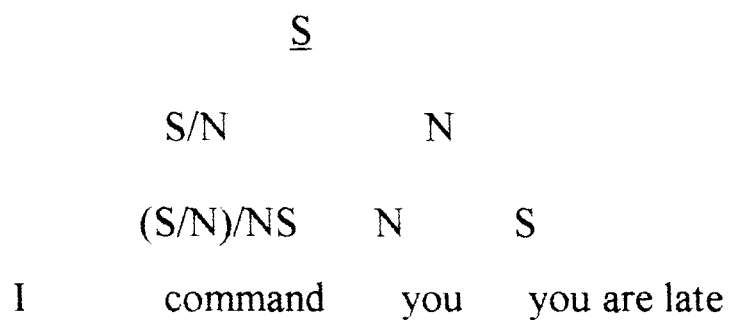
The view that non-indicatives can be reduced to indicatives is often presented in the literature as the possibility of a reduction of non-declaratives, that is commands, questions, to declarative sentences. In this context I take the two kinds of reduction, non-indicatives to indicatives and non-declaratives to declaratives, to be parallel. The considerations and criticisms presented here treat indicatives and declaratives as belonging to the same category of sentences.

The view that non-indicatives or non-declaratives should be treated as explicit performatives is put forward by generative semantics linguists like J.R. Ross (1970) and Lakoff (1972). D. Lewis (1970) endorses the view as well.

The suggestion is that, for instance, the non-declarative “Be late!” should be treated as a paraphrase of the explicit performative “I command you to be late.” and the non-declarative “Are you late?” should be treated as a paraphrase of “I ask you whether you are late.”. The semantic treatment of the non-indicatives “Be late!” and “Are you late?” is given by the semantic treatment of the indicatives “I command you to be late.” and “I ask you whether you are late.”²³. The explicit performatives “I command you to be late.” and “I ask you whether you are late.” are treated semantically like any declarative sentence. For instance, in D. Lewis account of

²³David Lewis (1970) suggests that a semantic treatment of declarative sentences be given via a categorically based transformational grammar with intensional semantic functions. This is an aspect of his semantic theory not of primary importance for the explicit performative reduction of the non-declaratives. An explicit performative treatment of non-declaratives could be accommodated by a Tarski-style truth theory for declaratives.

semantic treatment of declaratives on the basis of a categorically based transformational grammar the structure of the explicit performative “I command you to be late.” would be given by the following:



Though there is agreement among proponents of the view about the deep structure of non-declaratives, there is disagreement on the truth-evaluation of the explicit performative. So, Lakoff (1972) takes it to be the propositional content S that is to be true or false. This suggestion is acceptable in an analysis of a declarative sentence. For instance, an explicit performative analysis of the declarative sentence “It is raining.” would be “I state that it is raining.”. What is truth related is the propositional content ‘it is raining’. The sentence “It is raining.”, equivalent to “I state it is raining.”, would be true or false according to whether the propositional content is true or false.

But the suggestion has the unacceptable consequence that in the cases of non-declarative sentences it assigns irrelevant truth conditions. For example the suggestion for the case of the non-declarative “Close the door!” would be that it is equivalent in meaning to “I command you to close the door.”. The explicit performative “I command you to close the door.” is true or false according to whether the propositional content ‘you close the door’ is true or false. But this is not acceptable. The T-sentence it suggests is the following:

T: “I command you to close the door” is true if and only if you close the door.”

The semantic value of the explicit performative as specified by the above T-sentence cannot be accepted as equivalent to the original non-declarative “close the door!”. The ‘T-sentence’ does not give the truth conditions of “I command you to close the door.”.

D. Lewis’s suggestion is different in respect to the truth-evaluation of the

explicit performative. His suggestion is that the T-sentence specifying the semantic value of the explicit performative would be, for the example considered here, the following:

“I command you to close the door.” is true if and only if I command you to close the door.

The problem with this suggestion is that it does not apply in a uniform way to declaratives. The utterance “It is raining.” will be equivalent to “I assert that it is raining.” which is true if I do assert that it is raining even if it is actually false that it is raining.

But these are not the truth conditions we expect a theory of meaning to assign the declarative “It is raining.”. So ‘It is raining.’ is not equivalent to the explicit performative ‘I assert that it is raining.’.

Another thing to notice in the explicit performative account is that the explicit performative sentence carries an ambiguity not present in the non-indicative sentence. For instance, the explicit performative ‘I command you to be late.’ might amount to the issuing of a command and in this sense it is equivalent to the non-indicative ‘Be late!’. But it might be used as a description of what the speaker is doing, that is commanding. In this case it is not a paraphrase of the non-indicative ‘Be late!’. In uttering the non-indicative “Be late!” there is no the ambiguity of this kind. The utterance ‘Be late!’ cannot be used by the speaker as a statement of his performing a speech act. We will then have to modify the paraphrasis so that it is given by a less explicit performative term like ‘to tell’. “Be late!” would, then, be paraphrased as “I tell you to be late.”. But in this case we encounter the problems discussed in Lacoff’s and Lewis accounts.

3.3.2 An account of mood appealing to two notions of truth

A suggestion often present in the philosophical papers that offer an analyses of mood would be one that suggests a distinction in every indicative and non-indicative sentence between an indicative core, descriptive content, propositional content and a modal element or mood setter. We find this suggestion in Stenius (1967) and Davidson (1979). Both Stenius’ and Davidson’s proposals deny truth-evaluability to

non-indicative sentences and they attempt to offer a semantic analysis of sentences in various moods. In this section I consider Stenius' proposal and in the following section I continue with Davidson's paratactic analysis of mood.

Stenius considers the three following sentences which are in various moods: a) "You live here now." b) "Live you here now!" and c) "Do you live here now?". He suggests we call sentence radical what is common between them and modal element of each of those sentences what is different in each of them. The common element between the three sentences, what is here called sentence radical, and their different modal element can be shown more explicitly if we consider the following rephrasing of the above mentioned sentences: a') 'It is the case that you live here now', b') 'Let it be the case that you live here now!' and c') 'Is it the case that you live here now?'. If the sentence radical is specified by the that-clause, the rephrasing shows the three sentences to have a common element, since they contain the same that-clause. In addition the rephrasing shows a difference between the three sentences. Stenius suggests we call the difference modal.

The distinction between sentence radical and modal element suggested here conforms to a theory of meaning that is in certain respects equivalent to a picture theory of meaning. The sentence radical is like a picture and the modal element of a sentence specifies possible functions of the picture. The showing of a picture may serve various purposes. It may serve the purpose of showing to a viewer how things are, it may serve the purpose of showing to the viewer how things should be, it may serve the purpose of eliciting to the viewer a wonder whether things are as shown in the picture. We could think of more purposes the showing of a picture might serve. The suggestion here is that the sentence radical is like a picture and the modal element specifies various purposes the showing of a picture might serve. In the example considered above the sentence radical could be related to the picture that, difficulties with indexicals aside, would show you living here now. The modal element of the sentence "You live here now." would be related to the possible function of a showing of the picture of claiming that you live here now. The modal element of the sentence "Live you here now!" would be related to a possible purpose served by a showing of the picture of exhorting the hearer to be living here now. The modal element of the sentence "Do you live here now?" would be related to a possible purpose of showing

the picture of making the hearer interrogate about his place in relation to the place of the speaker at the time of the utterance.

With these background assumptions about the distinction in question we can inquire further about a semantic analysis of sentences in various moods. The distinction between sentence radical and modal element as suggested above allows for the possibility of sentences in different moods sharing the same descriptive content. The difference between these sentences lies, according to this assumption, on an extra-propositional element, an extra-sentence-radical element that amounts to a modification of the sentence radical. The semantic rules of this extra-propositional and extra-sentence-radical element need to be specified for a complete semantic analysis of sentences.

Let us call mood operator what accounts for the modification of the sentence radical in cases like the three sentences above. So, the three sentences, that according to the assumption here share the same sentence radical, can be represented as follows: a)p. b)p! and c)p?, where p stands for the common sentence radical, while the fullstop, the exclamation mark and question mark stand for the mood operator. A first thing to observe is that the mood-operator cannot function as a truth functional sentential operator like conjunction or disjunction since a sentence in the imperative mood does not have a truth-value and it is questionable whether a sentence in the interrogative mood has a truth-value. On the assumption that a sentence radical can be identified across sentences in various moods and, in addition it is a truth-bearer, it becomes a problem to show and explain how the mood operator functions semantically on the truth-evaluable sentence radical to produce a sentence that is not truth-evaluable. This problem is related to the question of what might be involved in understanding non-truth-evaluable sentences.

Stenius proposes the following answer to the question. One thing that is granted is that a sentence radical is a truth bearer; a sentence radical is 'true' when what is described by it really is the case. Given this assumption, he suggests the following rules specify the meaning/use of sentences. The rule for indicatives is 'produce a sentence in the indicative mood only if its sentence radical is true'. The rule for imperatives is 'react to a sentence in the imperative mood by making its sentence radical true'. There is room for discussion as to whether the above stated rules are

‘semantic’ rules. Does the indicative mood operator fail in its function if the sentence radical is not true? Does it, in this case, fail to convey a claim that the sentence radical is true and fail to make the sentence be understood as conveying this claim? Let us not dwell on this question here and use the term ‘semantic’ in a loose sense.

The point on which I intend to dwell on concerning Stenius’ proposal is that the suggested treatment of indicatives and non-indicatives requires the introduction of two concepts of truth. One concept of truth applies to the sentence radical. He calls it the descriptive truth. Another concept of truth applies to a sentence in the indicative. He calls it modal truth. So, as he puts it, the expression ‘speaking the truth’ is related to the modal concept of truth and means that one is following the above rule correctly. For one to follow the above rule correctly is to produce a sentence in the indicative only if the sentence radical is true, thus, an indicative sentence is true only if its sentence radical is true. Considering only the case of indicatives suggests that there is a generic concept of truth. If a sentence radical is a truth-bearer then it is semantically equivalent to the indicative sentence with it as a sentence radical. The truth-value of a sentence radical coincides with the truth-value (modal in this case) of the indicative sentence which has it as its radical.

The case is different for sentences in a mood different from the indicative. Stenius suggests that there are two concepts of truth because when the sentence radical of an imperative or of an interrogative is true, one does not say that the whole sentence is true. This is taken to suggest that in addition to the descriptive truth which attaches in a uniform way to sentence radicals of all sentences there is another concept of truth that relates only to the indicative sentences. So, according to the suggestion, an imperative or interrogative with a true sentence radical is a descriptively true sentence but it is not a modally true one. Modal truth is a value only indicatives are apt to enjoy.

The question here concerns the concept of truth that is suggested to be playing a central role in radical interpretation theory. Considering the suggestion made here that there is a distinction between descriptive truth and modal truth the question that arises is which one of the two concepts of truth is the one suggested to be central to interpretation. The attitude of holding a sentence true as long as it is the bridge for meaning and belief is related to a modal concept of truth. The attitude of holding a

sentence true is related to an analysis of indicatives; the utterance of a sentence in another mood, say the utterance of an imperative, is not associated with the attitude of holding a sentence true.

If we take into account that it is the attitudes as expressed in complete sentences that get meaning off the ground there seems to be a problem with identification of the descriptive content of sentences in a mood other than the indicative. It is just because in the case of indicatives modal truth and descriptive truth coincide that we can pass from an identification of the holding true attitude towards a sentence to interpretation of its descriptive content. While we are shown how to approach the descriptive content of an indicative, (according to the rule above if an indicative is modally true then it is descriptively true) we are left blind as to how we could identify the descriptive content of an imperative. We are not given access to the descriptive truth of imperatives.

A possible answer to this problem, at least as far as imperatives are concerned, could be given by considering the relation between obedience and descriptive truth of an imperative. The answer would be along the following line. If an imperative is obeyed then it is descriptively true. The suggestion deserves a more thorough investigation. What I will comment here is that use of this access to the descriptive content of an imperative presupposes a distinction between modally truth-evaluable utterances and obedience-evaluable utterances. The possibility of drawing a distinction between utterances that are modally susceptible of truth-evaluation and utterances that have obedience conditions could be seen as one of the subjects of this dissertation. Unless a possibility of distinguishing between the class of modally truth-evaluable and modally non-truth-evaluable sentences is given, the suggested access to the descriptive content of modally non-truth-evaluable utterances cannot be used. In this case the distinction between modally truth-evaluable and modally non-truth-evaluable utterances would be required for identification of a common sentence-radical/descriptive content across the two categories. Unless an account of the required distinction is provided on which a common sentence-radical/descriptive content would be identified the assumption implicit in Stenius account that sentences in various moods can share a common radical is not fully justified.

3.3.3 The paratactic analysis of mood

I continue with presenting Davidson's suggestion about a treatment of sentences in various moods. Davidson's suggestion about a treatment of mood is in the spirit of the paratactic analysis that he offers for indirect reports and for propositional attitude sentences. Here I will examine the paratactic analysis as offered for sentences in various moods. The paratactic analysis of indirect speech sentences and of propositional attitude sentences is discussed also in other contexts in this dissertation.

Davidson (1979, pp. 115-116) considers some constraints to which an analysis of mood should conform. A satisfactory theory of mood should a) account for the common element between, for instance, the sentences "You put on your hat." and "Put on your hat!", or "Do you put on your hat?". b) It should account for the common element between various sentences in the same mood; for instance, it should account for the common element between the sentences "Put on your hat!" and "Eat your food!". In addition, c) the analysis should be one that can be embedded in truth-conditional semantics.

Davidson proposes the following for a semantic treatment of sentences in non-indicative mood. Sentences in non-indicative mood, that is sentences which do not have a truth-value, should be decomposed into two sentences, one of which refers to the other and each of which has a truth-value (1979). So, the suggestion is that a non-indicative sentence should be considered to be decomposed into an indicative sentence plus a sentence, indicative again, that specifies a transformation of the first indicative sentence. Davidson calls the first sentence the indicative core and the second sentence the mood setter. A semantic treatment of non-indicatives consists in a specification of the truth conditions of the two indicative sentences into which, according to the paratactic analysis, it is decomposed. So, the suggestion is that the imperative sentence "Put on your hat!" should be considered to be decomposed into the two following sentences: "You put on your hat. This is an imperative".

There are many questions concerning the suggested analysis. In what follows here I mention some of the problems that a paratactic analysis of the phenomenon of mood faces. One of the questions that can be asked about the paratactic analysis of sentences in various moods is what kind of transformation the mood setter signifies

and what is the point of talking about its truth conditions.

One possibility is that the transformation that the mood setter signifies be a purely grammatical one. In this case the mood setter is taken to be specifying that the indicative core should be read as being in a certain grammatical mood. For example, according to this possibility, the transformation of the indicative core ‘you will put on your hat’ that the mood setter of the paratactic analysis of “Put on your hat!” signifies, is that the indicative core is to be read as an imperative. It is obvious that this interpretation of the transformation specified by the mood setter makes the mood setter redundant. The construction or the inflection of the verb of the original sentence “Put on your hat!” suggests that the sentence belongs to a certain grammatical category, namely, it is an imperative. There is no point in the suggestion of decomposing the sentence by the paratactic analysis.

Another possibility would be to take the mood setter to be specifying not a grammatical transformation of the indicative core but to be specifying the force of an utterance. It seems that this is the suggestion Davidson makes. There is a point that needs to be clarified here. In some contexts it seems that Davidson takes the mood setter to be specifying the force with which the indicative core is uttered.

“...we can give the semantics of the utterance of an imperative sentence by considering two specifications of truth conditions, the truth conditions of the utterance of an indicative sentence got by transforming the original imperative, and the truth conditions of the mood setter. The mood setter of an utterance of “Put on your hat” is true if and only if the utterance of the indicative core is imperatival in force.” (p. 120).

I discuss in a while what ‘imperatival in force’ might mean. At the moment suffice it to note that in this context the mood setter is taken to specify the force of the utterance of the indicative core.

This suggestion seems to imply that the utterance of the indicative core is not a forceless saying. Some comments Davidson makes in the same article (1979) seem to strengthen this view. He refers to the utterances of the mood setter and of the indicative core as two speech acts. (1979 p. 119) The term ‘speech act’ is rarely used for a forceless saying. The proposal seems to be that the speech act performed by uttering the indicative core is issued with an unspecified force. The speech act performed by uttering the mood setter is asserting that the indicative core was issued with a certain force.

It seems that the motive Davidson has for suggesting a paratactic analysis of mood is that mood should be dissociated from force. And this is a requirement that the paratactic analysis satisfies. The specification of force is given by an indicative sentence, the mood setter. The mood setter might be true or false according to whether the utterance was issued with the force it specifies or not. On the other hand, while, as it was argued in the first section of this chapter, mood does not provide an adequate criterion for force, mood is a feature of sentences. A theory of meaning should account for mood, if mood is considered as a semantic feature of sentences. And a theory of interpretation should account for mood, if mood modifies aspects of the understanding of an utterance. The view that I present here is that the paratactic analysis faces problems if it is considered to be an analysis of mood. It is not straightforward that the suggested analysis satisfies the constraints Davidson considers a theory of mood should be satisfying. The following offer a justification of my view.

A first problem with a paratactic theory of mood is that it needs to be supplemented by an account of the notion of ‘imperative force’ or ‘indicative force’. We saw that the term ‘imperative force’ cannot signify a grammatical category. Another possibility is that it is taken to signify an illocutionary force of an utterance. The suggestion then is that the imperative “put on your hat!” be analysed as “You put on your hat. This is a command”.

But could we accept this analysis as an analysis of mood? Does it show what two imperatives have in common? Interpreting the mood setter in the way suggested, that is, that it is a force specifier, results in the following problem. The imperative sentence “Put on your hat!” issued with the force of a command and the imperative sentence “Eat your food!” issued with the force of a request are not shown by the paratactic analysis to have anything in common. In the one case the analysis is “You put on your hat. This is a command.” and in the other case the analysis is “You eat your food. This is a request.”. This interpretation of the paratactic treatment does not satisfy the second of the constraints that Davidson suggests an analysis of mood should satisfy. The paratactic analysis, interpreted so that the mood setter specifies the force of the utterance, does not account for what is common between sentences in the same mood. An analysis of the phenomenon of mood that shows a common

element between various sentences of the same grammatical category seems to be a straightforward requirement. The paratactic analysis of sentences in various moods discussed above, that is the one according to which the mood setter is a force specifier, does not satisfy this requirement.

There is a possibility left open for an account of the mood setter, that is an account of the feature of the utterance the mood setter should be taken to specify. This is the following. The mood setter specifies that the utterance of the indicative core should be taken to belong to a certain category of utterances. A mood setter of the type “this is indicative in force” proceeded by an utterance of the indicative core classifies that utterance as belonging to a certain category of speech acts. It classifies it as belonging to the category of speech acts that includes assertions, reports, conjectures and other kinds of speech act that are often performed by use of indicative mood. A mood setter of the type “this is imperativ in force” proceeded by an utterance of the indicative core classifies the utterance as belonging to another category of speech acts. This category of speech acts includes commands, requests, exhortations and other kinds of speech act often performed by use of a sentence in the imperative mood.

In order for the second requirement imposed on an analysis of mood to be satisfied, that is the requirement that an analysis of mood should exhibit the common element between sentences in the same mood, the account suggested in the above paragraph is not adequate. The account needs to be supplemented by identity conditions of each of the categories of speech act that mood setters specify. Given the paratactic analysis of sentences in various moods, we agreed, on pain of rendering the mood setter redundant, that the mood setter does not specify a grammatical category. If mood setters are taken to specify categories of speech act or categories of forces then, for the second requirement to be satisfied, the categories are in need of identity conditions. It needs to be shown how the speech acts or forces specified by the mood setter of a paratactic analysis of indicative mood sentences are lumped together. Similarly for the case of imperative sentences it needs to be shown how the speech acts or forces specified by the mood setter of their paratactic analysis are identified as constituting one category of speech acts or forces. If identity conditions for each category of speech acts or forces is not given then the analysis in question does not

satisfy the second requirement imposed on an analysis of mood; it does not show the common element between sentences of the same mood. And, in this case, it can be questioned whether the proposed paratactic analysis of sentences offers an analysis of mood at all.

Despite the problem of supplementing the paratactic analysis, interpreting mood setters in the suggested way, that is as specifiers of categories of utterances, relates to the central problem of the dissertation. The identification of the category of speech acts to which the mood setter of the type “this is indicative in force” relates, that is the category that includes assertions, reports, conjectures and other speech acts often performed with use of indicative, is a requirement in this dissertation. It is the category of speech acts that includes utterances susceptible of truth-evaluation, as contrasted, for instance, to the category of utterances of “imperative in force” utterances which are not susceptible of truth evaluation. As discussed above, the presented paratactic analysis of sentences does not answer the question of an identification of speech acts as belonging to the indicative or to the imperative category. The mood setters of paratactic analyses of sentences name without identifying families of utterances, of forces or categories of speech act. An answer to the question of a classification of utterances into the required sorts is crucial for the tenability of the paratactic account of sentences in various moods.

We should mention here the following. According to Davidson though the indicative core and the mood setter are truth -evaluable, the utterance for which the parataxis of them offers an analysis might not be truth-evaluable. His account does not suggest a reduction of non-truth-evaluable utterances to truth-evaluable ones. I think that even if we accept Davidson’s suggestion for a paratactic treatment of non-truth-evaluable utterances the problem of the identification of the truth-evaluable utterances in a theory of interpretation remains alive. A scheme of evidential support of the theory that applies for the case of a truth-evaluable utterance does not apply for the case of a non-truth-evaluable utterance. A distinction between truth-evaluable and non-truth-evaluable utterances is necessary for the correct application of a scheme of evidential support. The paratactic analysis of sentences in various moods constitutes an attempt to accommodate non-truth-evaluable utterances into truth-evaluable semantics but it does not eliminate the requirement for an identification of the two

categories. In addition, for reasons explained in this section the attempt is not unproblematic. Something like the classification in question between truth-evaluable and non-truth-evaluable utterances might be required for the tenability of the paratactic account.

An additional comment we make here about the paratactic analysis of mood concerns the common content assumption. That is the assumption that indicatives and imperatives can share a common indicative core. The paratactic analysis of mood does not provide the resources for a distinction between mood. The possibility is open that an account of such distinctions would suggest a distinct core across sentences in various moods.

To sum up: Part of our criticism of both Stenious and Davidson's account of mood was that they require the truth-evaluable/non-truth-evaluable distinction. But this does not show the accounts to be untenable if supplemented by an account of the distinction. That is, a distinction between the utterances that are modally truth-evaluable and the ones that are not modally truth-evaluable in the Stenious case, and a distinction between utterances of mood setters that specify indicative force and the ones that specify imperatival or other kind of force in the Davidson case. So, the issue becomes whether an account of the required distinction could be combined with the suggested accounts by Stenious and Davidson. The issue will be given an answer in chapters 6 and 7 where a proposal about the distinction is made. Here we suggest that the account that takes a common truth-evaluable sentence radical or a common indicative core be put on hold until the issue of whether they combine with a proposal of the distinction is clarified.

CHAPTER FOUR

APPEALING TO BELIEF

4.1 Utterances and attitudes

In this chapter I consider a different approach to the possibility of distinguishing the truth-evaluable utterances of a language from the non-truth-evaluable ones. The approach considered here and in the following chapters of this dissertation is an attempt for the distinction on the basis of the attitude/state expressed or ascribed by the utterance²⁴. According to this approach, the systematic connections between utterances and mental attitudes/states of the speaker provide the ground for a possible classification of utterances. It will be on the basis of the kind of attitude to which an utterance of a subject, actual or possible, is related that enables a classification of the utterance into a truth-evaluable or non-truth-evaluable one. According to the present approach, if the attitude had by the speaker who utters the sentence “Did you notice that Jane is wearing her purple hat again?” is that of belief or partial belief, then the utterance is classified as a truth-evaluable one. This would result in a classification of the utterance within the domain of applicability of certain observational schemes that conjure up the principle of charity for an interpretation of the utterance. If the attitude had by the speaker who utters the sentence “You will not go to the party tonight (!)” is that of a desire then the utterance is classified as non-truth-evaluable one. It will not be on the basis of grammatical features of utterances that the required classification is based, but on the basis of attitudes/states speakers have related to utterances.

The discussion in the previous chapter touched upon the nature of the relation between utterances and attitudes. It was discussed whether the kind of systematic connection between utterances and attitudes or mental states should be seen as conventional. We saw that there might be difficulties with the view that the connection in question is conventional. However it was not doubted that there are systematic connections between utterances and kinds of attitudes. In uttering a sentence, a speaker might be expressing a certain attitude/state, he might be

²⁴ The terms mental state and propositional attitude will be used interchangeably in this context. The mental states the considerations here are about are mainly contentful mental states like belief and desire and in the relevant respects they need not be distinguished from propositional attitudes.

representing himself as if he had that attitude or as if he was in that mental state. In an interpretation of an utterance a certain attitude/state might be attributed to the speaker. All these suggestions about the nature of the relation between utterances, actual and potential, and attitudes require the existence of systematic connections between them.

The systematic connections between utterances and attitudes, cashed out in a scheme like ‘utterance *s* expresses state *m* of the speaker’ provide the ground for a possible classification of utterances into truth-evaluable and non-truth-evaluable ones.²⁵ According to this possibility for the required distinction there are states/attitudes systematically connected to truth-evaluable utterances. On the basis of these systematic connections, there is a way in which states/attitudes can be evaluated by the same range of values that utterances are evaluated. Belief states, for instance the belief that it is raining expressed by an utterance “It is raining.”, are typically states that are assessed as true or false. Though it is for the truth aptness of content that we call a belief state true or false, implying that the content of the belief state or the utterance that expresses the belief are true or false, there are states that are not assessed as true or false. So, while the class of truth-evaluable state/attitudes includes conjectures, opinions, suspicions, partial beliefs, there are states like desire, intention, preference or wish that are not assessed as true or false. Desires, preferences and wishes are assessed as satisfied/unsatisfied or as good/bad. My desire to succeed in a University exam expressed by my utterance “If only I succeeded in the University exam!” is not assessed as true or false. Succeeding in the University exam can be assessed as a good thing or a bad thing²⁶. Given then that there are systematic connections between utterances and attitudes, the suggestion here is that they can be exploited for a possible classification of truth-evaluable utterances. They can be used in the following way: utterance *s* is a truth-evaluable utterance if the state of the speaker that it expresses or it is taken to be representing is of a certain kind.

The suggestion stated above points to a certain direction for an answer to the problem of a classification of truth-evaluable utterances. But it does not constitute an answer itself. We do not know which is the category of states or attitudes that when

²⁵ The way the systematic connections between attitudes and utterances are envisaged here exploits the triangle principle Chapter 1.2.

²⁶ The possibility of truth-aptness of the content of an affective state will be considered in chapter 5 in the light

expressed or ascribed by an utterance classify the utterance as a truth-evaluable one. The suggestion that the category of states/attitudes related to truth-evaluable utterances includes beliefs, conjectures, opinions, suspicions does not provide an answer to the problem. The question is how we could classify a state/attitude as belonging to that category related to truth-evaluable utterances. This is the question of how we could classify a state as belonging to the category of which beliefs, conjectures, opinions, suspicions are members. And if we accept the claim that mental states can be evaluated by the range of values that sentences or utterances are evaluated, true/false, good/bad, fulfilled/unfulfilled, then the problem becomes how it is possible to distinguish the truth-evaluable states/attitudes from the truth-evaluable ones.

The question that will be discussed from this point onwards in this dissertation is the following: how is it possible, in the framework of a radical interpretation theory, to distinguish the category of states/ attitudes that are susceptible of truth evaluation from the ones that are not? We ask for a set of sufficient conditions that would enable a distinction between the two classes of states/attitudes; a distinction between the class of states systematically related to truth-evaluable utterances and the class of states that do not have this feature. If sufficient conditions for this distinction between states are granted then we can exploit the systematic connections that are taken to exist between states and utterances to result in a classification of utterances. This would be the required classification of utterances into truth-evaluable ones and non-truth-evaluable ones. The criterion for the distinction of state/attitudes between truth-evaluable and non-truth-evaluable should be such that it can be accommodated by a radical interpretation theory. This restricts the choice of the criterion mainly in one way. That is, that the criterion for the distinction in question should be such that its applicability would not require a prior semantic interpretation of language. It should be possible for the criterion to be applied either simultaneously with or prior to a semantic interpretation of the language. For developing further this line we need an analysis of states/ attitudes. An analysis of the notion of belief and of other state/attitude terms suitable to be accommodated by the present project is discussed in the rest of this chapter.

4.2 Truth-evaluable utterances and belief

The focus of our consideration in the following will be the possibility of a criterion for a classification of the truth evaluable utterances of a language based on types of mental state. Here an analysis of the notion of belief will be considered as a typical case of propositional attitude related to truth-evaluable utterances. The views considered here are generally attempts to classify the truth evaluable utterances of a language as the ones which are belief expressive or as the ones which are, in some way to be specified, belief related. That is, they are views according to which the truth evaluability of utterances is a characteristic of them exclusively relevant to the doxastic dimension of the utterer's mind. Acceptance or rejection of a criterion of this kind would depend on the way we conceive the notion of belief. Appealing to the philosophy of mind for an answer and considering theories of belief introduces us to a plurality of problems about belief. Since my original problem is a certain distinction among mental states I will not enter into detailed considerations about the nature of mental states and into the relation of the mental to the physical. The particular aspect of belief and other mental states that I intend to elucidate concerns the characteristic of some of them of being candidates for truth-value susceptibility.

Before discussing certain theories of belief I wish to make a few more comments on the constraints that such an approach to the problem of belief puts upon the theories under consideration. The first and most obvious one is that we cannot construe belief in a way that makes it depend on truth²⁷. What we originally seek is a way of characterising and classifying those utterances of a natural language which admit of truth-value. A theory of belief which could be proved useful for this particular inquiry should not rely essentially on the notion of truth. In this way we rule out as irrelevant and non-useful controversies over the equivalence between 'I believe p' and 'I believe p to be true' or emphases given to truth and falsity being characteristics of belief.

The second constraint put upon the theories of belief that we will consider is

²⁷ The objection that beliefs are the primary bearers of truth and sentences are truth-bearers in a secondary sense does not call off my project. In a way I espouse this view. The central question of the dissertation then becomes how is it possible to identify the belief kind of mental state.

that they should not make essential use of the relation between the notion of belief and the linguistic act of assertion. The assertive utterances of a natural language either constitute a subclass of the truth evaluable ones, or coincide with the truth evaluable utterances of the language. In the case that the characteristics 'truth evaluable' and 'assertible' delimit the same class of utterances the question of a criterion for classifying truth-evaluable utterances is the same question as that of a criterion for a classification of assertible utterances. Thus the statement that the utterances of a natural language which admit of truth evaluation are the utterances uttered in the linguistic act of assertion does not move us from the original problem. The question becomes how we could characterise and classify the assertive utterances of a natural language.

The other case is that the assertive utterances of a language constitute a subclass of the truth evaluable ones with other members being, for instance, utterances uttered by an actor on the stage, sentences occurring in logic texts that though they can be characterised as true or false they might not occur in order to be asserted; utterances that express conjectures, opinions or suspicions should be considered as utterances which without being assertions, are susceptible of truth-evaluation. It might be suggested that an account of assertion or a characterisation of the assertive utterances of a language could be given only within the class of truth-evaluable utterances. That is, the assertive utterances are a subclass of truth-evaluable utterances and so without a prior delimitation of this class, the characteristic 'assertive' would be ill determined.

However, this point needs further clarification. Here is a possibility I do not endorse. It might be suggested that we characterise the belief type of mental state as the type of mental state in which a person who has attained mastery of a language is in or is representing himself to be, in uttering an assertive utterance. At this point I leave aside cases of beliefs we have but which nevertheless we do not express linguistically. I assume for the sake of the argument that we get a grasp of the notion of belief and we acquire skill in attributing beliefs to others through our capacity for identifying assertive utterances. I assume further, for the sake of the argument, that the attribution of non-linguistically expressed beliefs to others is an aspect of our attribution of rationality to them. So we might envisage a theory which characterises

believing as a mental phenomenon identified through an identification of assertive utterances. This idea underlies some theories of belief. It does not turn up explicitly in theories of belief but it is used implicitly when belief is construed as a disposition to respond affirmatively to sentences.

It might seem that my objection to these theories of belief and my reason for introducing as a constraint upon a theory of belief that it should not rely on the notion of assertion is the following. Assertive utterances of a language are truth evaluable and so a characterisation of an utterance as an assertive one presupposes a characterisation of it as truth evaluable. Nevertheless my objection is slightly different. Proponents of such theories of belief could overcome the above objection by denying that such a preclassification is necessary. This objection would reject any requirement for a classification of utterances as truth evaluable ones prior to a classification of them as assertive ones by considering the class of truth evaluable utterances to be the union of some classes of utterances one of which is the assertive class. But this reply introduces a fair point only if it is possible to have access to the class of assertive utterances independently of the class of truth evaluable utterances. It requires a way for identifying assertive utterances independently of the truth-evaluable ones. The only way I can see this requirement to be fulfilled without relying on the notion of truth is by using grammatical features of the sentences to delimit the class of assertive utterances. As I argued in the previous section grammatical features of the sentences do not offer a good classification of utterances. I do not know of any other way of delimiting the class of assertive utterances that does not rely either on the notion of truth or on the notion of belief. So, my reason for not endorsing the suggestion is that belief be identified on the basis of assertive utterances is that I do not see an adequate way of identifying assertive utterances that does not rely itself on truth. So, either we end up with a circularity problem, or with no available means for the classification in question.

There is a third constraint on a theory of belief, a constraint that pervades the whole inquiry, namely that a theory of belief suitable for helping us with the problem of truth evaluation should be suitable to be accommodated by a radical interpretation theory. In chapter 1 of the dissertation we discussed the place of the holding true attitude toward sentences in a theory of interpretation like Davidson's early theory.

We saw how the holding true attitude is related to meaning and belief. In addition we discussed the importance of the possibility of identification of evidential relations between sentences in Davidson's later theory of interpretation. The evidential relations between sentences are accountable by a theory of partial belief for the subject. It was pointed out that identification of an utterance as belief expressing or as one standing to evidential relations to other utterances suggests applicability of the principle of charity. This is the principle according to which the speaker is attributed true beliefs according to the interpreter's standards or speaker and interpreter share their beliefs. An extension of the same principle suggests that the speaker's theoretical beliefs be supported by evidential relations similar to the ones the interpreter's theoretical beliefs are supported. A theory of belief that would be proved useful here should result in identification of belief as an attitude expressed in utterances in a way either prior or simultaneous with a semantic interpretation of the language. A theory of belief suitable to be accommodated in the particular inquiry considered here, that is the inquiry of a classification of truth-evaluable utterances in a radical interpretation theory, should not take for granted semantic aspects of utterances. Delimiting the class of utterances that are susceptible of truth evaluation is something that should be done at least simultaneously with, if not in advance of, a determination of truth conditions of sentences. As such, a theory of belief and of mental states that would be useful here should not be one that grants semantic aspects of mental states as known.

4.3 Belief as a relation to a sentence

In this section I present and discuss sentential analyses of belief and other propositional attitudes. It is popular, especially among philosophers of language, to construe belief as a two-term relation holding between a speaker and a sentence. This view of belief is put forward as part of an attempt to offer an analysis of belief ascribing sentences. The relation of believing as a relation holding between a person and a sentence has been characterised in various ways. Carnap (1947 and 1956) suggests that it is the relation of being disposed to respond affirmatively. Such a view of belief can be helpful in solving some problems about the analysis of belief ascribing sentences; in particular it can cope with problems related to the phenomenon of referential opacity. But the suggestion cannot, at least without further elaboration, prove useful to the present inquiry. I explain my reasons for holding this claim.

The suggestion that belief should be construed as a two-term relation holding between a person and a sentence has been criticised by Church (1950). I intend to leave aside this sort of criticism that is very well known and often repeated in the literature²⁸. I will comment on two points introduced by Carnap's analysis, which are relevant to the present inquiry. The first point I think requires clarification is whether 'sentence', in this context, stands for any sentence of a language, declarative - truth-evaluable- or not. The second point to be clarified concerns the nature of the relation.

I discuss the two points in order. In the case that the belief relation is construed as holding between a subject and a certain category of sentences only, i.e., declaratives or, generally, the truth-evaluable sentences, the suggestion begs the question of the present inquiry. We appeal to the notion of belief in the context of this dissertation in order to use this notion as part of a theory that would classify the declarative or the truth-evaluable sentences/utterances of a language. It was argued in the previous chapter that the category of declarative sentences and the category of truth-evaluable sentences/utterances could not be demarcated on the basis of grammatical features of the sentences. In order, then, for the view that belief is a two-term relation between a subject and a sentence to be accepted by the present context of inquiry, 'sentence' should be construed to range over all sentences of the language of the speaker. If we accept the view that belief is a two-term relation between a subject and a sentence, an account of the relation of believing a sentence would provide a solution to the central problem of this dissertation. The suggestion would be that a sentence to which a subject stands in a belief relation should be classified as a truth-evaluable sentence of the language of the speaker.

The second point that needs to be clarified concerns the possibility of an account of the relation of believing construed as holding between subjects and sentences. The problem here is that subjects and sentences can stand in a variety of relations, not all of which are of the kind of belief. Attitudes that could be construed as sentential and not be classified as beliefs are for instance, desires or wishes. The suggestion in Carnap's theory is that the relation of believing holding between a subject and a sentence is that of being disposed to respond affirmatively.

A similar suggestion is present in Quine's theory of radical translation. In radical

²⁸ One problem of construing belief as being a relation to a sentence is that it becomes impossible to attribute

translation a notion of assent to a sentence is employed. Affirmative response to a sentence, like assenting to a sentence, amounts to something like nodding to a sentence. My claim is that the relation of affirmative responding to a sentence, like the relation of assenting to a sentence and like the relation between a subject and a sentence of holding a sentence true that could be used to identify the relation of believing, are in need of identification themselves. The point is illustrated by Huw Price (1983):

“...I raise a difficulty for recent accounts of radical interpretation as a basis of truth theories, and hence theories of meaning, for arbitrary languages. For the linguistic activity of assent, or holding true, plays a crucial role in such accounts... unless factors of a different kind from speakers’ dispositions to hold sentences true are taken into account, there is liable to be an indeterminacy in the sense/force distinction drawn by such theories of meaning. This will not be the indeterminacy which Quine, Davidson and others have led us to expect in contexts of radical interpretation, but something more basic. For the problem will not be to decide what part speakers’ beliefs, and what part the sense they ascribe to an assertoric sentence, should play in accounting for their willingness to assent to that sentence: but rather to decide whether, with respect to a particular utterance, the categories of assertion and belief are appropriately applied. Quine and Davidson have shown us that to assign truth-conditions we must assign beliefs: but to see the problem in these terms, we must first assign assertoric force. For this decision, I shall argue, evidence as to speakers’ willingness to hold sentences true is far from sufficient.” (H. Price, 1983)

I wish to leave aside some aspects of Price’s claim in the above quotation and keep in consideration the following suggestion of his: that the notion of assent, used in Quine’s theory of radical translation, the relation of holding a sentence true, used in Davidson’s theory of radical interpretation and similarly the notion of affirmative response to a sentence suggested by Carnap’s account of belief ascribing sentences are notions which cannot be taken to be identifiable in advance of a theory of interpretation.

The complaint stated here in relation to Carnap’s suggestion that the relation of believing a sentence amounts to being disposed to respond affirmatively to the sentence is the following. In order for the account to be useful in the context of the present inquiry, we need to be able to distinguish among the various ways subjects respond to sentences the one that is the ‘affirmative’ way of responding to sentences.

With this problem in the background I proceed with discussing Davidson’s analysis of belief ascribing sentences.

Davidson suggests a paratactic analysis of belief and other propositional attitude ascribing sentences (1975). Similarly to the paratactic analysis suggested for

sentences in various moods and for indirect speech report sentences, a paratactic analysis of belief ascribing sentences suggests a decomposition of the sentence into two sentences. So, the belief ascribing sentence

"Galileo believes that the earth moves."

should be considered to be decomposed into two sentences in the following manner:

"Galileo believed that. The earth moves."

The demonstrative 'that' of the first sentence in the above analysis refers to the utterance following the utterance that contains the demonstrative. It is suggested that according to the paratactic analysis there is an ascription of the relation of 'samesaying' between the ascriber and the speaker/ subject. The problems with the paratactic analyses of belief and other propositional attitude ascribing sentences as well as with the paratactic analysis of indirect report sentences are various and have repeatedly been stated in the literature²⁹. I will not recapitulate these problems here. I select to discuss some features of the paratactic analysis of belief and other propositional attitude ascriptions that I consider to be closely connected with the problem of this dissertation.

One problem of a certain version of the paratactic analysis that is, in a way, special to the analysis of belief/propositional attitude ascribing sentences is the following: it presupposes that the subject of the belief relates to the sentence following the demonstrative sentence. But this sentence might be a sentence to which the subject of the belief is totally unrelated. It might be a sentence the subject does not understand and it might be the case that there is no occurrence of the subject's utterance of it. In the case of the example presented above the 'the earth moves' which is that to which the demonstrative refers is a sentence of English. In ascribing the belief that the earth moves to Galileo, the paratactic analysis suggests that Galileo stand in a certain relation to an English utterance made at the time of the belief ascription. But this is unacceptable for the reason that Galileo believed or did not believe that the earth moves independently of the occurrence of any utterance at the time of the present ascription.

Davidson's suggestion for paratactic treatment of belief ascribing sentences

²⁹Most of the problems of a paratactic analysis of indirect report sentences concern a paratactic analysis of belief ascribing sentences as well. See for instance, R. Haack (1971), Simon Blackburn (1975). See also Schiffer (1987), esp. ch. 5 and Loar (1981).

that might overcome the criticism stated above is given by the following:

"...When I say, 'Jones believes that snow is white' I describe Jones's state of mind directly: it is indeed the state of mind someone is in who could honestly assert 'Snow is white' if he spoke English, but that may be a state a languageless creature could also be in." (1975, p. 167)

But construing belief as 'the state of mind someone is in who could honestly assert 'Snow is white'', does not get us very far in the present context. It makes use of a notion of 'honest assertion' whose explication in the present context is one of the aims of the inquiry. A parallel suggestion for the case of a desire attribution to an agent would go along the following:

In uttering the sentence "Jones desires that we go to the movies." we describe directly Jones' state of mind. It is the state of mind someone is in who could sincerely say "Would that we go to the movies!" if he spoke English.

But one reason we appeal to a theory of propositional attitude attributions, in the context of this dissertation, is exactly the problem of accounting for the utterances that are honest assertions. These utterances would be classified as truth-evaluable utterances and would be distinguished from utterances that are, for instance, sincerely commanded and which would be deprived of truth-evaluation³⁰.

The paratactic analyst might suggest another version of paratactic propositional attitude attributions that would overcome the objection presented above. That is the objection that the subject might be totally unrelated to the utterance of the ascriber referred to by the demonstrative. The version of the paratactic analysis that would not be susceptible to this objection takes the relation between the subject and the utterance of the ascriber to be a non-causal relation. The utterance of the ascriber referred to by the demonstrative would express one of the beliefs of the subject. According to this suggestion the paratactic analysis of the propositional attitude report "Galileo believes that the earth moves.", that is the parataxis "Galileo believes that. The earth moves." specifies by use of an utterance of the ascriber a belief of Galileo. The utterance of the ascriber, "The earth moves.", expresses one of the beliefs of Galileo.

There are two comments I intend to make on this suggestion. My first comment is that this version of the paratactic analysis of propositional attitude

³⁰ If we take into consideration Gricean suggestions concerning the relation between utterances and states of the speaker, the account belief attributions we presently discuss is further weakened. The present account, at least without further elaboration, would not single out a state of the speaker identified as the belief that p.

attributions requires the relation of samesaying or the relation of content identity between utterances/states of the subject and the utterance of the ascriber. The holding of the relation of samesaying or of the relation of content identity between one of the states/utterances of the subject and the utterance of the ascriber is a necessary condition for correctness of the paratactic attribution. The principle of triangle that was presented in chapter 1 suggests that the attitude ascribed to the speaker be the same either it is described in the speaker's language or in the interpreter's language. But the principle does not, on its own, establish conditions under which the identity by the two descriptions holds. We expect a theory of interpretation to establish the required conditions. Since we abandon the suggestion that the relation between subject and the utterance of the ascriber is a causal one a theory of interpretation is required for a propositional attitude attribution. A theory of interpretation would account for the relation of samesaying or content identity between states/utterances of the subject and utterances of the ascriber. The best use we could make here of the suggested version of the paratactic analysis would be to take the logical form of propositional attitude attributions to be as the paratactic analysis suggests.

Taking into account these considerations would suggest the paratactic analysis of propositional attitude attributions to be a tool for radical interpretation theory. The proposal would be that interpretation theory has the aim of correct propositional attitude attributions. These attributions are of a paratactic form. Radical interpretation theory accounts for the relation of samesaying or content identity between utterances/states of subjects and utterances of the ascriber referred to by the demonstrative. The utterance of the ascriber that contains the demonstrative specifies the kind of the utterance or state of the subject which stands in a samesaying or content identity relation to the other utterance of the interpreter. It specifies a state as being of a kind of belief, conjecture, desire, wish and an utterance as being an assertion, report, order, exhortation.

I have one reservation concerning this proposal. My reservation about the suggestion presented above is based on the discussion made in chapters 2 and 3 of the dissertation. The considerations I presented in those chapters intended to show the assumption that a doxastic kind of utterance and an affective kind of utterance can share the same content or can be expressing the same locution not to be straightforward. I presented some grounds on which this assumption could be

challenged³¹. Taking into account these considerations and refusing to take the identity of content across doxastic and affective utterances for granted –but instead leaving the assumption open as a possibility to be confirmed by interpretation theory– has a consequence for the use of paratactic analysis in interpretation theory. Denying to make the assumption about identity of content across doxastic and affective utterances has as a consequence that we do not identify the relatum of the ‘samesaying’ relation or of the content identity relation independently of an identification of the utterance or of the state as doxastic or affective. The suggested paratactic form of a logical analysis of indirect speech and propositional attitude report sentences might require the possibility of such an identification. Nevertheless the paratactic analysis could be modified in a way that does not require the assumption. The modification could be made in the following way. The relation of samesaying or content identity would apply to utterances or states of the same kind, or built into the samesaying or content identity relation would be the kind of the saying of the state doxastic or affective. I leave the suggestion about the modification aside here. Let us keep from the present discussion that the pressures presented in chapters 2 and 3 might restrict the paratactic analysis of indirect speech and propositional attitude attributions.

There is a comment that is due here. The analysis in discussion takes the utterance of the ascriber to be expressing one of the propositional attitudes of the subject. So, the utterance of the ascriber "The earth moves." in the paratactic analysis "Galileo believed that. The earth moves." is taken to be expressing one of Galileo's beliefs. The comment on the paratactic analysis here is that it needs to be supplemented by a theory of belief or by a substantial theory of propositional attitudes. The paratactic analysis does not offer an account of belief attribution. All it suggests is exhausted in the logical form of the attribution sentence. In particular the analysis does not provide an answer to the main problem of the dissertation. It requires conditions for a distinction between beliefs and other propositional attitudes to be known. So, the paratactic analysis does not answer the question asked in this

³¹ In the following chapter, having introduced the notion of mental state, I present parallel considerations about the common content assumption across doxastic and affective states. The force of the present discussion for the case of a paratactic analysis of propositional attitude or mental state ascribing sentences will become more apparent in chapter 5 of this dissertation.

chapter, that is the question about a possible identification of the doxastic attitudes/states of speakers/agents. Now one might suggest that we keep the logical form of the paratactic analysis in talking about propositional attitudes and independently search for a substantial theory of propositional attitudes. The suggestion would be that the paratactic account of propositional attitudes should be supplemented by a theory of propositional attitudes that would identify the beliefs, the desires etc. While I do not object to this suggestion, I think there are conditions related to the logical form of propositional attitude sentences that we should let the most suitable theory of belief and other propositional attitudes determine. In particular, I think we should let the substantial theory of propositional attitudes that would be judged to be the best in the present context determine whether the content of the propositional attitude is separable from the kind of attitude and identifiable across attitudes of different kind. That is, we should not let the logical form of propositional attitude attributions predetermine conditions about identification of content across doxastic and affective states. Identification of content across attitudes will be shown to be possible or it will not be shown to be possible within a substantial theory of propositional attitude attributions.

And there is a last comment to make on the paratactic analysis of propositional attitudes. This comment shows a certain kind of insufficiency of the paratactic analysis in comparison to normal way of propositional attitude attributions. The comment concerns the case of reports of propositional attitudes in which the verb of the attitude is used both in an affective and in a doxastic sense³². An example that has been mentioned is the verb 'to agree'. A suggestion made in chapter 2 was that when this verb is used in a doxastic sense it is constructed with a clause in the indicative mood. When the verb is used in an affective sense it is constructed with a clause in the subjunctive mood. Consider now the paratactic analysis of the two possible uses of a propositional attitude ascription of the verb 'agree':

"John agreed that Mary will give the lecture tomorrow."

and,

"John agreed that Mary should give the lecture tomorrow."

The ambiguity in the use of the verb, in the above cases, is spelled out by

³² This problem of the paratactic treatment of propositional attitudes was pointed out to me by Dr Peter Milne

means of the indicative mood and the infinitive or modal construction of the that-clause. But a paratactic analysis of the above ascriptions would be the same for both cases:

"Mary gives the lecture tomorrow. John agreed that."

A possible attempt that could be made in order to defend the paratactic account from the aforementioned objection would be by postulating lexical ambiguity in the verb 'to agree'. I think that this attempt offers an unattractive option. The objection presented above was not one peculiar to the verb 'to agree'. A paratactic analysis would be insensitive to features of propositional attitude reports (shown by modifications made by indicative/ modal subjunctive or infinitive) of any propositional attitude verb that, like the verb 'to agree', can be used both doxastically and affectively. Thus, a paratactic analysis would be insensitive to the two possible reports of the verb 'to tell'. The verb 'to tell' in a doxastic context is constructed with the verb of the subordinate clause in the indicative. This is the case with the report "He told me that I will give the lecture". The same verb in an affective/directive context is constructed with the verb of the subordinate clause in modal subjunctive or in the infinitive. This is the case with the reports "He told me that I should give the lecture." or "He told me to give the lecture.". A paratactic analysis would represent both reports in the same way, i.e. "I give the lecture. He told me that.". The same holds for reports of fearing, hoping, deciding. My claim is that ordinary reports of these propositional attitudes can be made in two ways according to whether the context is doxastic or affective. Thus ordinary reports distinguish between the following: "He fears that he will give the lecture." is contrasted to "He fears that he should give the lecture." and "He fears to give the lecture.". "He hopes that he will give the lecture." is contrasted to "He hopes to give the lecture.". "He decided that he will give the lecture." (in the sense of concluding or coming to believe) is contrasted to "He decided that he should give the lecture." or "He decided to give the lecture.".

A paratactic analysis does not distinguish between what is contrasted in these cases. The suggested alternative of postulating lexical ambiguity in the propositional

attitude verb looks unattractive to me because lexical ambiguity will have to be postulated in all these verbs. Locating the semantic difference between the contrasted reports on the propositional attitude verb suggests that there are two meanings of the verbs ‘to tell’, ‘to fear’, ‘to hope’, ‘to decide’. The option looks unattractive to me for the additional reason that it leaves the systematicity with which propositional attitudes in doxastic context are constructed with indicative and the systematicity with which propositional attitudes in affective contexts are constructed with infinitive or modal totally unexplained.

In this section I presented and discussed sentential analyses of belief and other propositional attitudes. The comments I made on these analyses intended to show that a sentential analysis of belief and propositional attitude is, in the present context, insufficient to account for an identification of doxastic and of affective attitudes. A sentential analysis of propositional attitudes needs to be supplemented by a substantial theory of propositional attitudes in order for an identification of the doxastic states to be possible.

4.4 Belief as a disposition to act

In this section I consider a general claim about beliefs and its place in the present context of inquiry. There is a widespread claim in accounts of belief that belief should be construed as a disposition to action. Concerning belief attribution, believing that *p* is sometimes explained as being disposed to act as if *p* were true. Let us discuss the general line of the suggestion that belief is a disposition to action. There are problems for an account of belief which is based simply on the assumption that action is a criterion of belief. A first problem is that action is, in an analogous way, a criterion not only of belief but also of desire and a number of other mental states. A second problem arises from the view that mental states are interrelated and so cannot be individuated on the basis of action.

The first of these problems shows that the constraint that action is the only guide available for belief attribution is only the beginning of a theory in need of further elaboration. In following sections of the dissertation a more refined version of the suggestion about the way belief is related to action will be considered and the difference in the ways in which beliefs and desires issue in action will be elaborated.

The other problem I stated above concerns the interrelation of beliefs and desires. The problem was that since it might be the case that mental states are interrelated it will not be possible to individuate them. An instance of this objection is that we cannot have beliefs without having desires and vice versa and consequently behaviour issued by mental states would reflect a fusion of them. We could think of a holistic theory of mental states in which every single mental state, in some way, depends on or is related to the rest of them. The question addressed to such a theory of mental states is whether there are any units constituting the constructs of the theory. An answer to this question might be that any unit, any single mental state, is determined only through its relations to the rest of them. But does the interrelation of two or more items need to exclude any possibility of individuating them? There are properties which we attribute to objects in the physical environment which are not only coextensive but also conceptually interrelated, for instance the property of having mass and the property of being extended. This fact does not impede us from keeping the two concepts distinct and applying them to objects individually. If now we want to admit that the doxastic capacity of the mind is interrelated with the affective capacity of the mind in such a way that each single doxastic mental state depends on a number of mental states some of which are affective mental states, this commitment does not present a reason for an individuation of mental states at least into types, doxastic, affective, to be in principle impossible. On the contrary keeping a terminology of beliefs and desires requires a characterisation of the belief and the desire type of mental state.

Having introduced belief as a notion to be used for a classification of truth-evaluable utterances and having stated some of the constraints that my perspective puts upon the notion, I will proceed with considering a more elaborated account of belief state.

4.5 Belief as a functional state

I continue my discussion on the notion of belief by discussing an account for individuating the belief kind of state. The account to be considered here suggests that states have a functional role on the basis of which states can be individuated. The suggestion here is that it is possible to identify belief and consequently to provide

criteria for correctness of belief attributing sentences as contrasted to other propositional attitude attributing sentences on the basis of the aspect of mental states of being functional states or being described as functional states. According to the accounts of belief considered here, belief is construed as a category of physical or psychological state whose functional role or its position in a network of other states characterises it as belief. A belief ascribing sentence will be that propositional attitude ascribing sentence that attributes a mental state functionally individuated as a belief state. I consider in the following some general features of functionalistic theories of mental state and the possibility of classifying the belief kind of mental states in such a framework.

According to a functionalist theory of mental states, mental states are individuated by having a certain causal role; a mental state is characterised by its causal relations to sensory stimulations, to behavioral outputs and to other mental states (Lewis 1966, 1972, 1978, Putnam 1960) :

“Our view is that the concept of pain, or indeed of any other experience or mental state, is the concept of a state that occupies a certain causal role, a state with certain typical causes and effects. It is the concept of a state apt for being caused by certain stimuli and apt for causing certain behaviour. Or, better, of a state apt for being caused in certain ways by stimuli plus other mental states and apt for combining with other mental states to jointly cause behaviour. It is the concept of a member of a system of states that together more or less realise the pattern of causal generalisations set forth in common sense psychology.” (Lewis 1978)

These considerations are more or less straightforward and have an intuitive appeal for mental states like pain. Pain is construed as that psychological state that typically causes pain behaviour³³. The suggestion that contentful mental states like beliefs and desires are individuated by their functional role is in need of further elaboration. Without intending to let considerations about properties of mental states overwhelm us in the present discussion of belief/desire propositional attitude attributions some considerations are relevant here. For instance, there is an issue about whether propositional attitude attributions should be construed as attributions of non-relational properties individuated by functional role or whether they should be construed as attributions of relations. Pain is a psychological state that, at least on an unsophisticated account of it, it is non-relational. There are adequate grounds against the view that contentful mental states like beliefs and desires are non-relational

³³ See Ned Block (1980) for the advantages of functionalism about mental states over behaviorism.

properties attributed to subjects. Considering mental states to be non-relational properties attributed to subjects would result in considering the belief that snow is white and the belief that snow is yellow to be unrelated mental properties. The productivity and constituency of thoughts as well as certain generalisations concerning mental processes are requirements imposed on the notion of mental state that are difficult to be accommodated by a non-relational view of mental state attributions³⁴.

Nevertheless, contentful mental states like beliefs and desires are often considered to be individuated by their functional role in more or less the same way pain is individuated. It is suggested that it is the causal interaction of a mental state with certain sort of stimuli and certain sort of behaviour that characterises it as pain. Similarly, but more generally and in a manner respectful to the complexity of mental state interactions, it is suggested that it is the causal interaction of a state with certain sort of environmental stimulation, interaction with other mental states and causal interaction with behaviour that characterises the state as a belief that *p* or a desire that *q*. The functional interactions of a state to environmental stimulation, to other states and to behaviour have been considered to be of various kinds. Functional theories of mental states vary in this respect. Suffice it here to mention some proposals concerning the functional interactions between mental states. In computational functionalism the causal interactions of mental states assimilate an empirical computational theory of the mind (Fodor, Putnam). In Gilbert Harman's (1973) theory the causal interactions between states instantiate inferential properties. Hartry Field (1977) construes the interactions between states in terms of subjective conditional probability. At this stage I leave the details of the kind of functional interactions between states aside.

In the functionalistic accounts of the mind, mental states are construed as theoretical entities, which participate in a causal or functional network - computational, inferential, probabilistic. An explanation of an organism's behaviour is given in terms which refer to these theoretical entities which are specified by their position in a network, which is itself specified by sensory input, functional operations and behavioural output. The functional network in which mental states participate

³⁴Jerry Fodor discusses problems of the view that propositional attitudes are monadic functional states in Fodor

serves a two dimensional goal. It individuates the states and it provides the ground on which behaviour is explained. Beliefs, desires, conjectures, wants are theoretical terms referring to mental states attributed to an organism in explaining the behaviour of the organism. The position of those states in a network of states, that is their functional interactions with other states individuates them.

We said that mental states are individuated by their functional role and that they are postulated entities on the basis of which the behaviour of an organism is explained. We explain an organism's actions by attributing to it mental states. For instance, an explanation of why a person is carrying an umbrella could be the following: the person believes that it is raining and he does not want to get wet. The attribution of mental states which constitutes an explanation of behaviour takes a sentential form. In the example I gave what I considered to be an explanation of the person's action of carrying an umbrella was a couple of sentences: "He believes that it is raining.", "He does not want to get wet.". I assume that an explanation of action is given by a system of state/attitude attributing sentences.

The idea to be exploited in a functionalistic framework of states/attitudes concerning the classification of doxastic states/attitudes is the following. The doxastic type of state, belief, conjecture, assumption, suspicion, being truth related, are eligible for functional interactions that states like desires are not. The suggestion based on a functionalistic framework of states/attitudes would be that the kind of functional interactions to other states for which a state is eligible classify the state as been of the truth-evaluable kind. Eligibility for different kind of functional interactions of a state to other states classifies that state as one not susceptible of truth evaluation.

Some of the issues that turn up in a discussion about mental states concern an ontological analysis of mental states: what does a mental state attributing sentence ascribe to an individual; does it ascribe a property of the individual? If yes, what kind of property is this? If it is a relational property which are the relata? Some other problems I consider concern a semantic analysis of mental state attributions: attributing a contentful mental state to an individual is an attribution that introduces issues about semantic analysis of the attribution. The bearing that each analysis has on the other, in respect to the particular problem of individuating kinds of states is,

briefly stated the following: semantic analyses might provide us with specifications of semantic properties of what is ascribed to an individual. A semantic analysis should be such that it would be encompassed by an analysis of mental state ascribing sentences which attribute mental states to an individual and which constitute a system of explaining behaviour. If so, a mental state ascribing sentence should be offered a consistent analysis that specifies both i) the property of a mental state of being a doxastic or affective state and ii) semantic properties of what is ascribed.

In this section I pointed to a certain direction for a solution to the problem of a classification of truth-evaluable utterances. The direction is one according to which the required classification is grounded on the kind of attitude/state related (be ascribing or expressing) to the utterance. Taking belief as a typical case of states/attitudes from the truth-evaluable category, I considered the question of individuating belief. A sentential analysis of belief ascribing sentences was considered and it was suggested that the analysis needs to be supplemented by a substantial theory of attitudes/states. The functionalistic framework of states was presented and the possibility of individuating belief on the basis of the functional role of the state in a behaviour explaining network of states was hinted. A full account of the characteristic kind of role of doxastic state in an explaining behaviour system of states will be given in chapter 6. In the following chapter I put forward some considerations concerning the content of state/attitude attributions.

CHAPTER FIVE

IDENTITY OF CONTENT ACROSS ATTITUDES

5.1 Introduction

One respect in which attitudes/states are important in interpretation theory is that understanding of the linguistic and non-linguistic behaviour of agents is enabled by attributions of propositional attitudes. Psychological explanations are given in terms of states/attitudes. Belief and desire attributions constitute the framework within which actors are seen as rational agents³⁵. In discussing Davidson's paratactic analysis of propositional attitudes in chapter four of this dissertation it was suggested that certain matters of logical form of propositional attitude attributing sentences be decided on the basis of and together with more general considerations; that is, on the basis of and together with a substantial theory of propositional attitudes. The particular aspects of the logical form of propositional attitude ascriptions that I suggested a substantial theory of attitudes should decide concern the claim that content is identifiable across doxastic and affective attitudes. The starting point of the considerations I put forward in this chapter are, then, that certain matters of logical form of propositional attitude attributing sentences will be determined on the basis of a substantial theory of propositional attitudes.

With this considerations in the background I discuss here some issues concerning an analysis of state/attitude attributions. The issues are present in certain dyadic-relational analysis of state/attitude attributions. A dyadic-relational analysis in this context is one that treats believing, desiring, fearing and some other states/attitudes in terms of which behaviour is explained as relations that hold between an agent and something else. Attributing the belief that p to an individual is claiming that a specific relation, that of believing, holds between the individual and something else. The second term of the relation that the mental state sentence attributes is sometimes considered to be a sentence, a proposition, the belief/desire/fear object, the belief/desire/fear content etc.

³⁵ Considerations presented in chapter one suggest that an interpretation theory for a speaker's language is constructed and tested on the ground that, when combined with an interpretation of the speaker's actions - linguistic and non-linguistic-, shows the speaker/actor to be a rational agent. Attribution of propositional attitudes that constitute a rational system of attitudes amounts to the *rationalisation principle* that like a *charity principle* is essential for construction and testing of interpretation theory.

I must clarify that it is not the relational account of attitude attributions that is my target in this chapter. My target is a certain issue that often arises in relational treatments of attitude/state attributions. Here I intend to consider the legitimacy of the assumption made by propositional attitude theorists that doxastic (beliefs, suspicions, conjectures) and affective (desires, intentions) attitudes/states are different relations towards the same kind of objects or contents. Though someone can insist on the relational account of propositional attitudes while holding back assent to the assumption I question, most of the propositional attitude theorists that treat attitudes relationally do seem to make the assumption in question. My target in this chapter is the assumption that the content of a doxastic and an affective state/attitude can be type identical.

The suggestion that criteria for identifying content would grant identity of content across doxastic states, that is, beliefs, conjectures and other truth-evaluable states and affective states, that is, desires, intentions and some other non-truth-evaluable states is the subject of the present chapter. The challenge of the common content assumption is here made by showing there to be difficulties with identifying content across the two kinds of state/attitude. The way the common content assumption is challenged here does not suggest a refutation of the assumption. The challenge is directed at a condition that could ground the common content assumption. Namely the condition that common content can be identified across doxastic and affective states/attitudes. My considerations challenge an assumption that lends support to the common content assumption but they do not prove a refutation of the common content assumption.

The structure of the chapter is as follows. I consider some criteria for identification of content of a propositional attitude and I will examine whether on the basis of those criteria content can be identified across doxastic and affective attitudes. There are three criteria I consider here for identification of content. These are i) content is identified by functional role ii) content of a state/attitude ascription is a mental sentence iii) content is identified by the conditions of satisfaction of the state. I will show that there are difficulties, different in each case, with those three criteria as far as the required identification between doxastic and affective content is concerned. Dealing with case iii) requires the notion of direction of fit. So, as part of the

discussion of case iii) accounts of the difference in direction fit will be considered. I will point out that the suggested accounts of the difference in direction of fit combined with criterion iii) do not enable identification of content across doxastic and affective states. The considerations presented in this chapter suggest that the identification possibility, in the particular ways considered here, is not granted. This will suggest that the common content assumption does not gain support by the identification possibility assumption. I take these considerations to question the common content assumption as to whether the assumption is properly established. The relevance of this point to the general question of the dissertation, that is the possibility of a classification of truth-evaluable utterances/states, is that accounts for the classification in question that we will consider later will not be compelling simply on the ground that they take the common content assumption for granted.³⁶

5.2 Identity conditions of content of functionally individuated states

In this section I consider the possibility of individuating content of a state/attitude on the basis of functional role. Let us assume that sameness of functional role is a sufficient condition for sameness of content. I leave unspecified the sense of 'functional' here. I think that in what I am trying to say here, specifying the functional individuation of content as inferential, linguistic, probabilistic would not make any difference. The main point is that functional role is determined by interactions with other states, sensory input, behavioural output. An issue to which I draw attention is pointed by J. Fodor (1987). Fodor points to this issue for the purpose of objecting to meaning holism. My purpose is different but I will quote Fodor's passage which nevertheless relates to the point I intend to raise.

"...It's an embarrassment for attempts to construct content from functional role that quite different - indeed quite opposed- sorts of mental states can nevertheless share their contents. John hopes that it will snow on Tuesday (because on Tuesday he proposes to ski); Jane dreads that it will snow on Tuesday (because on Tuesday she proposes to pot petunias). John's hope that it will snow interacts with his belief that it will to cause elation; Jane's dread that it will snow interacts with her belief that it will to cause despair; so it's hard to see that the causal roles of the hope and the dread have much in common. But how, then, can it be their causal roles that determine their content?..." (Fodor, 1987)

³⁶ My purpose here is not to forestall the reader in making his own considerations but I state in advance that the granting or not of the assumption will have a crucial role to play for the choice of a suitable decision theory that grounds the thesis I propose in the dissertation.

I explain further what I think is the issue here³⁷. Consider the two following state/attitude attributing sentences: "I believe that John is giving the lecture." and "I desire John to be giving the lecture.". According to the common content assumption these two state/attitude attributing sentences would normally attribute attitudes/states that have the same content, if, say, contextual conditions are equal. Let us assume then that it is possible to individuate the content of the two mental states and represent it as <John, lecture giving>. Let us insist on the condition that sameness of content be determined by sameness of functional role. The question I ask is whether content individuated by functional role can be identified in the two ascriptions. We have assumed that functional role should be construed as the actual or dispositional properties of mental states a characterisation of which involves interactions with sensory input, other mental states and behavioural output, in such a way that the system of the interrelated mental states constitutes an explanation of an individual's behaviour.

The first point to notice is that the sensory input which causes the belief state that John is giving the lecture cannot be assumed to cause the desire state that John be giving the lecture. The individual perceiving John to be giving the lecture can, in a straightforward manner, be assumed to cause the belief by the individual that John is giving the lecture. In opposition, the desire state that John be giving the lecture cannot be charged with this sort of interaction with the same kind of environmental situations.

The second comment concerns the possible interactions of the mental states "I believe that John is giving the lecture." and "I desire John to be giving the lecture." with other mental states. My belief that John is giving the lecture might cause my belief that there is someone who is giving the lecture, it might cause my belief that John is not in his office at the time of the lecture, or it might cause my choice to attend the lecture, given that value John as a lecturer. My desire for John to be giving the lecture cannot be charged with causing my belief that there is someone who is giving the lecture or the belief that John is not in his office at the time of the lecture. In addition it does not interact in a similar way to my valuing John as a lecturer to

³⁷ I have my reservations for accepting all the points in Fodor's passage. My reservations are concerned with the relevance of emotion to functional individuation of content. See also chapter 2 for a discussion on

cause my choice to attend John's lecture.

Thirdly, if we think of the relation of these two mental states with behaviour we find again dissimilarities. My belief that John is giving the lecture interacting with my desire to attend John's lectures gives a mini explanation of the behaviour of mine to go to the lecture. My desire that John be giving the lecture does not function in a similar way in this network.

The suggestion we considered here was that sameness of functional role be a sufficient condition for sameness of content. The identity of content across doxastic and affective attitudes would be grounded in a functionalistic framework of states if it could be shown that the content of a belief and the content of a desire share the same functional role.

Fodor's point in the quotation above was to object to the possibility of functional individuation of content and to strengthen the language of thought hypothesis. I think that what these examples show is that the contribution of an attitude independent content to the functional role of a state which belongs to a system of states which explain behaviour is not straightforwardly individuated. Fodor accepts the attitude independence of content but rejects the possibility that it can be functionally individuated.

"To believe that such and such is to have a mental symbol that means such and such tokened in your head in a certain way: it's to have such a token in your 'belief box' as I'll sometimes say. Correspondingly, to hope that such and such is to have a token of that mental symbol tokened 'in your hope box'..." (1987, p. 17).

In addition he takes a functionalist's view towards the individuation of beliefs and desires.

The identification assumption challenged is, by using the metaphor with boxes, the assumption that a token of a mental symbol in the belief box and a token of a mental symbol in the desire box can be identified as been tokens of the same type. In the following subsection I investigate more thoroughly the place of the assumption in an account of propositional attitudes compatible with the language of thought hypothesis.

5.3 Relations to sentences

I continue by discussing the place of the assumption of the possibility of content identification across doxastic and affective states in a certain relational theory of belief, the sentential theory of belief. Theories that construe beliefs and desires as attitudes toward sentences or utterances of a public language and, thus, identify the content of the belief or a desire with a public language sentence are a special case in this discussion. Public language sentences are the only sort of sentences that do not require a theory beforehand in order to be identified³⁸. Considering beliefs or desires to be relations to public language sentences is unproblematic as far as the identification of the second relatum of the attribution is concerned. So, in this case, the assumption that content is identifiable independently of attitude holds almost trivially. But reflecting on natural language and taking into account the modality introduced by mood or other devices shows that, in the present suggestion of content, conditions for type identity between the content of a belief and the content of a desire rarely obtain. In the cases in which identity conditions do obtain it is because considerations independent of any features of the sentence are in force. The discussion presented in chapters 2 and 3 of the dissertation illustrates this point. In most cases the public language sentence associated with the belief that you will not go to the party tonight is different from the public language sentence associated with the desire for you not to go to the party tonight. The sentence of English associated with that belief is the sentence “You will not go to the party tonight.” while the one associated with that desire is the sentence “Do not go to the party tonight!”. But in chapter 3 we saw that there are cases in which the same sentence “You will not go to the party tonight.(!)” might be used to express a belief or a desire. Normally the sentence is used to express a belief but we saw a case the same sentence was used to express a desire; it was the case the sentence was used by an austere mother to give an order to her teenager son. If beliefs and desires are construed as relations to public language sentences the case above would be a case in which a belief and a desire share the same content. But cases like this are presented as counterexamples to a possible classification of utterances. Besides they are not the cases theorists of propositional attitudes have in mind in suggesting that beliefs and desires can share

³⁸ Of course they have to be distinguished from a mere uttering of noise.

the same content.

Another account of content sentential theories of propositional attitudes employ is one along the line of Davidson's theory or a mental representation theory. This line of inquiry requires a more elaborated account of propositional attitude attributions. These theories adopt the view that beliefs or desires are attitudes toward not a public language sentence or utterance but toward a sentence-like content or a sentence-like object. Thus, these theories require identity conditions for the content of the propositional attitudes.

A view of mental state attributions that does not give a uniform answer to the question of how the second relatum of the belief or desire relation should be construed is put forward by H. Field (1978). The question to which a uniform answer is denied is whether the relatum of the belief or desire be something very similar to a public language sentence or it be a sentence analogue, something resembling a sentence only in the respect of exhibiting sufficiently complex structure. The account of belief H. Field suggests is given along the following line:

1) X believes that p if and only if there is a sentence or a sentence analogue S such that X believes* S and S means that p. (H. Field, 1978)

X is the subject of the attribution, p the content of the that-clause and S a sentence of the subject's language or a sentence analogue. The relation believes* holds between the subject and a sentence of his own language or a sentence analogue. The suggestion is that a subject X believes* a sentence or a sentence analogue S only if the sentence were explicitly stored in the subject's head³⁹. I assume that an account of desiring that q can be given in a similar spirit. Let X be again the subject of the desire that q attribution, q the content of the that-clause, R a sentence of the subject's language or a sentence analogue and desires* a relation between the subject and a sentence of his own language or a sentence analogue. An account of desiring that q would then be given by 2).

2) *X desires that q if and only if there is a sentence or sentence analogue R such that X desires* R and R means that q.*

³⁹ Some difficulties concerning the possibility of storing infinitely many sentences could be dealt by distinguishing between core beliefs and consequences of core beliefs. For more details see Dennett (1975) and

Similarly to the belief attribution the suggestion for desiring* would be that a subject desires* a sentence or a sentence analogue R only if the sentence were explicitly stored in the subject's head.

The main reason for not restraining the relation of believing* or desiring* only to sentences but allowing for the possibility that believing* or desiring* be relations to sentence analogues is the requirement that the possibility of interpretation of languageless creatures be left open. I will leave aside the contention that application of beliefs and desires to a languageless organism requires a sufficiently complex system of internal representations; that is, it requires that sufficiently complex structural entities, the sentence analogues, be stored in the organism. The aspects of 1) and 2) in which I am interested here relate again to allowing believing* and desiring* be relations to sentence analogues. But my interest concerns specifically one point of this possibility. It is the step from taking believing* and desiring* to be relations to public language sentences -sentences of the subject's language- or relations to internal representations of those sentences to allowing believing* and desiring* be relations to sentence analogues that are construed as psychological entities that represent propositions or attitude independent contents of sentences. That is, what I consider to be an issue here is that we can take sentence analogues to be isomorphic to aspects of public language sentences -their propositional content- that are isolated from the doxastic or affective dimension of the sentence.

Field's proposal is that a belief or a desire attribution has two components, only one of which needs to be semantic. The belief that p attribution requires the following:

- a) belief* that R
- b) R means that p

Belief* that R, according to the suggestion, does not need to be a semantic relation. The attribution of the belief* that R amounts to the claim that a sentence or sentence analogue R is stored-as-belief in the subject. H. Field (1978) does not give us a clear account of a storing-as-belief and a storing-as-desire of a sentence or analogue. A rougher dispositional answer would account for the belief* that R on the

basis of the role of R in reasoning, deliberation etc. The materialist's answer, which seems to be one he favours, would require detailed neurophysical information about *how* sentences are stored-as-beliefs or stored-as-desires.

Let us now consider what is the role of the common content assumption in this proposal. The content of a belief that p would be identifiable (type identical) with the content of a desire that q, and the assumption would be justified, if the following conditions are met:

There is an R (sentence or sentence analogue) such that

- i) R is stored-as-belief in a subject's head
- ii) R is stored-as-desire in a subject's head
- iii) R means that p
- iv) R means that q

My claim is that the postulation of a sentence or sentence analogue R that meets i)-iv) is not solely on the constraints of the present suggestion on propositional attitudes guaranteed. The conditions of propositional attitude attributions suggested by 1), 2), a) and b) are consistent with a denial of the claim that there is an R that meets conditions i)-iv). So, the assumption that there is such an R meeting i)-iv), which amounts to the common content assumption is not straightforward in a sentential theory of propositional attitude attributions. The assumption in question, in the context of a materialist account of a sentential theory of attributions, will have to be referred to neurophysiology for confirmation.

5.4 A logical analysis of intentional state

I continue with considering the place of the assumption that content can be identified as type identical across doxastic and affective states/attitudes in another account of propositional attitudes. It is the account of propositional attitudes/states that is based on a logical analysis of states/attitudes like the one John Searle offers. There are two key notions in this analysis of an Intentional state; the notion of direction of fit and the notion of conditions of satisfaction. In the first part of this section I will point out that the assumption in question, namely the assumption that content can be identified as type identical across doxastic and affective state, requires an account of the difference in direction of fit between the two kinds of state. In the

second subsection I will consider attempts to offer an account of the difference in direction of fit. I will examine those accounts on the basis of whether they are sufficient for grounding the assumption in question and on the basis of whether they can be accommodated by the framework of the dissertation.

5.4.1 Identity of content across state in Searle's analysis of intentional state

Searle (1983) proposes an account of Intentional state that is not based on a functionalistic framework of the mind. He offers an analysis of Intentional state in terms of logical properties of the state⁴⁰. I start my presentation of Searle's analysis of Intentional state by quoting a relevant passage:

“...Every Intentional state consists of an Intentional content in a psychological mode. Where that content is a whole proposition and where there is a direction of fit, the Intentional content determines its conditions of satisfaction. Conditions of satisfaction are those conditions which, as determined by the Intentional content, must obtain if the state is to be satisfied. For this reason the specification of the content is already a specification of the conditions of satisfaction...”(1983, pp. 12-13).

Searle does not discuss the problem of type identity of content across Intentional states with a different direction of fit - I comment extensively on Searle's account of the notion of direction of fit in the following subsection- but it seems that the notion of conditions of satisfaction applies only to an Intentional state with a direction of fit. An Intentional state is a representation of its conditions of satisfaction in a psychological mode. Conditions of satisfaction consist of states of affairs in the world which, if obtaining, satisfy the Intentional state. Direction of fit and Intentional content are interrelated characteristics of an Intentional state.

Consider the belief that I will be at the party tonight and the desire to be at the party tonight⁴¹. One issue here is what are the conditions for type identity of content of two states with a different direction of fit. Another issue is how the content of two states with a different direction of fit can be identified as type identical. In the analysis we are presently considering a possible answer to the first question of type identity of content could be given along the following lines: Two states have type identical contents if they have the same conditions of satisfaction. The belief state that

⁴⁰ The suggestion that the term 'Intentional' be spelled with a capital I is Searle's own and I follow it the discussion of his views.

⁴¹ Here, as in other contexts of the dissertation, I consider beliefs and desires to be representative states from

I will be at the party tonight is satisfied if the belief is true, that is if I will be at the party tonight. The desire state is satisfied if the desire is fulfilled, that is again, if I will be at the party tonight. The sense in which the belief state that I will be at the party tonight and the desire state to be at the party tonight have the same conditions of satisfaction is that the same state of affairs in the world if obtaining would satisfy both of the states.

The above stated condition for type identity of content is plausible, but not adequate. The notion of satisfaction needs to be spelled out; we need to give an account of what it is for the belief that *p* and for the desire that *p* to have the same conditions of satisfaction. The conditions of satisfaction of a state are conditions which as determined by the state must obtain if the state is to be satisfied. It is in virtue of a characteristic of an Intentional state that it has the conditions of satisfaction it has. The notion of dependency that requires some explanation here is the one according to which a certain state of affairs constitutes condition of satisfaction *of* a certain Intentional state. Searle deals with this issue by pointing to an ambiguity of the term 'conditions of satisfaction'. The ambiguity of the term is like the ambiguity between a requirement and the thing required. I do not find the analogy precise enough and I attempt to offer a more precise version of the same idea in the following paragraphs but let us discuss the suggestion here using Searle's terms. I quote the relevant passage and I comment on it.

"...So, for example, if I believe that it is raining then the conditions of satisfaction of my belief are that it should be the case that it is raining (requirement). That is what my belief requires in order that it be a true belief. And if my belief actually is a true belief then there will be a certain condition in the world, namely the condition that it is raining (thing required), which is the condition of satisfaction of my belief, i.e., the condition in the world which actually satisfies my belief..."(1983, p. 13).

If an account of type identity of content is to be based on the notion of condition of satisfaction, the notion must be taken to signify a requirement characterising the state and not the thing required. A state of affairs (thing required) constitutes a condition of satisfaction *of* the belief that *p* and not *of* the belief that *q* because it satisfies a requirement determined by the belief that *p*. That the state of affairs that I will be at the party tonight satisfies the belief that I will be at the party tonight is determined by a characteristic of the Intentional state. The condition in the

world (thing required) which might obtain and satisfy the state cannot be a characteristic of the state itself and a specification of a condition in the world does not suffice on its own for characterising the Intentional content of a state⁴². The insufficiency of the suggestion that a specification of a condition in the world (thing required) suffices on its own as an account of the Intentional content of a state becomes more obvious in the cases of perception and intention. The content of my intention to be at the party tonight cannot be specified solely by the thing required - I will be at the party tonight. The thing required - I will be at the party tonight - satisfies the Intentional state of the intention to be at the party tonight and constitutes a condition of satisfaction of the state in virtue of being in a certain kind of dependence relation, often taken to be causal, to the Intentional state.

The satisfaction condition (thing required) is not part of the Intentional state. It will have to be a relation *to* the condition in the world which is essential to the Intentional state. Searle proposes this relation to be a requirement that the condition is satisfied implicit in the Intentional state. The suggestion that we consider this relation to be that of a requirement helps only to disambiguate the representation from the thing represented. It does not offer an account of the nature of this relation. The nature of this relation is the very problem of the Intentionality of the mind and it lies far beyond the ambitions of this dissertation to attempt an answer to it. The only issues concerning the relation that I bring into consideration here have to do with the degree up to which facts about a state being doxastic or affective might enter into the nature of the relation.

The account of type identity of content across states with a different direction of fit proposed above was that the content of a belief and the content of a desire are type identical if the two Intentional states would be satisfied by the same state of affairs. This account tells us in what content identity consists; it tells us that content identity consists in identity of conditions of satisfaction. The issue of identification of content across two states, which is the second question I consider here, would resolve

⁴² Other problems with the suggestion that a specification solely of the thing required suffices for a specification of the Intentional content relate to cases like disjunctive beliefs. The state of affairs (thing required) that it is raining satisfies the belief that *it is raining or it is snowing* and it satisfies the belief that *it is raining or (it is snowing and it is not snowing)*. But it is insufficient for specifying the Intentional content of either of the disjunctive beliefs.

by identification of the same conditions of satisfaction. Identifying the same conditions of satisfaction across a belief and a desire would suggest content identity between the two states. The question of identification of conditions of satisfaction between two states is a question that relates to the notion of satisfaction in the sense of a requirement implicit in the two states. The problem to which I point in the following has its source in a feature of the notion of satisfaction (as a requirement). This is that satisfaction (as a requirement) can be between satisfaction by truth or satisfaction by fulfilment. The claim I make in the following is that an account of the notion of satisfaction, a notion on which identification of content relies, requires an account of the difference in direction of fit between beliefs and desires.

Let us, then, examine more closely the situation. Let us consider the possibility of getting a condition for identifying content across the two kinds of state, doxastic, affective, based on the notion of condition of satisfaction taken to signify a requirement characterising the Intentional state. I said above that the analogy between the type of ambiguity between a requirement and the thing required and the type of ambiguity involved in the notion of conditions of satisfaction is not precise enough. A more precise version of the same idea could be given along the following lines: There is a function h which for every Intentional state determines a set of conditions in the world. An Intentional state I is mapped onto the set of conditions in the world which satisfy the state. Type identical Intentional states are satisfied by the same set of conditions in the world -network and background constraints being equal-. Distinct Intentional states can be satisfied by the same set of conditions in the world. Let us call function h the satisfaction function. A satisfaction function is internal to an Intentional state. That an Intentional state is satisfied by certain conditions in the world is a characteristic of the state which individuates the state.

Consider now a doxastic Intentional state and an affective Intentional state. A criterion for type identity of their contents based on the satisfaction function of the states could be the following: The two states have type identical contents if the values of the satisfaction function of each are the same.

The claim here is that unless the criterion supplemented by an account of the difference in direction of fit, it does not give us an affirmative answer to the question of whether the content of a belief and the content of a desire can be type identical.

Here are my reasons for holding this. We assume that there is a notion of satisfaction such that a satisfaction function is implicit in every Intentional state. The same content can be identified across a doxastic and an affective state if the same satisfaction function can be identified. Without begging the question concerning the issue of whether there is a general notion of satisfaction in relation to both doxastic and affective states, we can assume that the satisfaction conditions of a doxastic state are truth-conditions and that the satisfaction conditions of an affective state are fulfilment (good) conditions. Beliefs are satisfied by being true, desires are satisfied by being fulfilled. A doxastic state, say the belief that I will be at the party tonight is mapped onto a set of conditions in the world which make it a true belief. An affective state, the desire to be at the party tonight, is mapped onto a set of conditions in the world which would fulfil the state. We could assume that the nature of satisfaction is the same in both cases of beliefs and desires and that in the case of beliefs the satisfaction conditions are truth-conditions while in the case of desires the satisfaction conditions are fulfilment conditions. This would require a distinction to be drawn between doxastic and affective states independently of classification of satisfaction conditions as truth or fulfilment (good) conditions. In order to save the criterion for type identity of content suggested above we have to find a way of distinguishing the domain of beliefs from the domain of desires, while keeping a general notion of satisfaction. Having a way for distinguishing the domain of beliefs (doxastic states) from the domain of desires (affective states) would determine that satisfaction conditions on the first domain are truth conditions, while satisfaction conditions on the second domain are fulfilment conditions. In the case that the satisfaction-truth conditions of a doxastic state and the satisfaction-fulfilment conditions of an affective state coincide the two states have type identical contents. For a way of individuating content across the two category of states not to be question begging it should not rely on the truth-evaluability of doxastic states or the non-truth-evaluability of the affective states. Let us notice that the pressure here comes from the central question of the dissertation, the distinction between truth evaluable and non-truth-evaluable states/utterances.

The following subsection is devoted to some attempts to account for the difference between beliefs (doxastic states) and desires (affective states), a difference that in the literature is known as a difference in the direction of fit. The point in

considering accounts of the difference in direction of fit here is that such an account would ground the satisfaction by truth of the beliefs and the satisfaction by fulfilment of desires and it might contribute to an answer to the question of the dissertation. One requirement on an account of the difference in direction of fit that needs to be stressed here is that the account should not rely on the notion of truth or a notion of fulfilment (or good). The reason for the presence of this requirement concerning the issue here discussed is that we need an account of the difference in direction of fit that would distinguish the domain of states satisfied (evaluated) by truth from the domain of states satisfied by fulfilment (not evaluated by truth). This is what would enable the proposed criterion for type individuation of content across attitudes and it would be an acceptable account given the constraints of the dissertation.⁴³

5.4.2 Direction of fit

I consider first Searle's suggestion for an account of the difference in direction of fit. His account of the difference in direction of fit between beliefs and desires does not use directly the claim that beliefs (doxastic states) are the states susceptible of truth-evaluation while desires (affective states) are not. That is, that desires are evaluated by fulfilment or good. But still, it will be suggested here, the account exploits the truth-normativity of beliefs and the non-truth-normativity of desires. For this reason and from the present point of view it will be argued here that the account does not offer an interesting solution to the problem of the difference between doxastic (truth-evaluable) and affective (non-truth-evaluable) states. And, on this ground it will be pointed out here, Searle's account of the difference in direction of fit does not offer a non-question begging answer to the difference between satisfaction by truth and satisfaction by fulfilment required for the suggested criterion for the common content across the two kinds of state.

The discussion on the difference in direction of fit will be supplemented by a presentation of Humberstone's account (Humberstone, 1992). It will be pointed out

⁴³ I think that the requirement that an account of the difference in direction of fit should not rely on a notion of truth or a notion of fulfilment (or good) is important for any respectable account of the difference in direction of fit. On the one hand, taking truth or fulfilment (or good) for granted suggests a trivial answer to the question of direction of fit: states with the one direction of fit are the ones that admit of a truth-value, states with the other direction of fit are the ones that admit of fulfilment or are good. On the other hand, it would be significant for considerations about truth and about agency if an account of the difference in direction of fit had it as a

that Humberstone's account is not question begging concerning the truth-evaluability and non-truth-evaluability of the two directions of fit and that it is an attempt to account for the truth normativity of belief and the non-truth normativity of desire. Humberstone's account will be criticised on the ground that it is not adequate concerning the required distinction in direction fit.

Searle's account of the difference in direction of fit between beliefs and desires is based on the difference between cases that things go wrong, that is, it is based on conditions of failure of satisfaction.⁴⁴ The difference in direction of fit between beliefs and desires lies in the different location of the mistake in case things go wrong. According to the present suggestion an Intentional state has the "mind-to-world" direction of fit if, in case things go wrong it is the state that is at fault⁴⁵. A state has the "world-to-mind" direction of fit if, in case things go wrong, it is the world that is at fault. In the first case if a state fails to get the relevant value, for instance a belief fails to be true, it is a fault of the Intentional state, not of the world. That is, an Intentional state has the "mind-to-world" direction of fit if it is evaluated on the basis of whether it matches the world or not. In the second case an Intentional state has the "world-to-mind" direction of fit, if in case it fails to get the relevant value, for instance a desire fails to be fulfilled, it is a fault of the world, it is not a fault of the Intentional state; the world has to change so that the state gets the relevant value. In this case it is on the basis of whether or not the world matches the state that the state gets the relevant value.

There are some general comments, more or less significant, someone could make on the proposed account. An analysis of the notion of direction of fit based solely on the different ways in which an Intentional state can fail to get the relevant value does not get us very far. The account does not show the significance for a state to have a certain direction of fit. The condition for the distinction between the

corollary, and not as a presupposition, that states with a certain direction of fit are susceptible of a truth-value.

⁴⁴ Searle adopts and elaborates Elisabeth Anscombe's (1957) suggestion for an account of the difference in direction of fit.

⁴⁵ In the literature the difference in direction of fit is called in various ways. In Searle (1983) the contrast is between "mind-to world" and "world-to mind" direction of fit. In Humberstone (1992) the difference is betweenthetic and telic direction of fit. In Velleman (1998) the difference is between cognitive and conative attitudes. Here I use the titles doxastic, cognitive,thetic, "mind-to-world" interchangeably to signify the direction of fit of beliefs and the titles affective, telic, conative, evaluative, "world-to-mind" interchangeably to signify the direction of fit of desires.

direction of fit is not applicable in case the Intentional state gets the value by which it is evaluated, that is the case that nothing goes wrong, the case that nothing needs to be revised. The account does not point to the significance of the notion of direction of fit in an analysis of states that would contribute to a theory of explaining behaviour. A state being doxastic or affective, that is having the one direction of fit or the other, signifies a difference in the role of the state in the psychology of the agent. This is an aspect of the direction of fit left out by the present account. These points will reappear and will be discussed more extensively in the following parts of the dissertation. But here I intend to focus on one aspect of Searle's account of the difference in the direction of fit between beliefs and desires. This aspect of his account is what presents problems for the individuation of content across states with a difference direction of fit moreover for a possible accommodation of his suggestion for a distinction between doxastic and affective states by the context of the dissertation.

In order to illustrate the point I intend to raise let us say that for every Intentional state *I* there is a value *x* from a range of values {*t*, *v*} that the state either gets or it fails to get. In case an Intentional state *I* fails to get the value *x*, we might agree that it gets the value x^{146} . For an Intentional state to get a value x^1 , would mean that it fails to get the value *x*. Applying Searle's sufficiency condition for a state to have, say, the "mind-to-world" direction of fit, would yield the following: in case a state gets the value x^1 and it is the state that is at fault, then the state has the "mind-to-world" direction of fit. We could say that in this case *x* is *t*.

Let us notice that the part of the sufficiency condition 'it is the state that is at fault' begs the question concerning an account of the difference in direction of fit. We can see the insufficiency of the account by considering an Intentional state with an unknown direction of fit. Let us take an Intentional state with an unknown direction of fit and with the Intentional content signified by the that-clause 'I will be at the party tonight'. Searle's condition is that in case the state fails to be satisfied and it is the state that is at fault then the state has the "mind-to-world" direction of fit.

Applying Searle's condition for determining the direction of fit of the state requires

⁴⁶ Notice that the relation between *x*, x^1 and *v*, v^1 is suggested by Searle's criterion on the basis of failure of satisfaction.

the application of a certain condition of failure of satisfaction. But his suggested condition of failure of satisfaction is not applicable unless we can locate the fault. We have to be able to tell that in a certain case it is the state that is at fault and in a different case it is the world that is at fault. But we cannot do this unless we know the direction of fit of the state. It does not make sense to ask whether a state with an unknown direction of fit is at fault. The state with an unknown direction of fit with content specified by 'I will be at the party tonight' failing to be satisfied, that is it obtains that I will not be at the party tonight, does not indicate where the fault is located. It does not indicate whether it was wrong to hold the state or the way things happen is wrong. The location of the fault in case of failure of satisfaction that, as suggested would determine the direction of fit, requires the distinction between the direction of fit. If so, the suggested criterion of the difference in direction of fit applied in the present context would be question begging. Since an account of the difference in direction of fit is required here in order to distinguish between the truth-satisfaction of doxastic states from the fulfilment-satisfaction of the affective states, the criterion suggested would be question begging concerning this distinction as well. The possibility of locating the fault in case things go wrong requires the truth-normativity or non-truth-normativity of the state in question.

The considerations above were made under the assumption that there is a general notion of satisfaction which will be satisfaction by truth for the case of states with the one direction of fit and satisfaction of fulfilment for the case of states with the other direction of fit. The suggestion considered was that we could assume that there is a general notion of satisfaction as long as we have a distinction between doxastic and affective states.

It seems, then, that the required characterisation of the difference in direction of fit is not given by Searle. The required characterisation would be one which determines the difference in the direction of fit of Intentional states in a way which does not presuppose the notion of satisfaction by truth or by fulfilment. To clarify my claim at this point, I must mention that I do not object to all aspects of Searle's account. I do accept the claim that there is a difference in direction of fit between the belief that *p* and the desire that *p*. The claim that is of interest here is that Searle's attempt to distinguish between the direction of fit cannot be useful for the suggested criterion

for the common content across doxastic and affective states. An acceptance of the suggested criterion for identity of content across states with a difference in direction of fit requires an account of the difference in direction of fit not relying on the notion of truth or fulfilment. In addition his account of the difference in the direction of fit is not helpful as an answer to the central question of the dissertation, that is the problem of a classification of truth-evaluable states/utterances. An account of the notion of the direction of fit of an Intentional state that could contribute to a classification of the state as a truth-evaluable state cannot be based on conditions under which a state gets or fails to get a value. An Intentional state is evaluated by the values true or false and not by fulfilled or unfulfilled because it has the “mind-to-world” direction of fit. At least, this is the direction of explanation that accords with the line of this dissertation. Conditions of truth of a belief are conditions which if met, make the belief a good belief. But an Intentional state is a belief state independently of whether it meets the conditions of truth or not.

The account of the difference in direction of fit proposed by Searle and discussed above is a case of an account of the difference between beliefs and desires that exploits the truth-normativity of beliefs -and the non-truth-normativity of desires. Truth is a norm for beliefs but not for desires. An elaboration of this idea we find in L. Humberstone (1992). Humberstone’s account of the difference in direction of fit between beliefs and desires is based on the difference in background intentions had by an agent between having a belief and having a desire.

The truth-normativity of beliefs is suggested by the following⁴⁷:

“...it’s that unless one counts one’s (current) intentions in Φ -ing that p as thwarted if it is not true that p , one’s Φ -ing that p does not constitute believing that p .” (1992, p.73)

The normativity of desire to issue its own fulfilment is, by contrast suggested by the following:

“... On the “higher order” proposal for controlling intentions in the telic case, the intention is that it be the case that p , given the telic attitude toward p : intention, desire, or whatever.” (1992, p.75)

Formalised the background intentions for the two cases are the following:

Intend ($\neg Bp/\neg p$) (thetic case)

Intend (p/Wp) (telic case)

This account of the difference in direction of fit is not vulnerable to the criticism presented for Searle's account. The account of the difference in direction of fit suggested by Humberstone does not rely on the truth-value susceptibility of belief and the different value susceptibility of desire. The attempt here is to offer an account of the truth-normativity of belief and the different kind of normativity of desire. So this account of the difference in direction of fit, if shown to be adequate, could be used for deriving a distinction between the truth-evaluable states, or the states satisfied by truth and non-truth-evaluable states, or states satisfied by being fulfilled (or good). The use that could be made of this account of the distinction in question would be along the following line. In the case that the background intention of a subject in forming an attitude with content *p* is not to have the attitude unless *p* is true, then the attitude has a thetic direction of fit and it is susceptible of truth-evaluation. In the case that the background intention of a subject in forming an attitude with content *p* is that it be the case that *p*, then the attitude is not susceptible of truth-evaluation –it is satisfied by fulfilment or good. Let us notice that this account of the difference in direction of fit, if accepted, would not beg the question concerning the identity of content across states of a different direction of fit.

Surely the account Humberstone suggests of the difference in direction of fit can be criticised for appealing to a notion of background intention which is an obscure notion itself. Do background intentions have a direction of fit? What are the conditions for a correct attribution of a background intention of the one kind in contrast to a correct attribution of a background intention of the other kind? I leave these questions out of consideration here in order to focus on one point of criticism of Humberstone's account. My criticism of the account in consideration concerns its adequacy. In the following I present cases of states which are expected to be classified as doxastic but which do not satisfy the condition of Humberstone's account for having the thetic direction of fit. I take these considerations to show the account to be insufficient for the required distinction. The general suggestion I will introduce after criticising this account for adequacy is that an account of the difference in direction of fit needs to exploit the psychological dimension of the notion of direction of fit.

⁴⁷ Humberstone calls the one direction of fit thetic and the other telic.

Velleman (1998) considers a case of a cognitive attitude that is not truth-directed⁴⁸. This is the case of imagining or pretending as it occurs in children's play⁴⁹. In those cases children are in a state of regarding a proposition *p* as true while they lack the intention to stop being in that state on discovering that *p* is false. What classifies a state like this as doxastic/thetic in contrast to conative/ affective/telic is the role of the state in the behaviour of the subject. The children who pretend to be elephants behave as if they are elephants. They are in a doxastic kind of state with content specified by 'child *x* is an elephant' because they behave as if the proposition specified by 'child *x* is an elephant' was true. At the same time they lack the intention to give up being in the state of regarding the proposition as true on discovering that 'child *x* is an elephant' is not true. They regard the proposition as true while disregarding its truth-value.

Here, I claim that there are more occurrences of the 'regarding as true' role of a state accompanied with a disregard for the truth of its content. I take the 'regarding as true' role here to signify the role of the doxastic/thetic direction of fit. Of course this is not at all an explication of the role of the doxastic/thetic direction of fit. It merely amounts to an assumption that there is a unifying role in the psychology of agents for the states of that direction of fit.

Cases of states that are classified as doxastic but they might not be states for which their bearer has an intention to abolish in case their content is false are beliefs deeply embedded in a subjects repertoire of attitudes, beliefs that are beyond falsification and beliefs formed by wishful thinking⁵⁰. A subject might lack an intention to abolish a belief that *p* on discovering that not *p* if that would require a considerable revision in the rest of the subject's states. A belief formed by wishful thinking might persist inspite of evidence to the contrary. In the case of beliefs beyond falsification the criterion has no application. Though the first two cases are abnormal in a certain respect that needs to be clarified, there is a clear sense in which those cases are classified as beliefs. Nevertheless those states are not formed or accompanied by a

⁴⁸ Velleman uses the term 'cognitive' for the doxastic/thetic/"mind-to-world" direction of fit.

⁴⁹ Velleman's purpose in introducing this case is to distinguish beliefs from other sates of the same direction of fit. My purpose here is different since it is an account of the direction of of fit that I am after and not only of belief.

⁵⁰ These cases are not exhaustive of the category of beliefs that persist with disregard for truth. But I hope they are sufficient for my purposes here.

background intention not to be in order or to be abolished in case their content is not true. I take these cases to show the criterion of the background intentions suggested by Humberstone to be insufficient for grounding the thetic direction of fit.

The sense in which the cases of states presented above are classified as doxastic as well as the sense in which the cases of the deeply embedded beliefs and the beliefs formed by wishful thinking are abnormal needs to be explained. My view is that those states are classified as doxastic on the ground that they have a certain role in the psychology of agents. In particular, I suggest that those states are classified as doxastic on the ground that they have the role of being evidence *for* other states. In addition I suggest that those cases are abnormal on the ground that they persist with disregard for evidence that would discredit them.

A state formed by wishful thinking is classified as a belief because the state figures as evidence *for* other states. A state that *p* that is beyond falsification is classified as belief because it figures as evidence *for* other states. A belief deeply embedded in a subject's repertoire of attitudes lends support to many other states of the subject.

I must clarify that my view is not that a state figuring as evidence *for* other states is sufficient for accounting for the thetic/doxastic direction of fit. My suggestion would involve more than this condition and it will be elaborated in the following chapter. My purpose here is to show an insufficiency in Humberstone's account. Humberstone suggests that the role of evidence in the case of the thetic direction of fit is derivative of his own account.

"It appears, though, that the role of evidence in the present connection is derivative: it is hard to see how one could attitudinise with the aim of not believing something unless it was true other than by taking advantage of whatever evidence came once way." (*op. cit.* p.78)

My comment at this point is that Humberstone's account derives the role of evidence in the case of the thetic direction of fit as far as the grounding of attitudes is concerned. It leaves out of consideration the role of the states with the thetic direction of fit of grounding other states, their role of being evidence *for* other states. In the counterexamples considered above the role of evidence in relation to beliefs is frustrated as far as the grounding of beliefs is concerned. And this accounts for the abnormality of those cases. But as long as those states figure as evidence *for* other

states they fall into the doxastic category of states.

The presentation of my view concerning an account of the thetic direction of fit as well as an account of the difference in direction of fit between doxastic and affective attitudes will be completed in the following two chapters. Here I presented my criticism against two kinds of approach to the problem of a distinction in direction of fit. Concerning Searle's account my criticism was based on showing how the account relies on the truth-value susceptibility of doxastic states and a different value susceptibility of affective states. As such, the account is not useful for granting the difference between satisfaction by truth and satisfaction by fulfilment (or good). I claimed that an account of this difference is required for the possibility of an identification of content across doxastic and affective states. Humberstone's account which is an attempt to account for the truth-normativity of the thetic direction of fit and the different kind of normativity of the telic direction of fit is based on the different background intentions of the subject. I claimed that this account is insufficient for grounding the distinction in question.

My view is that an account of the difference of direction of fit needs to exploit a psychological dimension of the two directions. I will suggest that decision theoretic considerations can exploit and explicate the psychological dimension of the two kinds of state. In particular, I will suggest that the thetic direction of fit is accounted by the holding of evidential relations between states. The doxastic states (the states with a thetic/mind-to-world direction of fit) will be classified as the states that belong to the category of states the members of which stand in evidential relations to each other. The affective states (telic/world-to-mind/conative direction of fit) will be classified as the states that belong to the category of states the members of which stand in a preference relation to each other. The suggestion is that states classified as doxastic are satisfied by truth. States classified as affective are satisfied by fulfilment (or good). The direction taken in the following parts of the dissertation is that the relevant psychological dimension of the two categories of state can be accounted by decision theoretic considerations. After considering some theories of decision, I will propose a decision theory (Savage) that does not beg the question concerning the common content assumption. Full exposition of the proposal is given in chapters 6, 7.

To recapitulate: In this chapter I considered the assumption that doxastic and

affective states can share the same content. I discussed approaches for the possibility of identification of content across the two kinds of state. I pointed out that there are difficulties with those approaches as far as the identification of content across the two kinds of state is concerned. The result of my considerations in this chapter is that the assumption that content is identifiable across attitudes is not straightforward and this is a challenge to possible grounds of the common content assumption.

CHAPTER SIX

THE DECISION THEORY APPROACH

6.1 Introduction

The approach introduced in this chapter offers a systematic exposition of the notion of belief in its place in the context of rational action. The approach here attempts to grasp the notion of belief as it is rationality considerations on the system of attitudes of agents. The approach is based on assumptions about the notion of belief often made by decision theorists. The concepts central in decision theories are those of action, preference and belief. Decision theoretic considerations concerning the central concepts of the theory provide an account of these concepts. I have to state in advance that the presentation of decision theoretic considerations made in the following will involve some technical formulations of the relevant notions. I apologise to the reader for the tediousness the following the discussion might imply. An effort has been made so that the technically introduced notions be given an intuitive interpretation. The introduction of the technical apparatus is required for a formulation of the view I put forward in the dissertation and I hope that the significance of the technically introduced notions to the issues discussed up to now will emerge.

The basic aspect of belief the approach here considers is that belief admits of degrees. Though the theory of partial belief was originally introduced as an answer to the problem of probabilistic judgements, it has interest in its own right, perhaps as a part of a theory of belief. A probability judgement, like 'probably it will rain tomorrow' expresses, according to the subjective view of probability, a partial belief of the subject that it will rain tomorrow. The subjectivist's view of probability that introduces the notion of partial belief is contrasted to the objective view of probability. According to one objective view of probability, probabilities are the limits of relative frequencies and a probability judgement is a judgement about a feature of the external world, the frequency of an event. The objective view of probability has been criticised for its inability to account for all kinds of probabilistic judgements. There are few, if any, philosophers who think that the only worthwhile concept of probability is one that makes it an objective feature of the physical world. But there

are numerous philosophers who think we need more than one concept- at least one 'physical' and at least one 'epistemic'⁵¹.

The subjective view of probability accounts for probabilistic judgements by taking probabilistic judgements to express a partial belief of a rational subject and by taking probabilities to characterise the degree of the partial belief of the rational subject. The debate between the subjective and objective view of probability will be left out of consideration here. Since our primary interest here is in a theory of belief, we will be focused on the subjectivist's view of probability, according to which probability is degree of rational belief.

In this part of the essay the basic assumptions of a subjectivist decision theory will be stated, part of which is a theory of partial belief. The theories of partial belief considered here are the ones in which the assumption that partial belief is measured on the basis of overt behaviour is made. Historically they are theories of partial belief developed after the work of F. Ramsey on probability. So, theories of partial belief that assume the degree of belief to be identified by a feeling or complex of feelings are left out of consideration⁵².

Four theories will be considered here and they will be presented in order of historical development. The theories are i) Bruno de Finetti's theory of subjective probability, ii) Frank Ramsey's theory of belief and utility measurement, iii) Savage's axiomatised theory of decision, and iv) Jeffrey's non-standard decision theory. A presentation and critical examination of those theories will be attempted. The point of view of this presentation and criticism concerns a possible use that can be made of a theory of partial belief or a decision theory, in the context of interpretation theory, for disentangling two factors issuing in behaviour i.e. beliefs and desires. A more thorough investigation of the relation of a theory of partial belief or of a decision theory and interpretation will be presented in the following chapter. Here, I state in advance, that these theories are envisaged from the point of view of their possible accommodation by interpretation theory, and I ask to be excused for the insensitivity and neglect I show to many aspects of these venerable pieces of work.

⁵¹ Two philosophers who consider the only worthwhile concept of probability to be objective probability are Hans Reichenbach (1949) and Richard von Mises (1946). Philosophers who claim that we need more than one concept of probability include K. Popper (1990), R. Carnap (1951), D. Lewis (1980), D.H. Mellor (1971).

6.2 Bruno de Finetti's and Frank Ramsey's theories of subjective probability

In the theories to be considered here belief is measured on the bases of action. In the works of Bruno de Finetti and Frank Ramsey subjective probability is measured on the basis of betting behaviour of the subject. In the work of Bruno de Finetti we find the following suggestion:

“ It is a question simply of making mathematically precise the trivial and obvious idea that the degree of probability attributed by an individual to a given event is revealed by the conditions under which he would be disposed to bet on that event...Let us suppose that an individual is obliged to evaluate the rate p at which he would be ready to exchange the possession of an arbitrary sum S (positive or negative) dependent on the occurrence of a given event E , for the possession of the sum pS ; we will say by definition that this number p is the measure of the degree of probability attributed by the individual considered to the event E , or, more simply that p is the probability of E (according to the individual considered...)” (De Finetti, 1936, p. 101)

In the work of Frank Ramsey we find a similar idea:

“The old established way of measuring a person's belief is to propose a bet, and see what are the lowest odds he will accept.” (Ramsey, 1926)

A problem with a measurement of belief that takes as entries the odds a person would be willing to accept in a bet situation is that an independent assumption about the utility of money is needed. As a matter of fact money has a diminishing marginal utility. De Finetti takes into account this feature of monetary value and assumes that utility is proportional to money when stakes are small. Another feature of monetary value that complicates the situation is that money does not have the same value for everyone. Two subjects with the same degree of belief in p might accept different odds in a betting situation because of the different values they attach to the same amount of money. This problem is solved by Ramsey's method in which no assumptions about the independent utility of money is made. His method will be discussed shortly.

One central notion in many decision theories is that of coherence. It is central in the sense that an assumption of coherence is necessary in order for a measurement of degrees of belief and, in some theories of degrees of utility, to be possible. The notion of coherence in decision theory is usually defined in terms of the possibility of a Dutch book. The idea is that a subject is coherent if no Dutch book can be made against him; that is, the subject's evaluations which determine his choices are such

⁵² For instance, see De Morgan (1947).

that there is no case in which the subject chooses to enter into a bet or family of bets which is such that he must lose whatever happens. De Finetti, who chooses to talk about coherency of probability evaluations, (op. cit. p.103 fn) considers the following definition:

“...once an individual has evaluated the probabilities of certain events, two cases can present themselves: either it is possible to bet with him in such a way as to be assured of gaining, or else this possibility does not exist. In the first case one clearly should say that the evaluation of the probabilities given by the individual contains an incoherence, an intrinsic contradiction; in the other case we say that the individual is coherent.” (op. cit., p. 103)

In De Finetti's definition of coherency there is an implicit assumption about the utility of money. De Finetti is aware of the use of the assumption and admits that he presents an idealisation of actual situations. The assumption is that the subject desires to maximise monetary gains in betting situations and that he acts so as to maximise monetary gain. This is a constant parameter in the determination of the subject's choices. The subject is coherent if, given that he desires to maximise monetary gains, he has subjective probability evaluations that would never direct him in choosing a bet or family of bets that would result in certain monetary loss.

The above stated account of a coherent agent is very restrictive. It requires not only the assumption that a coherent subject is one who would make choices in order to maximise monetary gain, but, in addition it requires this desire to be the only desire a coherent subject has⁵³. Desires of a subject irrelevant to monetary gain do not enter as parameters in the definition of coherency. Nevertheless we can think of cases in which a subject lets a Dutch book be made against him because of the value he assigns to non-monetary consequences of his choices. In such cases we do not want to deny that the subject has coherent beliefs or that he is a rational agent⁵⁴.

A modified definition of coherency that would not require the restricting assumptions of the above definition is given by Ramsey. Ramsey's method is based on assumptions about 'goods' or 'bads' that might be the result of a bet to the person who accepts the bet. His theory deals jointly with subjective utilities and probabilities and his method results in a simultaneous belief and utility numerical representation. In

⁵³ My comment here concerns the interpretation of the betting behaviour model. Considerations of the kind that a coherent subject assigns utilities that can be represented by real numbers are irrelevant to whether we consider utilities here to correspond to monetary gain or loss. The real number representation of utilities cannot be taken for granted.

⁵⁴ F. Schick (1986) points to problems related to the Dutch book definitions of coherency.

Ramsey's theory of belief measurement the bets are not made in terms of money but in terms of goods and bads for the subject.

"If we call 'goods' the things that a person ultimately desires, the assumption is that goods are measurable and additive." (op. cit., p. 69)

The person's behaviour is governed by a law of mathematical expectation. That is, if p is a proposition about which he is doubtful, any goods or bads for whose realisation p is, in his view, a necessary and sufficient condition enter into his calculations multiplied by the same fraction, which is called 'the degree of his belief in p ' (p.70). There is, in addition, another couple of assumptions, psychological in character. The first is that a person's actions are completely determined by his beliefs and desires and the second is that a person's actions are conducted in a way which makes it most likely to the person that the objects of his desires will be realised. For measuring the degree of an agent's beliefs and the relative strength of his desires Ramsey invented a method based solely on the behaviour of the agent's showing preference or indifference in gambles. The method uses the notion of an ethically neutral condition, that is a condition whose truth or falsity does not affect the strength of the agent's desires. In addition it offers a measurement of degrees of belief and relative strength of desire by checking indifference of the agent between specified options.

The method, in more detail, proceeds as follows: the first step is to find an ethically neutral condition to which the agent assigns probability $\frac{1}{2}$. An ethically neutral condition p is assigned subjective probability $\frac{1}{2}$ by an agent in the case that the agent, while showing preference between a and b , he shows indifference between the options i) a if p is true, b if p is false and ii) b if p is true, a if p is false. An axiom of Ramsey's theory establishes that if this is the case for a pair of options a , b , it is true of all such pairs. The intuition is that an agent has a belief with degree $\frac{1}{2}$ if he is indifferent in any bet made upon it. Having offered a way for determining the degree $\frac{1}{2}$ of a belief, Ramsey's method proceeds with measurement of the relative strength of desire. A definition is given for four options to have equal difference in value. The difference in value between a and b is equal to the difference in value between c and d , if the agent is indifferent between the options i) a if p is true, d if p is false and ii) b if p is true, c if p is false, where p is an ethically neutral proposition believed to the

degree $\frac{1}{2}$. This definition enables a correlation of values with real numbers and it makes it possible to calibrate the desirability scale between consequences a and b, where b is preferred to a. Having defined the difference in value between any two consequences, Ramsey uses the intuition according to which the degree of belief is defined “by the odds at which the subject would bet on p, the bet being conducted in terms of differences of value as defined” (op. cit p. 76). This intuition is translated as follows: if an agent is indifferent between the options i) a for certain and ii) b if p is true and c if p is false, the agent’s degree of belief in p is the ratio of the difference in value between a and c to that between b and c. After this, it is possible to define the degree of belief in p given q. This measurement is based, as before, on indifference shown by the agent between specified options.

A definition of coherency that does not assume the independent utility of money could then be given along the following line: given a measurement of subjective utility of a subject, the subject has coherent evaluations if it is not possible to be presented with a book in which he would lose whatever happens. “Lose” here is defined relative to the subjective utility measurement.

The centrality of a notion of coherency in probability theory is due to the fact that it enables the representation of beliefs of a coherent subject by probabilities. This possibility is proved by a theorem of coherency. Roughly speaking, a theorem of coherency says that a subject is coherent if and only if his beliefs are represented by probabilities (and his utilities by real numbers satisfying certain conditions). Beliefs of a subject that would not permit the possibility of a Dutch book being made against the subject can be represented by probabilities; that is, by real numbers that satisfy the Kolmogorov axioms of Probability⁵⁵.

I will discuss the importance of representation theorems of probabilities in relation to the problem of classifying the doxastic states of an individual later in this chapter, in the context of Savage’s theory. Here, I make one comment on Ramsey’s and de Finetti’s theories.

⁵⁵ The Kolmogorov Axioms of Probability are:

1. The probability of an event E, P(E) is non-negative.
2. If S denotes the set of all possible events, then P(S) is equal to 1.
3. If two events E1 and E2 are mutually exclusive, then the probability that at least one of them occurs is the sum of the individual probabilities P(E1)+P(E2).

A theory like de Finetti's or Ramsey's results in a measurement of degrees of belief of an individual on the basis of betting behaviour of the individual. This seems to be a drawback of these theories for the possibility of using such decision theoretic considerations as those for a theory of understanding speech and behaviour of agents. In order to make use of a theory like de Finetti's or Ramsey's for measurement of degrees of belief of an individual (and of degrees of desire in the case of Ramsey's theory), the individual is invited to show preference or indifference between gambles. In addition, the individual has to accept this invitation. Now, the fact is that gambles or bets, in a literary sense, are artificial constructions. They are artificial constructions of situations in which a connection between a condition or conditions in the world and rewards (outcomes of the gamble) are established. These connections are established by the inventor of the gamble. Application of Ramsey's theory for subjective probability measurement would require essential use of gambles or bets envisaged as artificial constructions of situations between which the agent is invited to show preference or indifference. It would also, and more importantly, require the assumption that the agent does not show aversion to risk behaviour. The requirement of an essential use of gambles or bets envisaged in this way is restrictive as far as the possibility of accommodating the subjective probability theory by a theory of understanding behaviour of agents. Nevertheless, the theories can prove to be useful to an interpretation theory, if modified. The relation between decision theory and interpretation theory will be discussed in more detail in the following chapter. The modification of De Finetti's and Ramsey's use of the notion of a gamble that will be considered here is one suggested in Savage's axiomatised theory of decision. The sense in which gambles are used in Savage's theory is slightly different from the way gambles are used in de Finetti's or Ramsey's theories. In Savage's theory there is no use of an ethically neutral condition and of gambles constructed on an ethically neutral condition. In Savage's theory gambles are not artificial constructions of situations; they are 'gambles against nature'. The central notion of the theory, the notion of an act relates to decision situations individuals are confronted with in making choices in real life cases; an act resembles a gamble in abstraction only.

6.3 Savage's axiomatised decision theory

I continue then, by discussing the basic aspects of Savage's (1954) decision theory. Savage's theory is in some respects similar to Ramsey's. In both theories a subjectivist view of probability is taken, in both theories there is the assumption that degree of belief is measured on the basis of overt behaviour and both theories imply similar representation theorems. Savage's theory is differentiated from Ramsey's in the kind of goal the theory is designed to attain. Thus, while Ramsey's main interest was in a theory that would result in a measurement of belief, Savage's goal is an axiomatisation of decision theory, only part of which is a belief-measurement theory. In Savage's theory an axiomatic system concerning the notion of preference between acts is introduced. A relation of qualitative probability between events is introduced on the basis of the axiomatic system as well as a probability measurement of events. On the basis of the probability measurement introduced and by considerations very similar to Von Neumann and Morgenstern (1947) on the preference relation we get a utility measurement. Savage's theory combines de Finetti's theory of probability and Von Neumann's and Morgenstern's theory of utility.

Acts in Savage's theory are considered to be functions from states/events to consequences. Before the presentation of further aspects of the theory let us take an example to see how this suggestion can be comprehended. Let us consider my choice of taking my umbrella with me in going out this morning. That choice of mine is determined by my beliefs about the weather this morning and things that I want or do not want to happen. Taking my umbrella with me this morning is an action that could be seen as a function, as Savage's theory suggests, between states of the weather (events as subsets of states of the world) and states of mine by the following considerations. In case it rains the value of the action-function of taking my umbrella with me is that I do not get wet. In case it rains and it is very windy the value of the action-function of taking my umbrella with me is that I get wet and destroy the umbrella. In case of good weather the value of the action-function of taking my umbrella with me is that I carry a redundant umbrella all the way. Taking an umbrella with me is a function from states/events of the world to consequences. For every possible state of the weather -rainy, windy, good weather- that is, for the possible events of rain, wind and rain and good weather, the act of taking my umbrella will

result in a state of mine - not getting wet, destroying my umbrella, spoiling my walk by carrying a superfluous umbrella with me. In presenting Savage's theory in the following more will be said on the notion of action and on the constraints imposed by the theory on the domain of actions. Let the above comments suffice as a brief introduction into the notion of action as used in Savage's theory.

In Savage's theory there is no explicit use of the notion of gamble. The theory employs a rich terminology of states/events, consequences and acts. The aspect of Savage's theory in which I am interested here is the following: on the basis of conditions on relations between acts, that is, conditions that axiomatically determine the rational pattern of preferences of a subject between acts, it is possible to determine a unique probability function for the subject and a utility function for the subject that is unique up to a positive linear transformation⁵⁶. These functions provide the numerical interpretation of the relevant attitudes of an individual whose preferences exhibit the suggested rational pattern.

It is worth mentioning here that the notion of rationality in the context of decision theory is introduced by conditions that apply to couples of acts or options to be pursued by agents. In most decision theories the basic notion is that of preference, that is a binary relation between acts/options. The axioms of decision theories suggest conditions on the structure of the set of decision alternatives (set of functions, set of propositions, set of sentences) and on the preference relation between options from those sets. The axioms of decision theories are envisaged as criteria for consistent behaviour and rationality of agency. The rationality conditions in the context of the presently discussed decision theories (Savage's and Jeffrey's) delimit the relation of rational preference between acts/options of agents. Thereafter, in the present context, a rational agent is one who is attributed a preference relation between acts/options that satisfies the rationality conditions suggested by the decision theory. In the decision theories we will consider here the notion of a rational agent is consistent with the notion of a coherent agent discussed in the previous section. The rationality conditions we will consider on the binary relation of preference between acts/options of an agent show the agent to have attitudes with degrees that satisfy the Probability

⁵⁶ This means that if V is a desirability function for the subject, any function of the form: $aV+b$, where a, b are real numbers, $a > 0$, is a desirability function for the subject.

axioms.

The discussion of Savage's axiomatised theory of decision will be divided into three parts. First the basic primitives of the theory will be presented. Comments on the significance of their choice will also be made. In the second part of the discussion of Savage's theory the axioms of the theory and comments on them will be presented. In the third part of the discussion of Savage's theory the bearing the results of the theory, that is, the representation theorems of the theory, on a classification of doxastic and affective states of a rational agent will be elaborated.

The basic primitives of Savage's theory are the following:

- i) A set of states (S_i), sets of which are events (E_i). The set of events is a Boolean algebra⁵⁷.
- ii) A finite set of consequences x, y, \dots, z
- ii) Acts, that is functions from states to consequences.

Savage gives a rough and intuitive account of what each of those terms signifies. I, in turn, roughly reproduce it here in order to avoid a formalistic reading of Savage's theory. But it is important to bear in mind that my aim is a more precise account of each one of the basic primitives. That is, the aim is to show whether and how the distinctions between the basic primitives of the theory are substantiated by the theory. The question is whether the axiomatic basis of the theory implies a substantial distinction between states/events and consequences. The discussion here is about the feature of the theory of proving events and consequences to be susceptible of different kinds of characterisations. The difference in the kind of characterisation to which the two categories are susceptible is a difference in the numerical representation of events and consequences. Events, that is sets of states, will be characterised by being measured by a probability function while consequences will be characterised by having a numerical representation in an ordered space of real numbers. A subjectivist decision theory like this suggests that events and consequences are subjected to different kind of considerations by the rational agent; events as represented by a probability function are subjected to doxastic kind of consideration; consequences as represented by a utility function are subjected to a

different kind of consideration. It might be methodologically more appropriate not to use two kinds of term, a kind of term for states-events and a kind of term for consequences, until a distinctive characterisation of each kind is at hand. I will follow Savage in using distinct terminology but I will try to do so without begging the question for a substantial classification.

I continue with giving a rough intuitive account of the basic primitives employed in Savage's theory. The definition of a state (S_i) of the world is, for Savage, that a state is a description of the world.

"Term	Definition
The world	the object about which a person is concerned.
...	...

The **true** state in the world the state that does in fact obtain" (1954, p. 9).

A person who is faced with a particular decision situation is partially concerned with states that may possibly obtain. A state relates to an aspect of the world about which a person is uncertain when he is in a decision situation. The obtaining or not of a state that is relevant to a decision influences the outcome of a decision. Sets of states are events. Use of events instead of states provides in many cases a more convenient way of describing the situation a person is in when confronted with a decision. So, the temperature in Edinburgh being 16 C° this afternoon is a possible state of the world; the temperature in Edinburgh being warm enough for a walk, or being over 16 C° is a possible event. An event contains many possible states of the world. If one of the states an event contains is true, then the event obtains. In Savage's theory events are such that their obtaining or not is independent of the acts considered in the decision situation. This condition that holds in Savage's theory amounts to an independence condition. The independence condition is crucial in Savage's theory. A central notion of Savage's decision theory, namely the notion of preference between acts given an event, (definition 1, see below for details) is not well defined unless the independence condition holds. That is, the probabilities of events that are conditions for deliberation are independent of the acts to be chosen. The condition is one of the points at which Savage's theory is

⁵⁷ The set of events is closed under set theoretic operations.

differentiated from Jeffrey's. A more thorough discussion of the condition will be given shortly.

The second kind of term employed in Savage's theory is that for consequences. Savage's attempt to explain the intuitive meaning of this kind of term is even rougher than that of the case of states.

"In deciding on an act, account must be taken on the possible states of the world and also of the consequences implicit in each act for each possible state of the world. A consequence is anything that may happen to the person." (1954, p. 13)

Taking a consequence to stand for something that might happen to a person is far from sufficient as an account of the term. There are many cases of things of which it can be said that they might happen to a person. Perhaps a happier characterisation of the notion of consequence is that a consequence is brought about by the choices of individuals. So, while states/events are independent of the particular choice an individual makes, a consequence is brought about by the agent and thus it is absolutely dependant on the agent. While Savage's account of the notion of consequence is clearly insufficient, it shows something of the point of introducing terms for consequences and for states. The point of introducing additional terminology is that there are descriptions of the world relevant to a decision situation not subjected to doxastic considerations. A person is confronted with a decision situation as a potential agent, not as a passive recipient. A person who considers taking a course of action considers states of the world that condition the courses of action as being likely to occur. In addition the person considers states that he might bring about by taking that course of action. These states enter into a description of the decision situation the person is in when, in considering the taking of one of the available courses of action, the states are considered as consequences of the available courses of action.

The kind of consideration of the agent to which these states are subjected contributes to the decision making. Intuitively this kind of consideration is different to the kind of consideration to which events are subjected. Consequences are more liked or disliked, more or less desirable, more or less good. Events are more or less likely or probable to occur. The theory is invited to make this difference explicit and precise.

In stressing a difference between events and consequences I repeat here the following. While in Savage's theory events entering into a decision situation are such that their obtaining or not is independent of the particular action to be performed as a result of the decision making, consequences enter into a decision situation as absolutely dependent on the actions considered. The dependency of a consequence on action is an implicit one. As will be explained shortly, acts in Savage's theory are functions from states to consequences. At this point we see another characteristic of Savage's decision theory that marks a difference to Jeffrey's decision theory. That is, in Savage's theory there is act-consequence determinism. For every state/event acts determine consequences.

In Savage's theory a consequence is related to an act. An act is a function from possible states of the world to consequences. Here is how Savage introduces the notion:

"If two different acts had the same consequences in every state of the world, there would, from the present point of view be no point in considering them two different acts at all". (1954, p. 14)

An act may therefore be identified by its consequences in states. Or, more formally,

"an act is a function attaching a consequence to each state of the world". (1954, p. 14).

I think these remarks require further clarification. What is 'the present point of view', from which two different acts with the same consequences in every state of the world are considered not to be different? I think the point of view Savage takes is that the aspects of the acts about which decision theory is concerned are the ones represented by descriptions of decision situations. These are aspects of an action minimally sufficient for showing the agent who performs the action to be a rational agent. If all the beliefs I have about the weather this morning are that it is more probable that it is going to rain than not and if all my desires conditioned upon my weather beliefs are that I desire not to get wet, the present point of view identifies my choice of taking my umbrella with the choice of taking my anorak. According to my beliefs both choices will keep me dry either it rains or not. If, in addition, I have more beliefs about the weather this morning, say that it is more probable that it is going to be windy and rainy than calm and rainy, then, from the present point of view, the choice of taking my umbrella differs from taking my anorak, if walking with my

umbrella in the rain and wind is less desirable from walking in my anorak in the rain and the wind. The decision situation represented by a decision matrix in which only beliefs about rainy weather enter shows the agent who is indifferent to either option (umbrella or anorak) to be rational. A decision matrix with richer beliefs about the weather might show the agent who makes the one choice but not the other to be rational.

An example used by Savage for illustrating the relations in his terminological classifications is the famous omelette example. The example is the following: your spouse has broken five good eggs into a bowl and you decide to finish the omelette. There is a sixth unbroken egg at your disposal and you might decide to break it into the bowl or waste it altogether. The possible states relevant to the particular decision situation are two; either the egg is good or the egg is rotten. There are three acts among which you can make a choice. Savage uses table 1 to exhibit the situation. A description of a decision situation can be given by a decision matrix like the one presented by table 1.

Table 1

Act	State	
	Good	Rotten
Break into bowl	six-egg omelette	no omelette, and five good eggs destroyed
Break into saucer	six-egg omelette and a saucer to wash	five-egg omelette, and a saucer to wash
Throw away	Five-egg omelette and one good egg destroyed	five-egg omelette

Savage's theory accommodates decision problems that can be represented by a matrix like table 1. Decision problems that exhibit act-consequence determinism and event/state independence of act have a representation like table 1. The requirement on behalf of the theory that every decision problem needs to have a representation like table 1 has been subject to criticism. Considering decision problems that do not exhibit act-consequence determinism and event independence of acts would threaten the power of the theory. Decision problems that fail to exhibit those two characteristics do not have a representation like table 1 and thus cannot, in a straightforward way, be accommodated by Savage's decision theory. For the sake of making the presentation of the theory more comprehensive I offer the following general and brief discussion on the issue. Because my purpose here is to suggest Savage's theory I do not offer a detailed analysis of the alternative proposals I mention.

Both requirements of event-act independence and act-consequence determinism have been challenged and alternative theories have been suggested. The act-consequence determinism requires that there not be a probability distribution over the consequences of an act. For every state of the world each act determines one consequence. The requirement sounds exaggerated in considering some cases of decision. Savage considers the situation in which the agent might not know with certainty the consequence of an act in a state. For instance in the case of the omelette example described above it might be that one rotten egg results in a good omelette after all. The act of using the available egg in the case that the egg is rotten does not determine the consequence of a destroyed omelette. It results with some probability in a destroyed omelette and with some probability in a good omelette.

There is a suggestion made by Savage for reformulating the problem without giving up determinism. The suggestion is that the decision problem be redescribed so that one bad egg destroys the omelette and that one bad egg does not destroy the omelette are considered to be two more events. Thus, in the case that one bad egg destroys a five-egg omelette, the act of using a bad egg results in a destroyed omelette. In the case that one bad egg does not destroy a five-egg omelette, the act of using the egg results in a six-egg omelette. The debate concerning the acceptance of the act-consequence determinism requirement centres round the possibility of the

suggested reformulation in highly complex decision situations. So, Fishburn and Balch (1974) challenge the act-consequence determinism requirement on the ground that in highly complex decision situations a reformulation of the problem along the suggested line is not feasible. Fishburn (1970) proposes an alternative axiomatic decision theory, but for the particular requirement in every respect similar to Savage's. But, independently of the problem of the feasibility of reformulating the decision problem in highly complex situations, it is worth-noting that such reformulation is in principle possible.

The other requirement for an accommodation of decision problems by Savage's theory is that events are independent of the act to be chosen. ($P(E|A)=P(E)$ for an event E and consequence A that are entries in the same decision situation.) There are cases of decision situations in which the requirement of act-independence of events sounds plausible and straightforward and cases that it does not. In cases like the omelette example or like the act of taking my umbrella with me it is straightforward to assume that the state of the egg or the state of the weather are independent of the act chosen. Using the egg or taking my umbrella with me will not make it more or less likely that the egg is rotten or that the weather will be good. In contrast, there are decision cases in which the assumption of a dependency between act and state is motivated. For example, the state in which an agent-smoker lives a long life is dependent on the choice of continuing smoking rather than quitting smoking. In this case the probability of the state of living a long life is not independent on the choice of the agent. The point here touches on the debate about causal and evidential decision theories and it is beyond the limits of this dissertation to make justice to it⁵⁸. As a way of closing here the brief introduction to the significance of the determinism and independence requirement let me mention that, with an appropriate reformulation, every decision problem can be turned into one exhibiting both determinism and independence. In Jeffrey (1976) we find a more detailed discussion about how this can be done. Just for giving a clue, the idea applied to the smoking vs long life example mentioned above is that we reformulate the decision problem by redefining the states. So, the decision situation in which the states are those of living long and

⁵⁸ The relevance of the smoking example to causal considerations in the notion of deliberation is based on R. A. Fisher's (1959) interpretation of the case. The case, according to this interpretation, turns out to be a Newcomb-like situation. Some details are discussed in chapter 7.

dying early is replaced by one with the states of <living long and smoking, living long and quitting> and <dying early and smoking, dying early and quitting>. In Jeffrey (1976) we find that the reformulated problem is equivalent to the old one while both assumptions of determinism and independence hold⁵⁹.

I finish here my introduction to the primitives of Savage's theory. I continue with the presentation of the axioms of his theory. The axiomatic system of Savage's theory is presented in the following. Savage considers a basic relation between acts; the relation of *not preferred to*. Let \leq as occurring between acts stand for this relation⁶⁰. A very rough account of the case of preference had by an agent for an act f over act g is given by the following.

"if he were required to decide between f and g , no other acts being available, he would decide on f ". (1954, p. 17)

First I present some definitions that are used in Savage's axiomatic system. Then I present the axioms of the theory and comment on them.

Definition 1. (preference between two acts given an event): If f, g and f^1, g^1 are acts in F , such that for an event B

1) In $-B$, f agrees with g , and f^1 agrees with g^1 . (The notion of agreement is given by the following condition: for all s in $-B$ $f(s)=g(s)$ and $f^1(s)=g^1(s)$, where equality stands for identity and $-B$ is the complement of B .)

2) In B , f agrees with f^1 and g agrees with g^1 . (For all s in B , $f(s) = f^1(s)$ and $g(s) = g^1(s)$)

3) $f^1 \leq g^1$

Then f is not preferred to g given B ($f \leq g$ given B).

⁵⁹ An interesting discussion concerning the point of the suggested reformulation can be found in Bacharach and Hurley, 1991, p.14-17. A very illuminating discussion of the significance of the independent condition for determining a unified notion of good as located across persons, times and states of nature, can be found in John Broome (1992).

⁶⁰ My discussion on the notion preference concerns interchangeably the relation 'not preferred to' and 'preferred to'. An act f is not preferred to an act g if and only if the act g is preferred or is indifferent to the act f .

Definition 2. (preference between two consequences): for any consequences g, g^1 , $g \leq g^1$ if and only if, when f is an act such that $f(s) = g$ for every s in S and f^1 is an act such that $f^1(s) = g^1$ for every s , $f \leq f^1$.

Definition 2 uses the notion of a constant act. A constant act is one whose consequences are independent of the possible state of world, that is, it has the same consequence in every possible state of the world.

Definition 3. (event A is not more probable than event B): $A \leq B$, if and only if when consequences f, f^1 are such that $f^1 < f$ and acts f_a, f_b are such that

$$f_a(s) = f \text{ for } s \text{ in } A, f_a(s) = f^1 \text{ for } s \text{ in } -A$$

$$f_b(s) = f \text{ for } s \text{ in } B, f_b(s) = f^1 \text{ for } s \text{ in } -B$$

then $f_a \leq f_b$.

Definition 4. (qualitative probability): A relation \leq_p between events is a qualitative probability if and only if,

1) \leq_p is reflexive, transitive and connected.

Reflexivity: For every events A, B if $A \leq_p B$ and $B \leq_p A$ then $A = B$.

Transitivity: for every A, B, C events if $A \leq_p B$ and $B \leq_p C$ then $A \leq_p C$.

Connexity: for every A, B events $A \leq_p B$ or $B \leq_p A$ (or both).

2) $\emptyset \leq_p A \leq_p S$, for any event A and $\emptyset <_p S$.

3) for all events A, B, C if $A \leq_p B$, then $A \cup C \leq_p B \cup C$ when $C \cap (A \cup B) = \emptyset$

Definition 5. (probability measure): A probability measure on a set S (S is the universal event) is a function $P(B)$ attaching to each $B \subseteq S$ a real number such that the Kolmogorov axioms are satisfied:

1) $0 \leq P(B)$ for every B

2) if $B \cap C = \emptyset$, $P(B \cup C) = P(B) + P(C)$

3) $P(S) = 1$

The axiomatic system of Savage's theory is the following:

Savage's axiomatised theory of decision:

Axiom 1. The relation \leq , standing for *is not preferred to* between acts, is a simple ordering.

That is, for every f, g, h in F (F a finite set of acts)

- 1) Either $f \leq g$ or $g \leq f$
- 2) If $f \leq g$ and $g \leq h$, then $f \leq h$.

Axiom 2. If f, g and f^1, g^1 are acts such that:

- 1) in $-B$, f agrees with g , and f^1 agrees with g^1
- 2) in B , f agrees with f^1 and g agrees with g^1
- 3) $f \leq g$

then $f^1 \leq g^1$

Axiom 3. if $f = g, f^1 = g^1$ and B is not null, then $f \leq f^1$ given B , if and only if $g \leq g^1$, where f, f^1 are acts and g, g^1 are consequences.

Axiom 4. : If f, f^1, g, g^1 , are consequences, A, B are events, f_a, f_b, g_a, g_b are acts such that

$$f^1 < f \quad g^1 < g$$

$$f_a(s) = f, g_a(s) = g \text{ for } s \text{ in } A$$

$$f_a(s) = f^1, g_a(s) = g^1 \text{ for } s \text{ in } -A$$

$$f_b(s) = f, g_b(s) = g \text{ for } s \text{ in } B$$

$$f_b(s) = f^1, g_b(s) = g^1 \text{ for } s \text{ in } -B$$

$$f_a \leq f_b$$

then $g_a \leq g_b$.

Axiom 5. There is at least one pair of consequences f, f^1 such that $f^1 < f$.

Axiom 6. If $g < h$, and f is any consequence, then there exists a partition of S [finite number of mutually exclusive and jointly exhaustive events] such that, if g or h is so modified on any one element of the partition as to take the value f at every s there, other values being undisturbed, then the modified g remains less preferable than h .

Axiom 6'. If $B \prec_p C$, there exists $\{C_i, 1 \leq i \leq n\}$, a finite partition of S , the union of each element of which with B is less probable than C , for each i ,

$$1 \leq i \leq n, A \cup C_i \prec_p B \text{ }^{61}.$$

Axiom 7. If $f \leq g(s)$ given B ($g(s) \leq f$ given B) for every $s \in B$, then $f \leq g$ given B ($g \leq f$ given B), where $g(s)$ stands for both the act with a constant consequence and the value of the value of that consequence.

Axiom 1 says that for every pair of acts a rational agent has preference or indifference between them and that preference or indifference is transitive⁶². Axiom 2 along with axiom 3 amount to Savage's *sure-thing principle*⁶³. A *sure-thing principle* says that if an agent prefers an option f to an option g in both cases that an event B obtains and an event $\neg B$ obtains, then the agent prefers option f to option g . In Savage's words and in terms of the relation of not preferred to the principle is the following:

"...If the person would not prefer f to g , either knowing that the event B obtained, or knowing that the event $\neg B$ obtained, then he does not prefer f to g ." (p. 21).

The condition has an intuitive appeal as a rationality condition. In Savage's formulation of the principle reference is made to the not formally introduced notion of knowledge conditions of the obtaining of an event. Axioms 2 and 3 explicate the principle while avoiding reference to the undefined notion of knowledge of the obtaining of an event. Axioms 2 and 3 rely on the notions of preference given an event (as introduced by definition 1) and of preference between consequences (as introduced by definition 2).

Axiom 2 says that preference between acts does not depend on the states on

⁶¹ Axiom 6 makes sense under Theorem 1 (Savage, 1954, p. 32) and Theorem 2 (1954, p. 34) of Savage's theory. Theorem 1 (*op.cit.* p. 32) says that the relation \leq (definition 3) as applied to events is a qualitative probability \leq_p . Theorem 2 (*op.cit.* p. 34) says that if there is an n -fold almost uniform partition of S for arbitrarily large values of n , then there is one and only one probability measure P that almost agrees with \leq_p , where S is the universal event and \leq_p the qualitative probability. (For a definition of a probability measure *almost agreeing with* a qualitative probability see next page.)

⁶² There is extensive literature on the transitivity of preference. The discussion is not about challenging the condition as a rationality condition but about giving an account for the persistent violation of the condition by the agents. See Broome (1991) for a discussion.

⁶³ A *sure-thing principle* belongs to a family of conditions called dominance conditions or independence conditions. That the state that actually obtains be independent of the act chosen is what sustains these

proposition/sentence that I will choose to take my umbrella with me. Indeed, if Jeffrey's theory is not to compromise in generality by introducing distinctions between decisions and choices within propositional contents, the probability of a proposition specifying an action is the probability of the proposition specifying that the action is chosen. 'I take my umbrella with me' and 'I choose to take my umbrella with me' express the same proposition. Since choices are made on the basis of probabilities and desirabilities it seems that we have the result that my choice of taking my umbrella with me this morning depends on how desirable it is to take my umbrella and, surprisingly enough, on how probable it is that I choose to take my umbrella with me this morning. My belief that I will choose a certain course of action features as one of the conditions that explains my choice of the action.

The above example does not exhibit any accidental feature of Jeffrey's theory. Jeffrey's decision theory is an alternative to classical Bayesian theories (e.g. Savage's theory) in which no independence postulate of probabilities (of events) on actions is considered. Subjective probability conditions for deliberation are taken to be dependent on the act chosen. Moreover, the general rule in Jeffrey's theory is that probabilities dependent on the choices are the credence conditions for deliberation.

The point I raise here is that credence of acts featuring as conditions for deliberation seems paradoxical given some pre-theoretic accounts of deliberation. The pre-theoretic accounts of deliberation to which I appeal suggest the following. An agent in deciding which available course of action to pursue compares the available courses of action (on a desirability scale). In order for this comparison to take place the agent appeals to his subjective probability distribution over possible states (which in Jeffrey's theory include the actions). This suggests that the comparison between available courses of action requires a prior subjective probability distribution. On the other hand, in Jeffrey's theory, probabilities of acts featuring as conditions for deliberation seem to prejudge the outcome of deliberation.

The pre-theoretic feature of deliberation conjured up here constitutes the following condition. Namely, the condition is that in a decision situation in which one or more courses of action are envisaged, predictions as to which course will be chosen are not parameters for the decision making. If so, the degree of belief of the agent that he chooses an action does not contribute to the comparison between

actions on a desirability scale that determines the choice of the subject. Though it is not denied that an agent might have degrees of belief that he will take one course of action rather than another and that the agent might make predictions about the choices he will finally make, these are not beliefs relevant to the rationality of the choice of the agent.

The condition appealed here is cited in the literature as the ‘Levi-Spohn embargo’ (Bacharach and Hurley 1991). The requirement that decision theory should respect the above-mentioned pre-theoretic condition concerning the structure of a decision situation is clearly stated in Spohn (1977) and Levi (1991).

“The decision-maker should not assign either determinate or indeterminate credal probabilities to hypotheses as to how he will choose” (Levi, 1991, p. 99).

Spohn comments that the probabilities as to how an agent will choose are irrelevant to the choice of the agent. Levi strengthens this claim by noting the following:

“X might be concerned to predict his own choices, but when he does so, he is not functioning as a deliberate agent concerned to identify which of these options he is not rationally prohibited from making...X might very well take such a predictive explanatory attitude toward the choices of other agents or toward his own choices at future times” (Levi, 1991, p. 99).

In order to illustrate better the appeal of the discussed pre-theoretic condition I present a comment made by Spohn (1977) that shows the strangeness of letting credence of acts feature as conditions for deliberation. The point relies on a condition that constitutes common ground among decision theorists, that is, that subjective probabilities can be detected by the behaviour of the agent in betting situations with certain odds and various stakes. If we let credence of acts feature as conditions for deliberation we can consider betting situations for detecting the subjective probabilities of acts. In such a kind of betting situations the bets offered to the agent would be constructed on the condition that the agent chooses a certain act. This kind of betting situations is strange because all that mattered for accepting the various bets would be the various payoffs. If the payoff of choosing a certain act A is higher than the payoffs of choosing the alternative acts, the subjective probability of the choice A is very close to 1. By changing the payoffs of the same range of choices the subjective probability of A changes⁷⁷.

⁷⁷ It is worth noting here that the pre-theoretic condition that beliefs as to how the agent will choose should not

The condition that credences of acts should not figure as conditions for deliberation is violated in Jeffrey's modified decision theory (Jeffrey 1983) as well. In Jeffrey's modified theory the principle of conditional expected utility is replaced by the principle of ratifiability. *Ratificationism* is suggested in order to deal with Newcomb-like decision problems⁷⁸. A ratifiable decision is defined as a decision to perform an act of maximum estimated desirability relative to the probability matrix the agent thinks he would have if he finally decided to perform that act. *Ratificationism* is the maxim of making ratifiable decisions.

I come now to the second and more serious worry of Jeffrey's decision theory as far as the integration of Jeffrey's decision theory with interpretation theory is concerned. It results from the uniqueness conditions of Jeffrey's decision theory. As we saw in chapter 6, in contrast to classical decision theories (e.g. Savage's theory), the representation theorems of Jeffrey's decision theory do not issue in unique subjective probability and desirability functions for the preference ranking of an agent. The uniqueness conditions of Jeffrey's theory are the following⁷⁹.

P, D, p, d satisfy the Kolmogorov axioms of probability and the desirability axiom of Jeffrey's theory if and only if

For certain real numbers a, b, c, d
ad -bc is positive

enter as credential parameters in the choice-worthiness comparison of acts is generally respected by causal decision theories. It is definitely outside the scope of this dissertation to attempt an introduction of issues from causal decision theories. But it is worth mentioning that causal decision theories are developed with respect to certain structural conditions imposed on decision problems and promise good answers to many Newcomb-like problems. See Gibbard A. and Harper W. (1978), Lewis D. (1981), Skyrms B. (1982). Especially from Skyrms' theory we can raise the interesting question of whether Savage's structural conditions (act-independence of events and consequence determinism from acts) can accept a causal reading. Here, suffice it to say that Skyrms' theory is developed with respect to Savage's structural conditions.

⁷⁸ There are two boxes box1 and box2 and the agent can choose between the content of box1 and contents of both boxes. The agent knows that the box2 contains £1k and that £1m or nothing has been put by a Demon in box1. The conditional probabilities that there is £1m in box1 on choosing box1 and on choosing both boxes are respectively 0.9 and 0.1. According to the principle of maximising conditional expected utility it is rational for the agent to choose box 1. But this result clashes with the principle of dominance, since the choice of both boxes dominates the choice of box1. On the other hand, the ratifiable decision would be the one that since it is performed will have the highest desirability among performance of the alternatives.

The literature on Newcomb-like problems is huge. For a formulation of the problem see Nozick R. (1969). For a general discussion of Newcomb-like problems see Eells (1982). For problems of Jeffrey's theory concerning Newcomb-like decision situation see Gibbard A. and Harper W. *op. cit.*, Lewis D. *op.cit.*, Skyrms B. *op.cit.* For some elaboration of the notion of ratifiability see Skyrms B. (1990).

⁷⁹ The Jeffrey-Bolker axiomatised decision theory yields unique probability and desirability (to a linear transformation) representations in the case that the assignment of values (desirabilities) is unbounded both above and below. See Jeffrey (1983, p. ch. 6) for details.

$\text{des}A + d$ is positive

$c \text{ des } T + d = 1$

$\text{PROB } A = \text{prob } A (c \text{ des } A + d)$

$\text{DESA} = (a \text{ des } A + b) / (c \text{ des } A + d).$

That is, from the relation of preference between propositions as determined by Axioms 1-4 of Jeffrey's axiomatised theory of decision the theoretical constructs of subjective probability and desirability are not uniquely determined. For every pair prob, des , that satisfies the existence condition, i.e. prob is a probability and des satisfies the desirability axiom, such that des preserves the preference ranking, another pair PROB, DES is determined satisfying the existence and the uniqueness condition above such that DES preserves the preference ranking. The family of such pairs is infinite since for the pair PROB, DES there is another pair $\text{PROB}^1, \text{DES}^1$ that satisfies the existence and uniqueness condition with respect to the same preference ranking.

Let us now discuss the significance of this feature of Jeffrey's theory for the integrated Jeffrey-style radical decision theory. I remind the reader that decision theory is invited to contribute to radical interpretation theory by introducing subjective probability and desirability functions. Within a theory of interpretation the subjective probability function is used for determining relations of evidential support between sentences required for interpretation of sentences containing theoretical terms and interpretation of sentences not directly keyed to observation. The different representations of attitudes by the two functions enable a distinction between doxastic and affective attitudes of an agent. Since in Jeffrey's theory the probability function is not uniquely determined, but it belongs to an infinite family of probability functions as suggested by the uniqueness condition, the subjective probability function that in a Jeffrey-style radical decision theory mirrors the relations of evidential support between sentences becomes indeterminate. From the relation of preference between sentences that is the empirical primitive of the theory the evidential relations between the doxastic attitudes/sentences are not uniquely determined. It is left undetermined from preference rankings which probability function from the infinite family of probabilities that respect those rankings determines the relations of evidential support between the attitudes of the agent. The relevant indeterminacy is bequeathed to the

subjective conditional probability functions for the agent. That is, in Jeffrey's theory the preference ranking does not determine which one of the infinite probs would be used in every instance of applying *Bayes theorem*. This would result in an enormous indeterminacy concerning interpretation of sentences containing theoretical terms and sentences less directly keyed to observation.

Davidson (1990, fn pp. 323-324) comments that the feature of a Jeffrey-style radical decision theory of not yielding unique subjective probability representations from preferences is acceptable as another sort of indeterminacy that already exists and is expected in radical interpretation theory. But diminishing indeterminacy is a desirable property of interpretation theory. If so, an alternative radical decision theory that would result in unique probability representations of doxastic attitudes would be preferable to a radical decision theory that like a Jeffrey-style radical decision theory results in an infinite number of doxastic representations.

7.3 The Savage-style radical decision theory

Let us now consider the possibility of an integration of Savage's decision theory with radical interpretation theory. In a way similar to the way we considered a Jeffrey-style radical decision theory, i.e. as Jeffrey's decision theory in which the identity of propositions is not granted, a Savage-style radical decision theory will be Savage's decision theory in which the identity of acts is not granted. The identity of acts, which in Savage's theory are seen as functions from states/events to consequences is not in the data of the theory. The options among which the agent shows preference or indifference are represented by matrices in which the entries are uninterpreted sentences.

I start my discussion of the Savage-style radical decision theory by considering Davidson's and Jeffrey's objections. Davidson formulates his objection against Ramsey's theory and the essential use of gambles in that theory. Gambles are complex entities and they are difficult for the interpreter to identify. A gamble, though it can be represented by a sentence/proposition, involves a relation between sentences/propositions. The objection has it that the relation between sentences/propositions involved in gambles is causal. This is an aspect of gambles to which both Davidson (1984, 1990) and Jeffrey (1965, 1983) object. A gamble

specifies causal connections between a condition (the coin lands head up) and more or less desirable outcomes (you win £100, you lose £100). Since, in addition, Ramsey's theory requires the existence of arbitrary gambles, the theory would require attributions of arbitrary causal connections between the attitudes of agents. For instance, take the case that among the consequences are that there will be a thermonuclear war next week and that there will be fine weather next week. And among the conditions is that a coin lands head up. In this case, the agent needs to have preference for the gamble that there will be a thermonuclear war next week if the coin lands head up and there will be fine weather next week if not. In order for the agent to entertain this gamble he needs, the objection goes, to consider causal relations between the condition and the consequences. This is weird, because it seems to suggest that the agent needs to revise his beliefs about the causes of a war (Jeffrey, 1983, p. 157).

Another objection to Ramsey's theory, coming from Davidson, is that use of gambles in Ramsey's theory requires the existence of ethically neutral conditions; that is conditions whose obtaining or not has no value in itself. For an incorporation of Ramsey's theory into a radical perspective identification of ethically neutral conditions would be required.

Here we will discuss the objections by seeing how they can be reformulated so that they apply to Savage's theory. Since acts in Savage's theory are the structural equivalent of gambles in Ramsey's theory, the reformulation would go along the following line. Acts, as functions from states/events to consequences are complex entities embodying causal relations. Acts can be represented by complex sentences that specify causal relations between events and consequences. Events have to be identified as ethically neutral; the bearers of utility, acts/consequences and the bearers of subjective probability constitute distinct classes.

Let us first discuss the causal connection worry. The conditions that determine the notion of gamble in Ramsey's theory and the notion of act in Savage's theory do not need to involve causal connectives⁸⁰. Acts in Savage's theory are seen as functions between events and consequences so that the two following conditions

⁸⁰ For a non causal formulation of Ramsey's gambles see Levi I. (1966).

hold:

1. Act-consequence determinism: where C_i is consequence A is act and E_i an event such that $A(E_i)=C_i$,

$$P(C_i|A \cap E_i)=1$$

2. Act independence of events: where A is act and E is event,

$$P(E|A)=P(E)$$

Conditions 1) and 2) do not require causal connectives. So, the part of the Jeffrey-Davidson objection against gambles as notions that involve causal relations does not hold⁸¹.

Let us now discuss the objection against use of gambles/acts based on their complexity. The point I think we should consider in relation to this objection is that the complexity of the notion of gamble in Ramsey's theory and of the notion of act in Savage's theory mirrors the structural complexity of decision situations. The motivation for introducing structural conditions to the notion of act is that the structural complexity of the options corresponds to the structure of decision situations. The structural conditions inherent in the notion of act intend to grasp (more or less effectively) a distinction within the parameters of deliberation. The doxastic parameters that are outside the influence of the actions in question and the affective parameters that are absolutely dependent on the actions. The suggestion here is that a distinction between those two parameters of the notion of action mirrors a structural distinction in the notion of deliberation. If so, the complexity of Ramsey's gambles and of Savage's acts is not unmotivated.

The last thing we need to show in order to defend our proposal for a Savage-style radical decision theory is how Savage's theory can accommodate a radical perspective. In presenting Davidson's suggestion for a Jeffrey-style radical decision theory we saw how a Jeffrey-style radical decision theory is possible. This was shown by delivering the logical connectives of the language (Sheffer stroke) from the relation of preference or indifference between (uninterpreted) sentences.

There is a comment that needs to be made at this point. The possibility for Jeffrey's theory of accommodating a radical perspective was not shown on the basis

⁸¹ The observation that gambles/acts can be accounted without use of causal connectives does not clash with a comment made in the previous section, namely that Savage's theory can accommodate a causal reading.

of actual choices only. Davidson shows how the logical connectives of the language can be determined in a radical way by showing how the Sheffer stroke can be determined on the basis of preference or indifference between uninterpreted sentences (fn 76). The pattern of preferences between uninterpreted sentences that would determine that the sentences are composed by the Sheffer stroke involves both actual and counterfactual choices of the agent. The point of the suggestion is to show that from the relation of preference or indifference between uninterpreted sentences (the empirical primitive) the structural condition of Jeffrey's decision theory -objects of belief, of desire and preference form an atom free complete Boolean algebra- can be approached. The analogous argument required for showing the possibility of Savage's theory to accommodate a radical perspective would have to show that the structural conditions of Savage's decision theory can be tackled on the basis of the relation of preference or indifference between uninterpreted acts. The objects of preferences will, in the context of a Savage-style radical decision theory, be represented by matrices in which the entries are uninterpreted sentences.

The act-consequence determinism condition of Savage's theory can be addressed by representing acts as functions between sentences. The other structural condition of Savage's theory is the event independence of acts. In order to address from preference or indifference the independent condition we need to show that from preference or indifference we can determine that one sentence is irrelevant to another. That is, what is required for showing that independence structural conditions of Savage's theory holds is to be able to tell from a relation of preference or indifference only that the condition $P(E|A) = P(E)$ holds for two sentences A, E. Since a subjective probability function is not available in advance of the theory we will have to show that the irrelevance condition can be approached by the relation of qualitative probability. We need to be able to show from a relation of preference or indifference that the events (sentences figuring as arguments of act functions) E given A and E are equivalent with respect to qualitative probability.

The general task of showing two events (sentences) to be independent is not straightforward from the resources of Savage's theory. The reason is that it is not straightforward how to determine a personal probability relation 'not more probable than' between (B given C) and (G given H) for any four events B, C, G, H. The relation of preference between acts f, g given an event B that determines the relation

'not more probable than' between event A given B and event C given B is defined only for the same event B⁸².

In order to deal with this problem of a Savage-style radical decision theory we will appeal to the notion of conditional act and to Axiom 2' that introduces an extended notion of preference between conditional acts. In chapter 6 we saw how a notion of qualitative independence can be defined from the notion of preference between conditional acts. Since this condition is granted, it is possible to determine a notion of independence between (uninterpreted) sentences from the notion of preference between (uninterpreted) conditional acts.

The conditions to which we appeal for addressing the independence structural condition of Savage's decision theory requires the identification of logical connectives between sentences. For this problem I suggest we use a modification of Davidson's suggestion for determining the logical connectives of the Boolean algebra of propositions/sentences in the context of a Jeffrey-style radical decision theory. The modification of Davidson's suggestion about how to determine the Sheffer stroke (not both) from a relation of preference or indifference that needs to be made here would take as objects of preference or indifference functions between composed sentences (A|B, '|' the Sheffer stroke here) and appropriate consequences. These considerations intend to show the possibility for Savage's decision theory for accommodating a radical perspective. From the relation of preference or indifference between uninterpreted acts the structural conditions of Savage's decision theory can be approached.

7.4 A comparison between Jeffrey's and Savage's radical decision theories

In the two previous sections of this chapter the possibility of a Jeffrey-style radical decision theory and the possibility of a Savage-style radical decision theory were presented and discussed. Here I present a comparison between the two suggestions.

In section 7.1 of the present chapter we discussed the kind of contribution decision theoretic considerations are required to make to a radical interpretation theory. One respect in which decision theory is relevant to interpretation theory is that

⁸² The problem of qualitative probability with independence has been approached by B. O. Koopman (1940).

decision theoretic considerations can be used to elucidate pre-theoretical concepts like belief, desire and preference that are employed in understanding actions of speakers/agents. On this ground it has been suggested that the degree of affinity of the introducing notions of a decision theory to their pre-theoretic counterparts is relevant to the assessment of a decision theory. The assessment concerns the incorporation of a theory of decision to a theory of understanding. In discussing a Jeffrey-style radical decision theory we saw that the theory disregards some pre-theoretic intuitions about the notion of deliberation. These pre-theoretic intuitions constitute a condition that was named 'the Levi-Spohn embargo'. That is that credences of acts should not figure as conditions for deliberation. In contrast to Jeffrey's decision theory, Savage's decision theory respects the condition. The credences that constitute parameters for deliberation are independent of the acts to be chosen. Savage's theory respects the particular condition and in the relevant respect its introducing notion of deliberation - along with its belief, desire and preference components- has a greater degree of affinity than Jeffrey's theory to its pre-theoretic counterpart.

Another respect in which decision theoretic considerations are relevant to interpretation theory is that decision theories introduce probability and utility functions that can be used in order to cope with certain problems within interpretation theory. These problems are the determination of relations of evidential support between sentences required for interpretation of some sentences and a classification of the doxastic and affective attitudes of speakers/agents required for the applicability of the principle of charity and for the rational embedding of the attitude attributions. With respect to the first problem we saw that a Jeffrey-style radical decision theory introduces an enormous indeterminacy concerning the evidential relations between sentences for a subject that has a certain pattern of preferences. The subjective probability function that would determine the relations of evidential support between sentences is not uniquely determined in Jeffrey's theory of decision. On the other hand Savage's decision theory introduces a unique subjective probability function. The theory yields for an agent who has preference relations that conform to the rationality constraints of the theory between options that conform to the structural

constraints of the theory a unique subjective probability function. Such an agent, Savage's theory suggests, has a unique pattern of evidential relations that characterises his doxastic system of attitudes.

Another point of comparison between a Jeffrey-style radical decision theory and a Savage-style radical decision theory is the common content assumption. Jeffrey's theory requires the common content assumption. The structural condition of Jeffrey's decision theory is that belief, desire and preference are attitudes towards propositions. The domain of the objects of belief, desire and preference in Jeffrey's theory constitutes a complete, atom-free Boolean algebra (minus the null element). On the other hand, Savage's theory is liberal concerning the assumption. In Savage's theory the assumption for the domain of events is that it constitutes a Boolean algebra, while the assumption for the domain of consequences is that it is a finite ordered set. Subjects to doxastic-probabilistic considerations are events while subjects to value considerations indicated by the preference ordering of utilities are consequences. In addition, acts, which are functions from events to consequences, are subjects to expected value considerations indicated by expected utility preference ordering. The expected utility considerations that characterise the domain of act-functions embody both doxastic-probabilistic considerations –inherited from the doxastic component of acts which are events- and utility considerations –inherited from the value component of acts which are consequences.

Now in Chapters 2 and 5 of the dissertation it was pointed out that the common content assumption between doxastic and affective speech acts (chapter 2) and between doxastic and affective attitudes/states (chapter 5) could be challenged. We considered some conditions for identifying content of utterance/propositional attitude and it was pointed out that those conditions do not grant the common content assumption. Needless to say, the common content assumption is a useful theoretical tool for theories of understanding language and mind of agents that results in elegant theories of interpretation. The attempt in the dissertation has been to point out than an assumption considered uncontroversial and made almost uniformly by propositional attitude and speech act theorist might not be properly established. Taking those considerations into account suggests that Jeffrey's decision theory would not be compelling on the ground of the common content assumption solely. Thus, the aspect of Jeffrey's theory of taking the objects of belief, desire and

preference to constitute a single domain does not, on its own right, present us with a compelling reason in favour of the suggestion of a Jeffrey-style radical decision theory.

In discussing the logical analysis of attitude attribution sentences (chapter 4, 4.3) we suggested that we let a substantial theory about attitude attributions determine the common content condition across doxastic and affective attributions. The substantial theories of attitude attributions considered in the dissertation are elaborations of decision theoretic considerations. We considered Jeffrey's and Savage's theories of decision for an account of belief and desire attributions. In this chapter we presented some reasons for which Savage's decision theory is preferable to Jeffrey's theory as a candidate decision theory to be incorporated with radical interpretation. The reasons for the particular proposal are that Savage's theory does not violate the 'Levi-Spohn embargo' and it results in a unique subjective probability function for the agent/speaker. Now, Savage's theory does not require the common content assumption. Furthermore, concerning probability and utility representations, it suggests that the domain of doxastic-probabilistic considerations (events) and the domain of value-utility considerations (consequences) are not isomorphic. And we pointed out that there are certain structural differences between the domain of events and the domain of conditional acts. However these considerations do not constitute a proof that in Savage's theory the common content assumption is false. It might be possible, while keeping the spirit of Savage's theory (act-independence of events and act-consequence determinism) to modify Savage's theory so that it implies the common content assumption. In this case, given that Savage's theory is preferable to the alternative Jeffrey's theory, we will be offered with more comprehensive reasons for accepting the common content assumption. A substantial theory of attitudes would imply it. In case it is shown that such a modification of Savage's theory cannot be made and the theory implies that the common content assumption is false, we will have to distinguish between two kinds of content. In this case Savage's decision theory would ground the project interpretation theory in which there would be a distinction between two kinds of theoretical entities corresponding to attitude attributions. A substantial theory of attitudes would suggest that in a theory of interpretation two kinds of content are introduced. In case that, for reasons independent of the issue of the common content assumption in Savage's theory, a

proof of the assumption is offered, then acceptance or rejection of Savage's theory will be decided on whether it can accommodate the assumption or not.

7.5 Some closing remarks

The background suggestion adopted in the dissertation is that a theory of interpretation results in redescrptions of utterances and doings of speaker/agents. The utterance of a speaker "Gib mir wasser" is redescrbed as a certain saying. It is redescrbed as a saying to the hearer to be given water. A speaker utters the sentence "Es schneit". The utterance is redescrbed as a saying that it is snowing.

Redescrptions of this kind satisfying certain conditions enable understanding of the utterances of speakers. A condition that redescrptions need to satisfy is that the attitudes ascribed to the speaker by those redescrptions show the speaker to be a rational. They are attitudes that could be part of a system of attitudes that would show the preferences of the speaker/agent to satisfy certain constraints of rationality. The rationality constraints considered here on preferences are introduced by decision theoretic considerations.

The general problem that pertains the dissertation is the possibility of a distinction between truth-evaluable and non-truth-evaluable utterances in the context of radical interpretation theory. It is pointed out that the possibility of a distinction between truth-evaluable and non-truth-evaluable utterances is a substantial problem for a theory of radical interpretation. One projection of this problem is the issue of the applicability of the principle of charity. Another projection of the problem is the rational embedding of utterances. The suggestion made in this dissertation is that the required distinction could be given on the basis of a classification within the attitudes attributed to an agent. The classification within the attitude attributions is one between doxastic and affective attitudes. Utterances related to doxastic attitudes – expressing or ascribing a doxastic attitude- are classified as truth-evaluable utterances. Utterances related to affective attitudes – expressing or ascribing an affective attitude- are classified as non-truth-evaluable utterances.

A theory of interpretation that results in a redescrption of the utterance "Gib mir wasser" that would ascribe an affective attitude to the speaker related to the utterance would classify the utterance "Gib mir wasser" as one not susceptible of

truth evaluation. This utterance is one beyond the domain of applicability of the principle of charity. A theory of interpretation that suggests a redescription of the utterance “Es Schneit” ascribing a doxastic attitude to the speaker in making the utterance would classify the utterance as one susceptible of truth-evaluation. The theory that introduces doxastic classifications considered here introduces degrees of doxastic attitudes. In case the redescription ascribes a doxastic attitude with degree 1 (or close enough to 1) the utterance is classified as one within the domain of applicability of the principle of charity.

Needless to say there are many problems that are left unanswered in this dissertation. The detailed workings of the principle of charity were not discussed here. Moreover the workings of what was called *the extended principle of charity*, were left completely out of consideration. The question of whether a modification of Tarskian mechanisms could be possible for the value-evaluable (non-truth-evaluable) utterances was left out of consideration. And though the difference between doxastic and affective content were stressed, the systematic connections between them were not discussed. Adequate answers to those questions are required for a complete assessment of the views presented in the dissertation. In particular, answers to those questions are required for showing the adequacy of a Savage-style radical decision theory as it was introduced and interpreted in the context of this dissertation. The attempt that has been made here is to show the possibility of an alternative radical decision theory that promises good answers to some problems facing a theory of radical interpretation.

REFERENCES

- ADAMS, E. & ROSENKRANTZ, R.D. 1980. "Applying the Jeffery decision model to rational betting and information acquisition", *Theory and Decision*, 12, 1-20.
- ANSCOMBE, E. 1957. *Intention*. Oxford: Blackwell.
- ARMON-JONES, C. 1991. *Varieties of Affect*. Harvester Wheatsheaf.
- AUSTIN, J. L. 1962. *How to do things with words*. Oxford University Press.
- BACCHARAH, M. and HURLEY, S. (eds.). 1991. *Foundations of Decision Theory*. Basil Blackwell.
- BALCH, M / McFADDEN, D. L./ WU, S. Y. (eds.). 1974. *Essays on Economic Behaviour under Uncertainty*. North-Holland| American Elsevier.
- BALCH, M. and FISHBURN P. 1974. "Subjective Expected Utility for Conditional Primitives". In BALCH, M/ McFADDEN, D. L./ WU, S.Y. eds. 1974. pp. 57-69.
- BLACK, M. 1970. *Margins of Precision, Essays in Logic and Language*, Cornell University Press, Ithaca and London.
- BLACKBURN K., W. 1987. "Davidson on Force and Convention", *Analysis* 47, pp. 72-74.
- BLACKBURN, S. 1975. "The Identity of Propositions", in Blackburn, S., 1975. pp.182-205.
- BLACKBURN, S., (ed.). 1975. *Meaning, Reference and Necessity*, Cambridge, New York: Cambridge University Press.
- BLOCK, N., (ed.). 1980. *Readings in Philosophy of Psychology*, vol.1 Cambridge, Mass.: Harvard University Press.
- BOLKER, E. D. 1965. *Functions Resembling Quotients of Measures*, Ph.D. Dissertation, Harvard University.
- BOLKER, E. D. 1967. "A simultaneous axiomatisation of utility and subjective probability. *Philosophy of Science*, 34, 333-340.
- BROOME, J. 1991. *Weighing Goods*, Basil Blackwell.

CARNAP, R. 1947 and 1956. *Meaning and Necessity, A study in Semantics and modal logic*, The University of Chicago Press, Phoenix Edition.

CARNAP, R. 1951. *Logical Foundation of Probability*. London: Routledge and Kegan.

COLLINS COBUILD. 1990. *English Grammar*. Collins Publishers, The University of Birmingham.

CHURCH, A. 1950. "On Carnap's Analysis of Statements of Assertion and Belief", *Analysis* 10, pp. 97-99.

DAVIDSON, D. 1968. "On Saying That", in Davidson 1984, pp. 93-108.

DAVIDSON, D. 1970. "Mental Events", in Davidson 1980, pp.207-227.

DAVIDSON, D. and HARMAN, G. (eds.). 1972. *Semantics of Natural Language*, Holland. D. Reidel Publishing Company/Dordrecht.

DAVIDSON, D., 1973a. "Radical Interpretation", in Davidson 1984, pp.125-154.

DAVIDSON, D., 1973b. "In Defence of Convention T", in Davidson 1984, pp. 65-75.

DAVIDSON, D.1974a. "Belief and the Basis of Meaning", in Davidson 1984, pp.141-154.

DAVIDSON, D., 1974b. "On the Very Idea of a Conceptual Scheme", in Davidson 1984, pp.183-198.

DAVIDSON, D. 1975. "Thought and Talk", in Davidson 1984.

DAVIDSON, D. 1976. "Reply to Foster", in Davidson 1984.

DAVIDSON, D. 1979. "Moods and Performances", in Davidson 1984, pp. 109-121.

DAVIDSON, D. 1980. *Essays on Actions and Events*, Oxford Clarendon Press, Oxford.

DAVIDSON, D. 1982a. "Communication and Convention", in Davidson 1984, pp.265-280.

DAVIDSON, D. 1982 b. "Paradoxes of Irrationality" in Hopkins, J. and Wollheim, R. (eds.). 1982.

DAVIDSON, D. 1982c. "Expressing Evaluations", The E. H. Lindley Memorial

Lecturship Fund.

DAVIDSON, D. 1984. *Inquires into Truth and Interpretation*, Clarendon Press, Oxford.

DAVIDSON, D. 1985. "A new basis for decision theory", *Theory and Decision* 18. pp. 87-98.

DAVIDSON, D. 1990. "The Structure and Content of Truth". *The Journal of Philosophy*, Vol. LXXXVII, No 6, June 1990, pp.279-328.

DAVIDSON, D. 1991. "Three Varieties of Knowledge". In Griffiths A.P. (ed.) 1991, pp.153-166.

DE FINETTI, B. 1937. *Foresight: Its Logical Laws, Its Subjective Sources*, in Kyburg and Smokler (eds.) 1964. pp. 93-158.

DE MORGAN, A. 1847. *Formal Logic*. London: Taylor and Walton.

DENNETT, D. 1981. "True Believers" in Dennett 1987.

DENNETT, D. 1987. *The Intentional Stance*. Cambridge Massachusetts, London, England. Bradford book, MIT Press.

DENNETT, D. 1975. "Brain writing and mind reading", *Minnesota studies in the Philosophy of Science*, 7.

DUMMETT, M. 1976. "What is a Theory of Meaning? (II)", in *Truth and Meaning, Essays in Semantics*, Evans and Mc Dowell (eds.) 1976.

DUMMETT, M. 1981 (2nd Edition). *FREGE Philosophy of Language*, Duckworth.

EELLS, E. 1982. *Rational Decision and Causality*, Cambridge University Press.

EVANS, G. and McDOWELL, J. (eds.). 1976. *Truth and Meaning, Essays in Semantics*, Oxford: Clarendon Press.

EVANS, G. 1981. "Semantic Theory and Tacit Knowledge" in *Wittgenstein: to Follow a Rule*. Holtzman S. and Leich C. (eds). 1981. pp. 118-137.

FIELD, H. 1977. "Logic, Meaning and Conceptual Role". *Journal of Philosophy* 74, pp. 375-406.

- FIELD, H. 1978. "Mental Representation", *Erkenntnis* 13, pp. 9-61.
- FISHBURN, P. 1970. *Utility Theory for Decision Making*. John Wiley & Sons, Inc. New York
- FISHER, R. 1959. *Smoking: The cancer controversy*. Oliver and Boyd. Edinburgh and London.
- FODOR, J. 1985. "Fodor's Guide to Mental Representation" in Stich, S. (eds.) 1994.
- FODOR, J. 1987. *Psychosemantics*, The MIT Press.
- FREGE, G. 1892. "On Sense and Reference", in Geach & Black (eds.) 1970, pp. 56-78.
- FREGE, G. 1918. "Thoughts", in *Logical Investigations*, Oxford Basil Blackwell 1977, pp. 1-30.
- GAERDENFORS, P. and SAHLIN, N-E. (eds.). 1988. *Decision Probability and Utility*, Cambridge University Press.
- GARNER, R. 1972. "On Saying What is True", *Nous* 6, pp. 201-224.
- GEACH, P. 1965. "Assertion", *Philosophical Review* 74, pp. 449-465.
- GEACH, P. & BLACK, M. (eds.). 1970. *Frege Translations*, 2nd ed. Basil Blackwell, Oxford.
- GIBBARD, A. & HARPER, W. 1978. "Counterfactuals and two kinds of Epistemic Utility", in Gaerdenfors, P. and Sahlin, N-E, (eds.). pp. 341-77.
- GORDON, R. 1969. "Emotions and Knowledge", *The Journal of Philosophy* 66, pp. 408-413.
- GRANDY, R. 1973. "Reference, Meaning, and Belief", *The Journal of Philosophy* 70, pp. 439-452.
- GRIFFITHS, A. P. (ed.) 1991. *A.J. Ayer Memorial Essays*. The Royal Institute of Philosophy 1991. Cambridge: Cambridge University Press.
- HAACK, R., 1971. "On Davidson's Paratactic Theory in Oblique Contexts". *Nous* 5, N71, 351-361.
- HARMAN, G. 1973. *Thought*. Princeton. New Jersey: Princeton University Press.

- HOLTZMAN, S and LEIGH, C. (eds.). 1981. *Wittgenstein: To Follow a Rule* Routledge & Kegan Paul. London.
- HOOK, S. (ed.). 1960. *Dimensions of Mind*, New York.
- HOPKINS, J. and WOLLHEIM, R. (eds.). 1982. *Philosophical Essays on Freud*. Cambridge: Cambridge University Press.
- HUMBERSTONE, L. 1992. "Direction of Fit". *Mind* 101.pp. 59-83.
- HURLEY, S. 1989. *Natural Reasons*. Oxford: oxford University Press.
- JACOBS, R. A. and ROSENBAUM, P.S. (eds.). 1970. *Readings in Transformational Grammar*. Ginn and Company.
- JEFFREY, R. C. 1965. *The logic of decision*. McGraw-Hill Book Company.
- JEFFREY, R. 1976. "Savage's Omelet". In *PSA 1976*, Vol. 2. pp. 361-371.
- JEFFREY, R. 1983. *The Logic of Decision*. 2nd ed. Chicago IL. University of Chicago Press.
- KOOPMAN, B. O. 1940. "The axioms and algebra of intuitive probability", *Annals of Mathematics*, 41, pp. 269-292.
- KRANTZ, D., LUCE, D., SUPPES, P., TVERSKY, A., 1971. *Foundations of Measurement*. Vol.1. Academic Press. New York
- KYBURG, H. and SMOKLER, H. (eds.). 1964. *Studies in subjective probability*. New York, London, Sydney: John Wiley & Sons.
- LAKOFF, G. 1972. "Linguistics and Natural Logic", in Davidson, D., and Harman, G. (eds.) 1972.
- LEPORE, E. and VAN GULICK, R. 1991. *John Searle and his critics*. Oxford: Blackwell.
- LEWIS, D. 1966. "An Argument for the Identity Theory" in Lewis 1983.
- LEWIS, D. 1970. "How to Define Theoretical Terms", *Journal of Philosophy* 67, no.13, pp. 427-444.
- LEWIS, D. 1970. "General Semantics" in *Philosophical Papers* Vol.1, New York Oxford, Oxford University Press, 1983, pp.189-229.
- LEWIS, D. 1972. "Psychophysical and Theoretical Identifications", *Australian*

Journal of Philosophy, 50. pp. 249-258.

LEWIS, D. 1974. "Radical Interpretation" in D. Lewis 1983, pp. 108-121.

LEWIS, D. 1978. "Mad Pain and Martian Pain" in Lewis 1983, pp.122-132.

LEWIS, D. 1980. "A Subjectivist's Guide to Objective Chance" in Lewis 1986 pp. 83-132.

LEWIS, D. 1981. "Causal Decision Theory". *Australasian Journal of Philosophy*, 59. pp. 1-30.

LEWIS, D. 1983. *Philosophical Papers* Vol.1, New York, Oxford: Oxford University Press.

LEWIS, D. 1986. *Philosophical Papers* Vol. 2, New York, Oxford: Oxford University Press.

LEVI, I. 1966. "Review of Jeffrey's Logic and Decision" *Synthese* 16, pp. 239-244.

LEVI, I. 1991. "Consequentialism and Sequential Choice", in Baccharach M., and Hurley S. (eds.). 1991.

LOAR, B., 1981. *Mind and Meaning*. Cambridge: Cambridge University Press.

LUCE, R.D., 1968. "On the numerical representation of qualitative conditional probability", *Annals of Mathematical Statistics*, 39, pp. 481-491.

McDOWELL, J. 1977. "On the Sense and Reference of a Proper Name", *Mind*, 86. pp. 159-185

MELLOR, D. H. 1971. *The Matter of Chance*. Cambridge: Cambridge University Press.

MELLOR, D. H. (ed.), 1990. *F. P. Ramsey, Philosophical Papers*. Cambridge: Cambridge University Press.

MILLER, S. 1991. "Davidson's paratactic analysis of mood", *Journal of Pragmatics* 15, pp. 1-10.

NOZICK, R. 1969. "Newcomb's problem and two principles of choice", in Recher N. (ed) 1969.

NOZICK, R. 1993. *The Nature of Rationality*. Princeton: Princeton University Press.

PEACOCKE, C. 1986. *Thoughts: An Essay on Content*. Aristotelian Society Series. Volume 4. Basil Blackwell. Oxford.

PLATTS, M. 1980. *Reference, Truth and Reality*. Routledge & Kegan Paul. London, Boston and Henley.

POPPER, K. 1990. *A World of Propensities*. Bristol: Thoemes.

PRICE, H. 1983. " 'Could a Question Be True?': Assent and the Basis of Meaning", *The Philosophical Quarterly*, vol. 33. pp. 354-364.

PUTNAM, H., 1960. "Minds and Machines", in *Dimensions of Mind*, Sidney Hook, (ed.). 1960, pp. 148-179.

PUTNAM, H., 1975. *Mind, Language and Reality*. Cambridge. Cambridge University Press.

QUINE, W.V.O., 1960. *Word and Object*, The M.I.T. Press, Cambridge, Massachussets.

RAMBERGH, B., 1989. *Donald Davidson's Philosophy of Language*. Basil Blackwell.

RAMSEY, F. 1926. "Truth and Probability" in D.H. Mellor ed. 1990, pp. 52-109.

REICHENBACH, H. 1949. *The Theory of Probability: an Inquiry into the Logical and Mathematical Foundation of the Calculus of Probability*. 2nd edition. Berkeley: University of California Press.

RESCHER, N. (ed.). 1969. *Essays in Honour of Carl Hempel*. Dordrecht: Reidel.

ROSS, J. R. 1970. "On Declarative Sentences" in *Readings in English Transformational Grammar*, Jacobs, R. A. and Rosenbaum, P. S. (eds.) 1970.

SAVAGE, L. J. 1954. *The Foundations of Statistics*. John Wiley & Sons, inc., New York. Chapman & Hall, Limited, London.

SCHICK, F. 1986. "Dutch Bookies and Money Pumps". *The Journal of Philosophy*. Vol. 83. pp. 112-119.

SCHIFFER, S. 1987. *Remnants of Meaning*. A Bradford Book, The MIT Press, Cambridge Massachusetts, London, England.

SCOTT, D. and SUPPES, P. 1958. "Foundational Aspects of Theories of Measurement", *The Journal of Symbolic Logic*, Vol. 23, pp. 113-128.

SEARLE, J. 1969. *Speech Acts*. Cambridge: Cambridge University Press.

SEARLE, J. 1977. "Locutions and Illocutions". *Phil. Review*, 1977, p.405-424.

SEARLE, J. 1983. *Intentionality*. Cambridge University Press.

SEARLE, J, and VANDERVEKEN, D. 1985. *Foundations of Illocutionary Logic*. Cambridge University Press, Cambridge.

SKYRMS, B. 1982. "Causal Decision Theory". *Journal of Philosophy* 79. pp. 695-711.

SKYRMS, B. 1990. "*Ratifiability and the Logic of Decision*". Midwest Studies in Philosophy. Vol. XV. The University of Notre Dame Press.

SPOHN, W. 1977. "Where Luce and Kranz do really generalize Savage's decision model". *Erkenntnis* 11, pp. 113-34.

STALNAKER, R. 1984. *Inquiry*. Cambridge, Mass. London: MIT Press, A Bradford Book.

STENIUS, E. 1967. "Moods and Language-Game". *Synthese* 17, pp. 254-274.

STICH, S. and WARFIELD, T. (eds.). 1994. *Mental Representation, a reader*. Oxford, Blackwell.

STOCKER, M. 1996. *Valuing Emotions*. Cambridge. Cambridge University Press.

TARSKI, A. 1956. "Concept of Truth in Formalised Languages", in *Logic Semantics and Metamathematics*. Hackett Publishing Company, 1983.

VAN STRAATEN, Z. 1980. *Philosophical Subjects: Essays Presented to P.Strawson*. Oxford University Press.

VELLEMAN, D. 1998. "How Belief Aims at the Truth". Unpublished Paper.

VON MISES, R. 1946. *Notes on Mathematical Theory of Probability and Statistics*. Cambridge: Massachusets.

VON NEUMANN, J & MORGENSTERN, O. 1947. *Theory of games and economic behaviour*, Princeton, Princeton University Press.

WIGGINS, D. 1980. "What would be a Substantial Theory of Truth?", in *Philosophical Subjects: Essays Presented to P. Strawson Z.*, Van Straaten, (ed.), pp.189-221.

which the acts have identical consequences. The condition can be seen as a rationality condition on the presumption that the state that actually obtains does not depend on the act chosen. From axiom 2 it follows that the relation of preference between acts given an event is a simple ordering.

Axiom 3 introduces a relation, that is a kind of correspondence, between preference among acts and preference among consequences.

Axiom 4 and axiom 5 introduce a relation of being *not more probable than* between events that, using axioms 1, 2, 3 it can be shown to be a qualitative probability. Axiom 5 says that an agent is not indifferent to every pair of consequences. Axiom 4 in Savage's words explicates the following condition:

“ The assumption that on which of two events the person will chose to stake a given prize does not depend on the prize itself” (1954, p. 31)

According to axiom 4 the relation of being *not more probable than* between events defined in terms of preference between acts does not depend on a particular act. The condition is intuitively plausible as a rationality condition again on the presumption that the state that actually obtains does not depend on the act chosen.

Axiom 6 –not straightforwardly introducing a rationality constraint on preferences- enables the numerical representation of the agent's degrees of belief by a countably additive probability measure. As it has already been mentioned it can be shown that the relation of *being not more probable than* holding between events is a qualitative probability. By the introduction of the rather technical axiom 6 it can be shown that there exists a unique probability measure P that strictly agrees with the qualitative probability relation *not more probable than*. This notion is accounted by the following condition:

For a probability measure P on S and a qualitative probability relation \leq_p on S we say that P strictly agrees with \leq_p if and only if

for every B, C , $P(B) \leq P(C)$ if and only if $B \leq_p C$.

P almost agrees with \leq_p if $B \leq_p C$ implies $P(B) \leq P(C)$

Axiom 6 implies the simpler axiom 6' which, together with axioms 1-5, is again sufficient for introducing a unique probability measure on events and a unique

up to a linear transformation utility measure on consequences for the cases of finite acts, that is, acts with a finite number of consequences.

For the set of events E and a qualitative probability \leq_p , by axiom 6' it follows that $\langle E, \leq_p \rangle$ is infinite. This constitutes condition 1.

Condition 1: For any pair $\langle E, \leq_p \rangle$, satisfying axiom 6', E is infinite.

Proof:

- a) It follows from 2), 3) of Definition 4, that if $A \subseteq B$ then $A \leq_p B$.
- b) If $A \leq_p B$, $C \leq_p D$ and $A \cap C = B \cap D = \emptyset$, then $A \cup C \leq_p B \cup D$.

Proof: By 3) (definition 4) $A \cup (C \setminus B) \leq_p B \cup (C \setminus B) = C \cup (B \setminus C) \leq_p D \cup (B \setminus C)$, hence $A \cup (C \setminus B) \leq_p D \cup (B \setminus C)$.

By 3) (definition 4) $A \cup C = (A \cup (C \setminus B)) \cup (B \cap C) \leq_p (D \cup (B \setminus C)) \cup (B \cap C) = B \cup D$

- c) By 2) (definition 4) $\exists E_0 \in E$ such that $\emptyset \prec_p E_0$. By Axiom 6', there is a partition $\{C_i, 1 \leq i \leq n\}$ such that for each i , $1 \leq i \leq n$, $C_i \prec_p E_0$.

For some i , $1 \leq i \leq n$, $\emptyset \prec_p C_i$.

Proof: Suppose that for each i , $1 \leq i \leq n$, $C_i \approx_p \emptyset$, where $C_i \approx_p \emptyset$ iff $\emptyset \leq_p C_i$ and $C_i \leq_p \emptyset$.

By repeated application of b) we have that $C_1 \cup C_2 \cup \dots \cup C_j \leq_p \emptyset$, $2 \leq j \leq n$.

But $S = C_1 \cup C_2 \cup \dots \cup C_n$, hence $S \leq_p \emptyset$.

This contradicts $\emptyset \prec_p S$.

Let E_1 be one of the C_i for which $\emptyset \prec_p C_i$ such that $\emptyset \prec_p E_1 \prec_p E_0$.

We repeat the argument forming a sequence $\{E_n\}_{n \in \mathbb{N}}$ such that

$\emptyset \prec_p \dots \prec_p E_{n+1} \prec_p E_n \prec_p \dots \prec_p E_1 \prec_p E_0$.

So $\langle E, \leq_p \rangle$ is infinite. #

In addition the following condition can be proved.

Condition 2. For every event B and for every real number ρ such that $0 \leq \rho \leq 1$,

there exists an event C such that $P(C) = \rho P(B)$, where P is the unique probability measure that almost agrees with \leq_p .

A subjective conditional probability function can also be introduced for any condition D , $D \neq \emptyset$. This is done by considering a qualitative probability relation between events given D . The relation $E|D \leq F|D$, for E, F events is introduced by preference between acts f, g given D .

Axiom 7 is necessary for measurement of utility of a generalised notion of act. It is of no interest for the subject we are mainly concerned here and I will not comment on it.

The suggestion is that axioms 1-5 of Savage's theory are axioms of rationality. The axioms reflect rational relations between preferences, where a preference is defined as a relation between acts and acts given events. I will leave aside here considerations and problems related to the kind of interpretation given to a decision theory; that is that the theory provides normative constraints on preference and that it is empirically falsified. The issue with which I am primarily concerned is whether, and if so, how, it is possible from rationality axioms that govern preferences between acts and acts given events to draw conclusions about any distinctive kind of relations that hold between events or states. The relations would be of a kind distinctive for events or states if they suffice to characterise distinctively the category of events/states. In the cases of rational decision the relations in question would hold between what the event terms stand for and will not hold between what the consequence terms stand for. If it is shown that from the rationality axioms of Savage's theory it follows that there is a certain distinctive kind of relations that hold between events, Savage's choice of considering two categories of terms will be vindicated.

Let us then complete our presentation of Savage's theory with a brief discussion of the way the concept of utility is introduced. A probability function introduced by Axioms 1-6 provides a way of measuring events. A utility function is introduced for a numerical representation of the relation of preference among acts. Savage follows Von Neumann and Morgenstern (1947) in the introduction of the notion of utility. In the special cases of acts called gambles⁶⁴, a utility is a function U associating real

⁶⁴ A gamble $\sum p_i f_i$ is an act f for which there is a partition $\{B_i: i\}$ of S such that $P(B_i) = p_i$ and $f(s_i) = f_i$ for s in

numbers with consequences in such a way that

if $f = \sum \rho_i f_i$ and $g = \sum \sigma_j g_j$, then $f \leq g$ if and only if $\sum \rho_i U(f_i) \leq \sum \sigma_j U(g_j)$.

It is proved that there is a utility function and that it is unique up to a linear transformation. That is, for any two utility functions the one is an increasing linear function of the other.

The proof of the above does not require any further postulates than axioms 1-6. As mentioned already, Savage considers a further postulate, axiom 7, which enables him to account for utility in the cases of unbounded acts. I leave the whole discussion aside on the ground that an individual in a decision situation normally considers only a finite number of consequences.

For any act (gamble) f ($f = \sum \rho_i f_i$, $i \in I$, I finite) the quantity $V(f) = \sum \rho_i U(f_i)$ is the subjective expected utility of the act f . If f_i 's are the finite number of consequences of act f , then $\{B_i, i \in I\}$ is a partition of S such that for every s in B_i $f(s) = f_i$ and ρ_i 's are the probabilities of the events B_i . The quantities $U(f_i)$ are the utilities of the consequences f_i . As in Ramsey's theory, so in Savage's theory there is the premise that the psychological law that governs the choices of individuals is the maximisation of subjective expected utility (MSEU); agents choose acts that have the highest subjective expected utility V .

It is important to mention here that in Savage's theory there is a distinction between utility and expected utility.⁶⁵ Utility U is assigned to consequences while expected utility V is assigned to acts. Utilities are independent of acts (the value of a consequence is independent of the act that brings it about). Expected utilities of acts are determined jointly from the utilities of consequences of the act and the subjective probability of events. Expected utilities represent actions which embody the independent from preferences notion of probability and utility. An interpretation of the notions here is that probabilities represent beliefs, utilities represent values and expected utilities represent preferences. Given the notion of act as it is introduced here, Savage's theory is interpreted as *strict consequentialist*⁶⁶. It is a feature of

Bi. The f_i 's are a finite sequence of consequences and p_i a corresponding sequence of non-negative real numbers such that $\sum p_i = 1$. Equivalent to the class of gambles is the class of acts which with probability one are confined to a finite pair of consequences

⁶⁵ In Jeffrey's theory, that will be presented in the following section, utility and expected utility are fused into a notion of desirability.

⁶⁶ The term should not be confused with its use in ethics. See Levi (1991) for details.

Savage's theory that the value of a consequence does not depend on the act that brings it about. Rather the value of an act depends on the values of its consequences.

With these considerations in mind, what remains to be done is to check whether Savage's theory satisfies the requirement discussed previously. That is, the requirement for a distinctive characterisation of events. The crucial step in Savage's theory for the introduction of a specific relation that can hold between events is Axiom 4. Because of Axiom 4 the definition of *not more probable than* between two events does not depend on a particular act, that is on a particular correlation between events and consequences. It only requires Axiom 5. That is that there are at least two consequences between which a person has preference.

So it seems that part of the requirement stated above is satisfied. When certain restrictions of rationality are met on the preference relation between any acts/consequences, then we can tell for any events A, B either that A is not more probable than B or that B is not more probable than A. But there is an additional question that has to be answered for the requirements discussed previously to be satisfied. This is whether this kind of relation that might hold between events in cases of rational decision is distinctive of events. A similar question for a characterisation of consequences would be whether there is a characterisation the theory yields for consequences and whether the holding of this relation is distinctive of consequences. In Savage's theory consequences are determined by acts and the relation of preference between consequences is defined from the relation of preference between acts. To answer both of those questions we have to explore the relation *is not more probable than* as holding between events and the relation *not preferred to* as holding between consequences.

Let us then discuss how Savage's terminological distinctions have been vindicated. That is, whether the relations holding between events (*not more probable than*) and the relations holding between consequences (*not preferred to*) introduce different kind of characterisations of each category, which, in addition are distinctive of each category. The answer to the question of a different characterisation of each category will be given by the representation theorems of the theory. Measurement of one domain by a probability function characterises the category of events and distinguishes it from the category of consequences, represented and characterised by a utility function. The representation theorems of the theory suggest the following.

In the domain of events equipped with the qualitative relation *not more probable than* the representation theorem establishes a unique quantitative probability distribution that preserves the qualitative relation. That is, for any two events, A, B , A is *not more probable than* B if and only if $P(A) \leq P(B)$.

In the domain of consequences equipped with the relation *of not being preferred to* as defined by the theory, the representation theorem establishes a utility function from consequences to real numbers equipped with the usual ordering, that preserves the structure (preference ordering) of the domain of consequences.

So, the domain of events and the domain of consequences admit of characterisations. The two domains have different formal representations. The terminological distinction Savage has employed is vindicated by the theory. The relation *not more probable than*, being represented by a probability function induces the structure of the domain of events. The domain of events is a probability space, an algebra closed under union and complementation with a unique probability measure on it. The structure of the domain of consequences is induced by the relation *not preferred to*, being represented by a utility function in a finite subset of real numbers with the usual ordering. The domain of consequences is an ordered space. In the case we consider a finite number of consequences, the domain of consequences is a finite ordered space.

Let us now consider the question of whether these characterisations are distinctive of each category. That is, let us consider the question of whether there could be an isomorphism between the domain of events with the probability structure and the domain of consequences with a preference ordering. An isomorphism F between $\langle E, P \rangle$ the domain of events E with the probability structure P , and $\langle C, \leq \rangle$ the domain of consequences C with a utility ordering would be an one-to-one correspondence between E and C that would also preserve the structure.

A problem with the possibility of an isomorphism between $\langle E, P \rangle$ and $\langle C, \leq \rangle$ is that the two domains have different cardinality. By condition 1, $\langle E, \leq_p \rangle$ is infinite and so is $\langle E, P \rangle$. Since the domain $\langle C, \leq \rangle$ is finite, by condition 2, $\langle C, \leq \rangle$ cannot be structurally isomorphic to $\langle E, P \rangle$. The reason is that for every $E_i \in E$, and for every real number p , $0 \leq p \leq 1$, there is an $E_j \in E$ such that

$P(E_i) = p P(E_j)$. This is not the case for the domain of consequences $\langle C, \leq_p \rangle$ ⁶⁷.

A point that could be raised here is the following. The set of consequences is finite and it does not satisfy condition 2. But consequences in Savage's theory can be seen as acts. The utility of consequence C , $U(C)$ is the expected utility of the constant act that has the value C in all states. The issue here is that if the domain of events is infinite, the domain of preferences between acts is infinite too. (For every event B , and acts f, g we can define $f \leq g$ given B). The question then is whether the domain of events and the domain of consequences could be isomorphic, if consequences can be seen as acts. Our answer to this question involves two points. The first is that if the set of consequences is finite, then preference between constant acts is finite too. However, we could extend the preference ordering so that it includes conditional acts. A preference ordering on conditional acts is infinite, if the set of events is infinite. The second point in our answer is that in this case too, there are structural differences between the domain of events and the domain of conditional acts.

In order to formulate our answer we need to introduce the notion of conditional act and to extend the preference relation so that it includes conditional acts.

Definition 1': Given an act f and a non-empty event A , the conditional act $f|A$ (f given A) is the (partial) function that agrees with f on A and is undefined outside A .

Definition 1'': $f|A \cup g|B$ is the partial function that agrees with f on A , agrees with g on B and is undefined outside $A \cup B$. It is only defined if either $A \cap B = \emptyset$ or f and g agree on $A \cap B$.

Axiom 2': for f, g acts and A, B non-empty events,

a) $f|A \leq g|A$ if and only if $f \leq g$ given A

b) for A, B such that $A \cap B = \emptyset$,

$f|A \leq g|B$ if and only if $f|A \leq f|A \cup g|B$ if and only if $f|A \cup g|B \leq g|B$

c) where f is an act of constant value, (i.e. a consequence) for any non-empty events A and B , $f|A \leq f|B$.

The extension of the preference ordering introduced by Axiom 2' enables comparison between conditional acts. For instance, it enables comparisons between

⁶⁷ Notice that there is no isomorphism between the two domains even if we take the set of consequences to be denumerably infinite.

acts like staying at home in case it rains and staying at home in case it does not rain. The subject prefers staying at home in case it rains to staying at home in case it does not rain if and only if the subject prefers staying at home in case it rains to staying at home and, he prefers staying at home to staying at home in case it does not rain.

With these conditions in the background, let us discuss our first point in our reply to the possibility of an isomorphism between events and consequences. Our point is that comparisons between consequences and constant acts are finite. Let C_1 and C_2 be consequences, then $C_1 \leq C_2$ if and only if $f_{C_1} \leq g_{C_2}$ where f_{C_1} and g_{C_2} are the constant acts that take as values C_1 and C_2 respectively in all states. Now, for every event A we can consider the preference $f_{C_1}|A \leq g_{C_2}|A$. Since the domain of events is infinite the domain of preference on conditional acts is infinite too. But notice that this does not imply that the domain of preferences on constant acts is infinite. The reason is that the expected utility of a constant act f_C and of a conditional constant act $f_{C|A}$ coincide with the utility of its consequence C . So, the possibility of extending the preference relation so that it applies in a infinite domain is not a threat to the original point that there cannot be an isomorphism between $\langle C, \leq \rangle$ and $\langle E, P \rangle$.

Let us now consider the domain of conditional acts with the extended preference relation, $\langle F_C, \leq \rangle$. Axiom 2' b) entails that for A, B events such that $A \cap B = \emptyset$, $f|A < g|B$ if and only if $f|A < f|A \cup g|B$ if and only if $f|A \cup g|B < g|B$. Comparing this property of the domain $\langle F_C, \leq \rangle$ with the property of $\langle E, \leq_p \rangle$ introduced by axiom 6', we notice the following structural difference. For the case of $\langle F_C, \leq \rangle$ the following holds. For every $A, B \in E$ such that $A \cap B = \emptyset$, and for every $f|A, g|B \in F_C$, if $f|A < g|B$, then there is an act $h \in F_C$ expressible just in terms of $f|A$ and $g|B$ such that $f|A < h < g|B$. The equivalent of this condition need not hold in $\langle E, \leq_p \rangle$. By axiom 6' it holds that for every A, B events, if $A < B$ then there is a C event such that $A < C < B$. But any such C is not expressible just in terms of A and B . This marks a structural difference between the domain of events, $\langle E, \leq_p \rangle$ and the domain of conditional acts $\langle F_C, \leq \rangle$ ⁶⁸.

The result of the above considerations is that the axiomatic system of Savage's decision theory introduces distinctive characterisations for the category of events and

the category of consequences and conditional acts. The category of events and the category of consequences are defined as two distinct not isomorphic relational systems.

It is important to notice here that with definitions 1', 1'' and axiom 2' we can introduce a notion of qualitative independence between events on the basis of preference between conditional acts. While it seems that in Savage's theory a notion of qualitative independence between events cannot be introduced from a relation of conditional preference, the extension of preference on conditional acts we have considered here enables that.

For this we consider the following:

Definition: $A|B \leq_p C|D$ if and only if $f_A|B \leq f_C|D$ for all f_A, f_C acts, c_1, c_2 consequences such that $c_2 \leq c_1$ and

$$f_A(s) = c_1 \text{ for every } s \text{ in } A$$

$$f_A(s) = c_2 \text{ for every } s \text{ in } -A$$

$$f_C(s) = c_1 \text{ for every } s \text{ in } C$$

$$f_C(s) = c_2 \text{ for every } s \text{ in } -C$$

At this stage we could give up the terminology of events and consequences and suggest that events, consequences and acts can be represented by sentences. A decision situation is ascribed to an individual by a system of sentences that has the form of a decision matrix like table 1. In a decision matrix like table 1, the events and consequences are represented by sentences. Acts are represented by functions between (event) sentences and consequence (sentences).

On the assumption that the agent makes 'rational' choices, that is, choices that conform to the axioms of the theory, the theory induces a unique subjective probability function and a utility function unique up to a linear transformation. Taking sentences to be related to the attitudes of the agent, that is, that sentences express or ascribe attitudes, it becomes possible to distinguish between the sentences relevant to a certain decision situation, the ones that are belief related and the ones that are preference related. The distinction between belief related sentences (sentences related to the doxastic attitudes of the agent) and the preference related sentences (sentences

⁶⁸ Dr Peter Milne's contribution to formulating this point has been substantial.

related to the affective attitudes of the agent) is introduced by the different representations between the sentences representing events and the sentences representing consequences.

More explicitly the situation is as follows. The domain of belief related sentences is classified as the domain of sentences with the structure induced by the relation *being more probable than* represented by the probability function for the agent. The belief related sentences or simply the beliefs of an agent constitute a relational system with a probability measure. A sentence that belongs to a system of sentences which are related to each other by *being more probable than* is a belief related sentence. The way sentences relate to beliefs is by a way of expressing or by a way of ascribing a belief. The principle of triangle enables that a sentence that belongs to a certain probabilistic/doxastic/evidential system of sentences relates to the same belief of the agent either in the case it expresses it or in the case it ascribes it. The domain of preference related sentences (preference ascribing or preference expressing sentences) is classified as the domain of sentences with a structure induced by a utility function. The preference related sentences, the sentences that are related to the affective attitudes of the agent or simply the desires of the agent constitute a relational system of sentences with a preference ordering.

I make some brief comments here on the suggested classification of sentences as doxastic or affective. The issues will be further developed in the following chapter where I consider the co-operation of decision and interpretation theories. One feature of the suggested classification is that it is holistic. A sentence is classified as doxastic if it is integrated into a relational system of sentences with a probability measure. We could call a relational system with a probability measure a doxastic relational system. A sentence is classified as affective if it is integrated into a relational system of sentences with a preference ordering. We could call a relational system with a preference ordering an affective relational system. So, a classification of a sentence as doxastic or affective depends on the relational system of which the sentence is a member. The classification is holistic in the respect that a sentence is classified as doxastic or affective on the basis of its relations to other sentences of the system of sentences of which it is a member. Only sentences that are members of the one or the other relational system are identified here as doxastic or affective sentences.

In the following chapter I discuss more extensively the bearing these

considerations have on the possibility of classifying the doxastic and affective states of an agent in a theory of radical interpretation. There I will discuss into more detail the possibility of accommodating Savage's theory of decision along with its introducing classifications into a theory of radical interpretation. Here I continue with discussing an alternative decision theory. This theory, proposed by Richard Jeffrey, is the most popular decision theory among philosophers.

6.4 Jeffrey's decision theory

In this section I give a short presentation of and commentary on Jeffrey's decision theory. The reason for presenting Jeffrey's theory is that it is the decision theory most used by philosophers. Davidson in considering the possibility of an integrated theory of decision and interpretation suggests a version of Jeffrey's theory for an integration of decision theory with interpretation theory⁶⁹. A point we should bear in mind in discussing Jeffrey's decision theory here is that it is a theory in which the assumption that beliefs and desires can share the same content is taken. Subjective probabilities and utilities are attributed to the same kind of entities, to propositions. As with the discussion of an integration of Savage's theory with radical interpretation, the suggestion of an integration of Jeffrey's theory with interpretation will be given in the following chapter where I consider the possibility of an integration of decision and interpretation theory.

Jeffrey (1965) compares his theory to Ramsey's. His proposal is that instead of taking the objects of preferences to be gambles, essential for Ramsey's probability measurement of 'propositions' and utility of outcomes, we consider a unified theory of decision. The theory is unified in the sense that all attitudes are attributed to one kind of entity, to propositions. His use of the notion of proposition is not extensively specified; more or less he takes a proposition to be what a sentence means. Probabilities and desirabilities are attributed to propositions and acts are represented by propositions.

"An act is then a proposition which is within the agent's power to make true if he pleases, and the necessary proposition would correspond to no acting: to letting what will be, be" (1965, p. 73).

An agent believes or partially believes a proposition, he desires a proposition

⁶⁹ Davidson compares a Jeffrey-style radical decision theory to a Ramsey-style theory. He does not consider the

and he prefers one proposition to others. Partial belief and strength of desire, measured by subjective probability and utility are attitudes to propositions. To believe that you will play tennis tomorrow is to have one sort of attitude towards the proposition that you will play tennis tomorrow. To desire to play tennis tomorrow is to have another sort of attitude towards the same proposition, the proposition that you will play tennis tomorrow. On the other hand, desirabilities of propositions are reflected in preference rankings.

“ In general, to say that $\text{des}A > \text{des}B$ is to say that A is ranked higher than B in the agent's preference ranking of proposition. To say that A is ranked higher than B means that the agent would welcome the news that A is true more than he would the news that B is true. A would be better news than B.” (1965, p. 72)

Taking T, that is the necessary proposition to be no news, bad news and good news are defined as to whether they stand below or above it respectively in the preference ranking.

Jeffrey's theory, which in its mathematical details is heavily indebted to the work of Bolker (1965), as cited in Jeffrey (1965) and Bolker (1967) is presented in the following.

Jeffrey's theory of decision:

Let D be a set of propositions and \geq a relation of preference on D .

Axiom 1. The relation \geq on D is transitive and connected.

Axiom 2. The field of the relation \leq is a complete, atomless Boolean algebra of propositions from which the impossible proposition F has been removed.

Axiom 3. If $A \cap B = F$ then

- a) if $A > B$, then $A > A \vee B$ and $A \vee B > B$, and
- b) if $A = B$, then $A = A \vee B$ and $A \vee B = B$.

Axiom 4. Given $A \cap B = F$ and $A = B$, if $A \vee C = B \vee C$ for some C where $A \cap C = B \cap C = F$ and not $C = A$, then $A \vee C = B \vee C$ for every such C .

Some comments on Axiom 2 are due here. A set of propositions that is closed under negation, conjunction, disjunction and implication is a *Boolean algebra*. A Boolean algebra of propositions is atom-free, if, for every proposition P there is

another proposition Q in the algebra that is not equivalent to P , but which implies P . A Boolean algebra is complete if it contains all finite and infinite conjunctions and disjunctions of its members. Among the axioms of Jeffrey's theory of decision Axiom 2 has a special status. Axiom 2 imposes a structural condition on the domain on which the preference relation is introduced. The condition that preference is a relation on a complete, atom-free Boolean algebra constitutes the structural condition of the theory; it states that the set of preferences satisfies certain formal constraints that independent of the constraints imposed on the preference relation by axioms 1, 3, 4.

In Jeffrey's axiomatised theory the following *existence theorems* are proved.

Existence theorems (Representation Theorems):

There are functions p, d (from propositions to real numbers), such that:

- a) p is non-negative,
 - b) p is normalised, that is $p(T) = 1$, T the necessary event,
 - c) p is additive, i.e. if $XY = F$ then $p(X \vee Y) = p(X) + p(Y)$,
- and
- d) If $p(XY) = 0$ and $p(X \vee Y) \neq 0$, then
- $$d(X \vee Y) = (p(X)d(X) + p(Y)d(Y)) / (p(X) + p(Y)).$$

The Uniqueness Conditions

In addition, any functions P, D (from propositions to real numbers) satisfy a)-d) if and only if

there are certain real numbers a, b, c, d such that

- i) $ad - bc$ is positive
- ii) $desA + d$ is positive
- iii) $c des T + d = 1$

and

$$P(X) = p(X)(cd(X) + d)$$

$$D(X) = (ad(X) + b) / (cd(X) + d)$$

Condition d) is called the *desirability axiom*. It is equivalent to the following

condition: $[\sum_{A \in P} p(A) d(A)] / p(X) = d(X)$, where P is a partition of X . Here is how the desirability of any proposition can be calculated by use of the desirability axiom.

If we consider a proposition Q and the propositions $S_1 \& Q, S_2 \& Q, \dots, S_n \& Q$ mutually exclusive and jointly equivalent to Q the following follows:

1. $d(Q) = (\sum_{i=1, \dots, n} p(S_i \& Q) d(S_i \& Q)) / p(Q)$
2. $d(Q) = \sum_{i=1, \dots, n} p(S_i | Q) d(S_i \& Q)$

An interesting case to consider is the case in which the proposition Q above stands for an act. If we can think of $(S_i)_{i=1, 2, \dots, n}$ as a events, then the equation 2 above would express the equivalent of Savage's subjective expected utility

$(V(Q) = \sum_{i=1, 2, \dots, n} P(S_i) U(Q_i))$. A significant difference between the two is that in Savage's theory the event-act independence condition hold, which suggests that in the Savage's equivalent of 2 it holds that $p(S_i | Q) = p(S_i)$. So, while in Savage's theory the psychological principle that guides choices is the maximisation of expected subjective utility (MESU), in Jeffrey's theory the equivalent principle is the maximisation of conditional expected utility (MCEU). The weights that average the desirability are probabilities conditional on the act⁷⁰. Another difference between Savage's principle of maximising subjective expected utility and Jeffrey's principle of maximising conditional expected utility is the following. While in Savage's theory the expected utilities are computed on values of consequences –I remind the reader that in Savage's theory the value of consequence is independent of the act-, in Jeffrey's principle the difference between expected value and value vanishes.

Another point worth mentioning concerns the uniqueness conditions of the representations. As we can see from the uniqueness condition of Jeffrey's theory the condition does not imply a unique probability representation of propositions. The uniqueness condition in Jeffrey's theory determines an infinite family of pairs of functions. The significance of this feature of Jeffrey's theory for a possibility of integrating it with interpretation theory will be discussed more extensively in the following chapter.

⁷⁰ A modification of the guiding principle in Jeffrey's theory will be presented in chapter 7. The principle of maximising conditional expected utility is replaced by the principle of ratifiability. For some details see section 7.2

6.5 A comparison between the theories

In the final section of this chapter I comment on a comparison that can be made between Jeffrey's decision theory on the one hand, and Savage's and Ramsey's theories on the other. One feature of Jeffrey's theory is that the only primitives of the theory are propositions. Jeffrey's theory is more general than Savage's by imposing less structural conditions on the problems it accommodates and, in this sense, Jeffrey's theory is insensitive to the structure of a decision situation⁷¹. As we saw, the accommodation of a decision problem by Savage's decision theory requires that certain structural conditions be satisfied. These conditions are that acts, on pairs of which preference is defined, are functions from states/events to consequences. In addition the probability distribution over events does not depend on the act to be chosen and acts determine their consequences. In contrast, Jeffrey's theory requires only one condition of a decision situation in order to accommodate it, that is, that the decision alternatives between which an individual has preference form a complete, atom free Boolean algebra.

Jeffrey's decision theory is considered to be more advantageous than Savage's and Ramsey's for being both more general and simpler. Eells (1982) and other decision theorists point out, among other things, that Jeffrey's theory is general enough to accommodate Savage's theory as a special case. I consider here briefly one by one the advantages Eells claims Jeffrey's theory to have.

The first comment concerning a comparison of Jeffrey's theory with Savage's theory is that it is possible by using resources only from Jeffrey's theory to reformulate the basic assumptions of Savage's theory. Borrowing the interpretation of Savage given by Adams and Rosenkrantz (1980), the following assumptions implicit in Savage's theory can be reformulated in Jeffrey's theory. The assumptions are i) outcome functionality ii) act independence (of the obtaining or not of a state) iii) path-free desirabilities (the desirabilities of the outcomes not being influenced by the act that brings them about as results). When these assumptions hold, it is possible to show that Jeffrey's conditional expected utility rule can be stated in such a way that probabilities of the states are not conditional on the relevant act and the desirabilities apply only to outcomes.

⁷¹ See chapter 7, section 7.2 for an elaboration of this claim.

The point here probably deserves a more thorough investigation than the one I give. Generality is not a surprising feature of Jeffrey's theory given that the theory is little constrained by structural conditions. Two relevant interrelated questions here are a) how arbitrary and restricting are the assumptions i)-iii) and b) is it possible within a theory of Savage's style to formulate decision situations in which assumptions i)-iii) seem to fail? For example, Allan Gibbard and William Harper (1988) show how with resources from Savage's theory probabilities of states that have to be construed as act-dependent can be accommodated. Their theory makes use of epistemically act-independent counterfactuals. More generally, as it has already been argued, every decision problem can be reformulated in a way suitable to be accommodated by Savage's theory. Decision problems that do not satisfy conditions i)-iii) can be reformulated into equivalent, with respect to the preference ordering, decision problems that exhibit event-independence of acts and act-consequence determinism (Jeffrey 1976). By taking into account these considerations we can see that Savage's decision theory, though it requires of decision problems to satisfy certain structural conditions, it does not fall short of generality and strength.

The second advantage of Jeffrey's theory, pointed by Eells is relevant to a comparison with Ramsey's theory. The advantage is that in Jeffrey's theory there is no need of use of gambles. Eells's comment here points to a disadvantage of Ramsey's theory and I admit to being sympathetic to his criticism. The criticism is that in order to employ Ramsey's theory of belief measurement we have to assume that a subject who considers two outcomes, must also consider the gamble where which of the outcomes results turns on an ethically neutral condition believed to degree $\frac{1}{2}$. And if we assume that the connections linking the 'propositions' and the outcomes of a gamble are causal, then we will have to attribute to the subject strange causal beliefs that he might not have⁷². I think the problem here relates to the fact that gambles are artificial inventions. Use of gambles has turned out to be a useful device for decision theories; it is a useful device for measuring strength of attitudes. Because gambles envisaged as constructed devices for measuring strength of attitudes are artificial constructions, a theory that could prove to be sufficient without use of artificial gambles would be preferred. But here we should mention that the objection

⁷² For the causal connection worry see chapter 7, 7.3.

has no force against a theory in which gambles are envisaged as abstractions from decision situations. An interpretation of the case of a person being in a decision situation literally as accepting a gamble might be misleading. We should stress here that use of a notion of decision situation, that is, of a system of attitudes relevant to the making of a choice being identified in abstraction as a gamble is not vulnerable to the artificiality objection. And thus, Eells objection to Ramsey's theory does not carry over to Savage's theory. As we saw, Savage treats consequences as constant acts in order to establish a preference relation between them. But this move does not result in any bizarre attitude attributions to a subject who is taken to be having preferences between consequences.

The third and last advantage of Jeffrey's theory over Savage/Ramsey discussed by Eells concerns the distinction between states and consequences. The account Savage gives of the distinction between the two categories of terms is very rough. "States" are states of the world, while "consequences" are states of the person (what might happen to a person). Eells suggests this distinction to be problematic because 'states of the world' and 'states of a person' constitute the same domain. States of the person are states of the world and what happens to the world could be considered to be something that happens to the person. I think that the point here is not a fair one. Blurring or ignoring a distinction between states and consequences as Jeffrey suggests and Eells endorses amounts to blurring or ignoring a distinction inherent in the concept of an agent. The distinction is between states of the world that are independent of agency and states of the world that are brought about by choices of agents. In the occurrence of the former there is no parameter related to preference had by an agent about the state. In the occurrence of the latter there is a parameter that is related to preference. The distinction between states that are independent of agency and states absolutely dependent on agency is explicated Savage's theory by the conditions of event-independence of acts and of act-consequence determinism.

CHAPTER SEVEN

DECISION THEORY AND RADICAL INTERPRETATION

In this chapter I discuss the possibility of an integrated theory of decision and interpretation. I begin by making some comments on the kind of contribution that decision theory can make to interpretation theory. I then consider the relative merits of Jeffrey's and Savage's decision theories in this role, and give some grounds for thinking that Savage's theory will do better.

7.1 Integration of decision with interpretation

I start this section by reminding the reader of the central issues discussed in the first chapter concerning the integration of decision theory and interpretation theory. Such an integration is envisaged by Davidson as an answer to two problems that arose in his earlier theory⁷³. The first problem concerns the identification of the doxastic and affective attitudes of agents, an identification which is required for an understanding of their actions and speech. The second problem concerns the possibility of interpretation of sentences that contain theoretical terms, and so are less directly keyed to observation. In the next few paragraphs I review my discussion of the two problems.

The first problem

The importance of an answer to the first problem can be seen by considering two issues. The first issue is related to the principle of charity. Employment of a theory that is unable to distinguish doxastic and affective attitudes will result in a systematic misapplication of the principle of charity. The principle of charity is the principle according to which sentences are true by the interpreter's standards when *held true* by speakers. The principle of charity is indispensable for interpretation and consequently the need to identify each kind of attitude is central. Essential to this point is the assumption that sentences related to doxastic attitudes are *held true* and are truth-evaluable while sentences related to affective attitudes are not *held true* and are not truth-evaluable. The second issue is related to another principle that is indispensable for interpretation, namely, the principle of rationalisation. Two

⁷³ For instance, see Davidson, 1980, 1985, 1990.

sentences *s* and not *s* can both be related to attitudes of a rational agent provided they are held with different attitudes: the first with an affective attitude (*s* is related to a desire) and the second with a doxastic attitude (not *s* is related to a belief). Identifying the kind of attitude involved in each ascription is thus necessary for determining whether the ascription can be rational or not.

The second problem

The second problem of Davidson's theory of radical interpretation whose solution invites integration with decision theory is the problem of interpretation of sentences less directly connected to observation and of sentences containing theoretical terms. In addition to observational sentences interpretable by use of observational schemes like (O),

(O) "Es regnet" is held true by a German speaker s at time t and at place x and it is raining at the vicinity of the speaker at time t,

there is a whole range of sentences in the repertoire of speakers that are less closely connected to observation. These sentences are keyed to observation in a less direct way than observational sentences. The way theoretical sentences are keyed to observation is that evidence for or against their truth is provided by their relations to observational sentences. For example, the observational sentence "It is cloudy." uttered at place *x*, at time *t*, provides evidence for the truth of the sentence "It is going to rain.". If we assume that the holding true attitude towards an observational sentence is related to a full belief of the subject toward the sentence, the interpretation of less connected to observation sentences requires identification of evidential relations to observational sentences. Relations of evidential support or evidential dependence are required not only for the interpretation of sentences that contain observational terms but are less closely connected to observation than sentences like "This is a horse.", "It is cloudy.". They are required for the interpretation of sentences that contain theoretical predicates. Confirmation of those sentences depends on relations of evidential support with observational sentences.

Davidson gives us an idea how the interpretative enterprise could then be completed.

"...It will be found, for example, that the agent is caused to award a high probability to some

sentence when and only when it is raining in the vicinity: the interpreter will then enter as a hypothesis (possibly to be abandoned when more structure becomes apparent) that this is a sentence that means that it is raining. More evidence for this interpretation will come from the probabilities; rain perceived under poor conditions for observation will cause lower probabilities; a down-pour experienced in the open is a probability of 1, or something close. More evidence still will accumulate as further sentences are given tentative interpretations. Thus a sentence interpreted as meaning that there is a patten on the roof, if held true (given a high probability) ought to increase the probability of the sentence interpreted as meaning that it is raining. In this way, by marking what the speaker takes as evidence for the truth of a sentence, it is possible to interpret sentences and words of an increasingly abstract and theoretical nature." (1982, pp. 16-17)

Davidson's suggestion is that a theory of interpretation needs a theory of degree of belief for determining dependence relations between sentences:

"What is needed for an adequate theory of belief and meaning, then, is not merely knowledge of what causes a speaker to hold a sentence true, but knowledge of the degree of belief in its truth. It would then be possible to detect degrees of evidential support by noting how changes in the degree of credence placed on one sentence were accompanied by changes in the degree of credence placed on other sentences." (1980, p. 7).

And

"Theory of meaning calls for a theory of degree of belief in order to make serious use of relations of evidential support" (1990 p. 322).

The suggestion above takes a more specific form by accommodating *Bayes* theorem. Evidential relations between sentences are specified by conditional probabilities that satisfy *Bayes* theorem. The evidential support that, for instance the sentence "It is cloudy." brings to bear on the truth of the sentence "It is going to rain." is specified by *Bayes's theorem*:

$$\text{pr}(h \text{ given } e)/\text{pr}(h)=\text{pr}(e \text{ given } h)/\text{pr}(e).$$

That is, evidence is relevant to the truth of a sentence, a hypothesis, if and only if the truth of the sentence is relevant to the evidence in the same way. The degree of evidential dependence between sentences is determined by conditional probabilities that satisfy the theorem. Interpretation of sentences less connected to observation depends on conditional probabilities which reveal the evidential relations between sentences. Conditional probabilities that reveal the relations of evidential support between sentences of a language are specified by the subjective or a subjective probability function that measures degrees of belief in the sentences of the language. The way conditional probabilities are related to a probability function is the following:

$$P(A|B)=P(A \cap B)|P(B), \text{ provided } P(B) \neq 0$$

In the case that the event B is the whole probability space conditional probability and probability coincide. If, then, a subjective probability function were

available for a speaker/agent, it would be possible to determine the subjective conditional probabilities between two (or more) sentences. It would, then, become possible to determine the evidential dependence relations between two (or more) sentences for the speaker/agent.

In chapter 6 of this dissertation we saw the relevance of decision theoretic considerations to determining a subjective probability function and a utility function. A subjective probability function yields degrees of belief and shows the doxastic attitudes of the agent to form a doxastic (probabilistic) relational system. A utility function provides a numerical representation of the affective attitudes of the agent and shows the affective attitudes to be members of a preference relational system. An integration of decision theory with interpretation would then promise an answer to the two problems stated above; the problem of classifying the doxastic and the affective attitudes of the subject and the problem of determining relations of evidential support between sentences. Were a decision theory to be accommodated by interpretation theory, the integration would result in a subjective probability function and a utility function of the speaker/agent. An answer to the two above-mentioned problems would then be given in more or less the following way. A classification of an attitude of a subject as doxastic or affective would be given on the basis of the numerical representation of the attitude. Doxastic attitudes would be the ones that have a numerical representation via a probability function and are members of a doxastic relational system, while affective attitudes would be the ones that have a numerical representation via a utility function and are members of a preference relational system. Relations of evidential support between sentences related to attitudes of the subject would be determined by a subjective conditional probability function derived from a subjective probability function.

The kind of contribution of decision theory to the integrated theory

An issue that has to be clarified here concerns the kind of contribution of decision theory to interpretation in the envisaged integrated theory. Davidson's own views about the issue seem to accord with the suggestion that decision theory imposes rationality constraints on the attitudes of the agent/speaker. Attribution of rationality to an agent whose mind, actions and language are under interpretation is

indispensable for understanding the linguistic and non-linguistic actions of the agent⁷⁴. In the case of full beliefs of an agent this might be cashed out in terms of logical consistency. In the case of partial beliefs rationality conditions amount to satisfaction of the axioms of probability. Rationality of desire is cashed out by the condition that desires are ordered so that, at least, the transitivity constraint is not violated. And there is a rationality constraint on the combination of degrees of belief with degrees of desire prescribed by the decision theoretic principle of maximising expected utility. It seems, then, that norms of rationality dictated by decision theory stand on a par with norms of logical consistency. Decision theory dictates norms of pattern in the network of attitudes of the agent whose language and actions are under interpretation. Norms of decision theory along with norms of logic are norms of pattern.

There is a clarification that is due here. It is, that though according to the suggestion above decision theory imposes structural norms on the interpretation enterprise the contribution is not supposed to be normative or prescriptive. The conditions imposed by decision theory are not meant to be suggestions about how people ought to believe, prefer, and act. The suggestion is that unless a certain pattern can be recognised in the system of attitudes of speakers/agents, it becomes questionable whether we can consider speakers/agents to have any beliefs or preferences or, to be performing actions. The relevance of decision theoretic considerations to an illustration of common sense notions for understanding agents like deliberation, belief and desire is pointed out by D. Lewis as well. In a passage where he considers the contribution of decision theory to interpretation theory he suggests the following:

“Decision theory is not an esoteric science, however unfamiliar it may seem to an outsider. Rather, it is a systematic exposition of the consequences of certain well-chosen platitudes about belief, desire, preference and choice. It is the very core of our common-sense theory of persons, dissected out and elegantly systematised” (1974, p. 114).

In line with these considerations a suggestion that could be made here is the following. The contribution of decision theoretic conditions on the range of preferences of an individual would be twofold for the interpretative enterprise. In addition to yielding numerical representations of attitudes and a subjective probability function for the agent (desirable as answers to the above-mentioned problems of

⁷⁴ For a discussion of the importance of rationality attribution to agents see also Davidson 1982(b).

interpretation), decision theoretic considerations would provide the framework within which an agent's actions, beliefs and desires are identified as such. If we envisage decision theory in this way, the decision theoretic introduction of doxastic and affective classification would not be a mere operational classification. The decision theoretic considerations on the preferences of individuals along with the introducing notions of doxastic and affective attitudes would be capturing essential aspects of the relevant attitudes. In this sense decision theory is not a prescriptive or a descriptive theory of agency but it is part of a core theory of persons.

Full acceptance of this suggestion would require dealing with issues that are beyond the scope of this dissertation. The position taken in this dissertation in relation to the above-presented suggestion is the following. Affinity between the notions of deliberation, preference, doxastic and affective attitudes as introduced by decision theoretic considerations with the pre-theoretic notions of deliberation, preference, belief and desire as figuring in a common-sense attitudes about persons would be a positive assessment of the decision theory considered. As a result, my position here is that the decision theory that introduces relevant notions close enough to their pre-theoretic counterparts is preferable as a candidate decision theory to integrate with a theory of interpretation to a decision theory in which there is lesser extent of coincidence.

Pattern and content

As was mentioned above the contribution of decision theory to interpretation theory concerns the pattern of the attitudes attributed to agents, the subjects of the interpretative enterprise. But in addition to identifying a pattern in the range of attitudes required so that agents can be seen as having beliefs, preferences and thoughts, understanding agents requires ascriptions of content. We have seen how the principle of charity, that is the principle according to which a sentence held true by an agent be true by the interpreter's standards, results in attribution of content for the range of held true sentences. As has already been discussed, determining the range of utterances on which the principle of charity is applicable is a pressing problem for a theory of radical interpretation. (chapter 1, 1.2.1).

The line of thought followed in this dissertation rejects the possibility of

determining the range of utterances on which the principle of charity applies on the basis of syntactical features of the uttered sentences (Chapter 3, 3.2). The suggestion followed up here is that the range of utterances on which the principle of charity is applicable would be determined on the basis of the attitude of the agent in making the utterance (Chapter 4, 4.1- 4.3). The attitude of the agent related to an utterance is, in turn, determined on the basis of rationality conditions on the range of actions of the speaker/agent (Chapter 6, 6.3- 6.4). The rationality conditions on the range of actions of the speaker/agent delimit a rational relational system on actions, a pattern of rational preferences between actions. What has been suggested here is that rationality conditions on the preference ranking between actions as introduced by decision theoretic considerations result in a classification of the doxastic attitudes. Certain attitudes from this class, namely the attitudes of fully or strongly believing, had by agents in making utterances constitute the range of the applicability of the principle of charity. So, the suggestion here is that on the basis of the pattern of attitudes attributed to agents applicability conditions of the principle of charity are determined. Applicability of the principle of charity provides conditions for ascription of content.

A distinction between two kinds of condition for understanding is suggested by Davidson as well:

“The ineluctable normative element in interpretation has, then, two forms. First, there are the norms of pattern: the norms of deduction, induction, reasoning about how to act, and even how to feel given other attitudes and beliefs. These are the norms of consistency and coherence. Second, there are the norms of correspondence, which are concerned with the truth or correctness of particular beliefs and values. This second kind of norm counsels the interpreter to interpret agents he would understand having, in important respects, beliefs that are mostly true and needs and values the interpreter shares or can imagine himself sharing if he had the history of the agent and were in compatible circumstances” (1985, p. 92).

In the light of those considerations we can see the subject of this dissertation as being the relation between two kinds of condition for interpretation; the relation between conditions of pattern and conditions of content. In this dissertation it is pointed out that conditions of pattern and conditions of content are not independent of each other. Conditions of content (charity) depend on conditions of pattern (rationality). This point can be reinforced by further considerations. These considerations relate to conditions of content ascription concerning non-truth-evaluable attitudes of speakers/agents. Let us retain the term *principle of charity* for

the principle according to which sentences *held true* by speakers be true by the interpreter's standards. And let us call *extended principle of charity* the principle which a) for the case of the *held true* sentences the principle coincides with the *charity principle* and b) for the *held good* sentences it is the principle that the sentences be good by the interpreter's values. The *extended principle of charity* is not foreign to Davidson's philosophy. It is worth citing here a well-known passage from Davidson's earlier articles that illustrates the *extended principle of charity*. The passage discusses the possibility of a theory of understanding a speaker/agent.

“In our need to make him make sense, we will try for a theory that finds him consistent, a believer of truths, and a lover of the good (all by our own lights, it goes without saying).” (1970, pp. 222)

The extended principle of charity provides conditions for content attribution of both truth-evaluable and non-truth-evaluable utterances. The principle, as it stands, requires a distinction between beliefs and desires. It is with respect to this aspect of the *extended principle of charity* that conditions of pattern are relevant. The pattern of the attitude attributions introduces two distinct notions (belief and desire) to which the *extended principle of charity* applies.

The *extended principle of charity* can be interpreted as suggesting that there are two distinct key concepts for the attribution of content to beliefs and desires. In the case of beliefs the key concept of content attribution principle is truth, while in the case of desires the key concept of content attribution principle is goodness. A very rough way of seeing the line of the suggestion would be the following.

(O) Speaker x utters “Es regnet” and the interpreter believes truly that it is raining at the vicinity.

(O') Speaker x utters “Gib mir wasser” and the interpreter desires the speaker to be given water.

Application of the *extended principle of charity* would suggest the following:

(M) “Es regnet” is a saying that it is raining.

(M') “Gib mir wasser” is a saying to be given water.

The suggestion requires a detailed account of the workings of the charity principles for a theory of understanding that is not given here. An attempt to provide

such an account would go beyond the limits of this dissertation⁷⁵. The discussion in the dissertation is centred round the relation between the two kinds of condition, both indispensable for interpretation, namely conditions of rationality and conditions of correctness. The suggestion here is that conditions of rationality determine the applicability of the charity principles.

At this point it might be help the reader to be reminded of some issues discussed in the dissertation that are related to the present considerations. These issues are:

a. The *sufficiency worry* (chapter 1, 1.2.1). This is the issue of whether one key concept (like that of truth or assertibility) is sufficient for a theory of interpretation. The line of thought we decided to adopt concerning the sufficiency worry was that it would be on the basis of anthropological principles, like conditions of rationality, that the sufficiency of one key concept will be granted. The *extended principle of charity* can be interpreted as challenging the sufficiency of one key concept for interpretation. Rationality conditions require two notions indispensable for understanding, the notions of truth and goodness. The *sufficiency worry* poses the question of whether one key concept would be sufficient for understanding both doxastic and affective attitudes.

b. Logical analysis of propositional attitude sentences. In discussing the paratactic analysis of propositional attitude sentences we concluded that certain aspects of a logical analysis of propositional attitudes will be decided by a substantial theory of propositional attitudes (chapters 3, 4). The particular aspect of a logical analysis of propositional attitudes that as suggested will be decided on a substantial theory of propositional attitudes is the common content assumption.

The points a. and b. are reintroduced in 7.4.

The resume of the discussion in this section is that an integrated theory of decision and interpretation promises an answer to the problem of a classification of the attitudes on which the *principle of charity* or the *extended principle of charity* are applicable. In addition it would provide an answer to the problem of interpreting less close to observation sentences and sentences containing theoretical terms. What remains to be discussed is how an integrated theory of decision and interpretation can

⁷⁵ See, e.g. Susan Hurley 1989, ch.5, for a discussion of the issue.

be incorporated. Given the constraints of a radical interpretation theory the identity of meanings of sentences to be interpreted and the identity of the contents of attitudes of speakers/agents are not in the data of the theory. In effect, in a radical decision theory the identity of options among which an agent shows preference or indifference is not granted.

7.2 Jeffrey's radical decision theory

Let a Jeffrey-style radical decision theory be Jeffrey's decision theory constrained so that the identity of propositions towards which a speaker/agent has attitudes not be in the data of the theory. Let us then see some details of Davidson's suggestion for a Jeffrey-style radical decision theory. Pace Davidson, let the assumption for a Jeffrey-style radical decision theory be that all attitudes (belief, desire, preference) are attitudes towards uninterpreted sentences. This is the assumption that frames the radical perspective Jeffrey's theory is invited to accommodate.

The suggestion has a premise. The premise is that all (uninterpreted) sentences constitute a complete, atom-free Boolean algebra. Davidson's suggestion for a replacement of propositions by sentences in Jeffrey's decision theory requires that the structural condition of Jeffrey's decision theory for propositions hold for sentences. The premise is that the set of sentences that constitute actual and potential utterances of speakers/agents of the language under interpretation should be a complete, atom-free Boolean algebra (minus the null element).

Under this premise, the integration of Jeffrey's decision theory with a theory of radical interpretation requires the identification of the logical connectives for the language under interpretation. Given the identification of the logical connectives for the language we can get the workings of Jeffrey's decision theory off the ground.

It has already been mentioned that Davidson, in his later papers, suggests that the empirical primitive of a theory of radical interpretation be the attitude of preference had between the truth of two sentences (chapter 1). This is to be contrasted with Davidson's suggestion in earlier papers that the empirical primitive of the theory be the attitude of *holding true* a sentence. The attitude of *holding true* a sentence embodies belief and meaning. The attitude of preference between the truth of

two sentences embodies belief, desire and meaning. On this ground the suggestion of considering the empirical primitive of the theory to be the attitude of preference between the truth of two sentences is a more promising one as far as the capacity of the interpretation theory is concerned.

Following this line, the possibility of a Jeffrey-style radical decision theory can be shown by showing how on the basis of the attitude of preference between two sentences the logical connectives of the language under interpretation can be identified. Davidson shows this by showing how the Sheffer stroke (not both) for the language can be identified solely on the attitude of preference between two sentences⁷⁶. As all logical connectives of a language can be defined from the Sheffer stroke, it has been shown that Jeffrey's decision theory can accommodate a radical perspective. Applying the axiomatic system of the Jeffrey-Bolker decision theory on the preference relation between sentences yields subjective probability and subjective utility representations of the sentences; beliefs and desires are determined. A Jeffrey-style radical decision theory answers the problem of classifying the doxastic and the affective attitudes of a speaker/agent and yields a subjective probability function required for determining relations of evidential support between sentences.

My criticism of a Jeffrey-style radical decision theory is twofold. Firstly, I attempt an assessment of a Jeffrey-style radical decision theory on the basis of the degree of affinity of the introduced notions of deliberation, belief and desire to their pre-theoretic counterparts. And secondly, I point to a problem of a Jeffrey-style radical decision theory resulting from the uniqueness conditions of the theory.

I discuss first the first line of criticism. In presenting Jeffrey's decision theory in chapter 6 we encountered the view that the theory is simpler and more general than alternative theories. All attitudes are attitudes towards the same kind of entities, i.e. propositions/sentences. Possible courses of actions as well as states of the world and consequences of the action are represented by propositions/sentences that have numerical representation by probability and desirability (utility). Here I present a view

⁷⁶ In Davidson (1990, appendix) this is shown in the following way. Let ' \cdot ' be a truth functional operator. Consider the following condition:

If $des(s) > des((t|u)|((t|u)|(t|u)))$ then $des((t|u)|((t|u)|(t|u))) \geq des(s|s)$, and
if $des((t|u)|((t|u)|(t|u))) > des(s)$ then $des(s|s) \geq (des((t|u)|((t|u)|(t|u)))$.

If this condition holds for all sentences s, t, u and for some s, t $des(s|s) \neq des(t|t)$ then the only interpretation is

according to which the generality of Jeffrey's decision theory shows the theory to be insensitive to some features of a decision situation. These are structural features of decision situations that are indicated by pre-theoretic intuitions about deliberation. The following clarify further this claim.

One respect in which there is a point in assigning probability or desirability to a sentence/proposition is to point to the kind of relevance the sentence/proposition has in deliberate behaviour of speakers/agents. Assignment of probabilities to a class of attitudes amounts to assignment of a certain role of those attitudes in the rationalisation of an action; assignment of utilities (desirabilities) to another class of attitudes amounts to assignment of a different role of those attitudes in the rationalisation of action. Decision theoretic considerations intend to capture the doxastic and affective role of certain ascriptions of attitudes. It was with respect to the problem of disentangling two co-operative forces that issue in deliberate behaviour, beliefs and desires, that decision theoretic considerations were put forward here.

Let us apply now some considerations from Jeffrey's theory of decision on a simple example. Consider the case of my action of taking my umbrella with me this morning. The courses of action under consideration here are specified by the sentences/propositions "I take my umbrella" and "I do not take my umbrella". Jeffrey's theory would suggest that the choice between the two options is determined by how desirable and how possible the two options are for the subject. That is, the choice is determined by the subjective probabilities and desirabilities of the propositions that I take my umbrella and that I do not take my umbrella. There is a sense in which the probability of the proposition/sentence that I take my umbrella with me relates to the probability of the proposition/sentence that I will choose to take my umbrella with me. I believe that I will take my umbrella with me if I believe that I will choose to take my umbrella with me. Let us leave out of consideration here cases in which I fail to carry out the action intended, if, for instance, while intending not to take my umbrella unbeknownst to me someone brings me into a state of actually carrying an umbrella. So the probability assigned to the proposition/sentence that I will take my umbrella with me relates to the probability assigned to the

proposition/sentence that I will choose to take my umbrella with me. Indeed, if Jeffrey's theory is not to compromise in generality by introducing distinctions between decisions and choices within propositional contents, the probability of a proposition specifying an action is the probability of the proposition specifying that the action is chosen. 'I take my umbrella with me' and 'I choose to take my umbrella with me' express the same proposition. Since choices are made on the basis of probabilities and desirabilities it seems that we have the result that my choice of taking my umbrella with me this morning depends on how desirable it is to take my umbrella and, surprisingly enough, on how probable it is that I choose to take my umbrella with me this morning. My belief that I will choose a certain course of action features as one of the conditions that explains my choice of the action.

The above example does not exhibit any accidental feature of Jeffrey's theory. Jeffrey's decision theory is an alternative to classical Bayesian theories (e.g. Savage's theory) in which no independence postulate of probabilities (of events) on actions is considered. Subjective probability conditions for deliberation are taken to be dependent on the act chosen. Moreover, the general rule in Jeffrey's theory is that probabilities dependent on the choices are the credence conditions for deliberation.

The point I raise here is that credence of acts featuring as conditions for deliberation seems paradoxical given some pre-theoretic accounts of deliberation. The pre-theoretic accounts of deliberation to which I appeal suggest the following. An agent in deciding which available course of action to pursue compares the available courses of action (on a desirability scale). In order for this comparison to take place the agent appeals to his subjective probability distribution over possible states (which in Jeffrey's theory include the actions). This suggests that the comparison between available courses of action requires a prior subjective probability distribution. On the other hand, in Jeffrey's theory, probabilities of acts featuring as conditions for deliberation seem to prejudge the outcome of deliberation.

The pre-theoretic feature of deliberation conjured up here constitutes the following condition. Namely, the condition is that in a decision situation in which one or more courses of action are envisaged, predictions as to which course will be chosen are not parameters for the decision making. If so, the degree of belief of the agent that he chooses an action does not contribute to the comparison between

actions on a desirability scale that determines the choice of the subject. Though it is not denied that an agent might have degrees of belief that he will take one course of action rather than another and that the agent might make predictions about the choices he will finally make, these are not beliefs relevant to the rationality of the choice of the agent.

The condition appealed here is cited in the literature as the ‘Levi-Spohn embargo’ (Bacharach and Hurley 1991). The requirement that decision theory should respect the above-mentioned pre-theoretic condition concerning the structure of a decision situation is clearly stated in Spohn (1977) and Levi (1991).

“The decision-maker should not assign either determinate or indeterminate credal probabilities to hypotheses as to how he will choose” (Levi, 1991, p. 99).

Spohn comments that the probabilities as to how an agent will choose are irrelevant to the choice of the agent. Levi strengthens this claim by noting the following:

“X might be concerned to predict his own choices, but when he does so, he is not functioning as a deliberate agent concerned to identify which of these options he is not rationally prohibited from making...X might very well take such a predictive explanatory attitude toward the choices of other agents or toward his own choices at future times” (Levi, 1991, p. 99).

In order to illustrate better the appeal of the discussed pre-theoretic condition I present a comment made by Spohn (1977) that shows the strangeness of letting credence of acts feature as conditions for deliberation. The point relies on a condition that constitutes common ground among decision theorists, that is, that subjective probabilities can be detected by the behaviour of the agent in betting situations with certain odds and various stakes. If we let credence of acts feature as conditions for deliberation we can consider betting situations for detecting the subjective probabilities of acts. In such a kind of betting situations the bets offered to the agent would be constructed on the condition that the agent chooses a certain act. This kind of betting situations is strange because all that mattered for accepting the various bets would be the various payoffs. If the payoff of choosing a certain act A is higher than the payoffs of choosing the alternative acts, the subjective probability of the choice A is very close to 1. By changing the payoffs of the same range of choices the subjective probability of A changes⁷⁷.

⁷⁷ It is worth noting here that the pre-theoretic condition that beliefs as to how the agent will choose should not enter as credential parameters in the choice-worthiness comparison of acts is generally respected by causal

The condition that credences of acts should not figure as conditions for deliberation is violated in Jeffrey's modified decision theory (Jeffrey 1983) as well. In Jeffrey's modified theory the principle of conditional expected utility is replaced by the principle of ratifiability. *Ratificationism* is suggested in order to deal with Newcomb-like decision problems⁷⁸. A ratifiable decision is defined as a decision to perform an act of maximum estimated desirability relative to the probability matrix the agent thinks he would have if he finally decided to perform that act. *Ratificationism* is the maxim of making ratifiable decisions.

I come now to the second and more serious worry of Jeffrey's decision theory as far as the integration of Jeffrey's decision theory with interpretation theory is concerned. It results from the uniqueness conditions of Jeffrey's decision theory. As we saw in chapter 6, in contrast to classical decision theories (e.g. Savage's theory), the representation theorems of Jeffrey's decision theory do not issue in unique subjective probability and desirability functions for the preference ranking of an agent. The uniqueness conditions of Jeffrey's theory are the following⁷⁹.

P, D, p, d satisfy the Kolmogorov axioms of probability and the desirability axiom of Jeffrey's theory if and only if

For certain real numbers a, b, c, d

ad -bc is positive

desA + d is positive

decision theories. It is definitely outside the scope of this dissertation to attempt an introduction of issues from causal decision theories. But it is worth mentioning that causal decision theories are developed with respect to certain structural conditions imposed on decision problems and promise good answers to many Newcomb-like problems. See Gibbard A. and Harper W. (1978), Lewis D. (1981), Skyrms B. (1982). Especially from Skyrms' theory we can raise the interesting question of whether Savage's structural conditions (act-independence of events and consequence determinism from acts) can accept a causal reading. Here, suffice it to say that Skyrms theory is developed with respect to Savage's structural conditions.

⁷⁸ There are two boxes box1 and box2 and the agent can choose between the content of box1 and contents of both boxes. The agent knows that the box2 contains £1k and that £1m or nothing has been put by a Demon in box1. The conditional probabilities that there is £1m in box1 on choosing box1 and on choosing both boxes are respectively 0.9 and 0.1. According to the principle of maximising conditional expected utility it is rational for the agent to choose box 1. But this result clashes with the principle of dominance, since the choice of both boxes dominates the choice of box1. On the other hand, the ratifiable decision would be the one that since it is performed will have the highest desirability among performance of the alternatives.

The literature on Newcomb-like problems is huge. For a formulation of the problem see Nozick R. (1969). For a general discussion of Newcomb-like problems see Eells (1982). For problems of Jeffrey's theory concerning Newcomb-like decision situation see Gibbard A. and Harper W. *op. cit.*, Lewis D. *op.cit.*, Skyrms B. *op.cit.* For some elaboration of the notion of ratifiability see Skyrms B. (1990).

⁷⁹ The Jeffrey-Bolker axiomatised decision theory yields unique probability and desirability (to a linear transformation) representations in the case that the assignment of values (desirabilities) is unbounded both above and below. See Jeffrey (1983, p, ch. 6) for details.

$$c \text{ des } T + d = 1$$

$$\text{PROB } A = \text{prob } A (c \text{ des } A + d)$$

$$\text{DESA} = (a \text{ des } A + b) / (c \text{ des } A + d).$$

That is, from the relation of preference between propositions as determined by Axioms 1-4 of Jeffrey's axiomatised theory of decision the theoretical constructs of subjective probability and desirability are not uniquely determined. For every pair prob , des , that satisfies the existence condition, i.e. prob is a probability and des satisfies the desirability axiom, such that des preserves the preference ranking, another pair PROB , DES is determined satisfying the existence and the uniqueness condition above such that DES preserves the preference ranking. The family of such pairs is infinite since for the pair PROB , DES there is another pair PROB^1 , DES^1 that satisfies the existence and uniqueness condition with respect to the same preference ranking.

Let us now discuss the significance of this feature of Jeffrey's theory for the integrated Jeffrey-style radical decision theory. I remind the reader that decision theory is invited to contribute to radical interpretation theory by introducing subjective probability and desirability functions. Within a theory of interpretation the subjective probability function is used for determining relations of evidential support between sentences required for interpretation of sentences containing theoretical terms and interpretation of sentences not directly keyed to observation. The different representations of attitudes by the two functions enable a distinction between doxastic and affective attitudes of an agent. Since in Jeffrey's theory the probability function is not uniquely determined, but it belongs to an infinite family of probability functions as suggested by the uniqueness condition, the subjective probability function that in a Jeffrey-style radical decision theory mirrors the relations of evidential support between sentences becomes indeterminate. From the relation of preference between sentences that is the empirical primitive of the theory the evidential relations between the doxastic attitudes/sentences are not uniquely determined. It is left undetermined from preference rankings which probability function from the infinite family of probabilities that respect those rankings determines the relations of evidential support between the attitudes of the agent. The relevant indeterminacy is bequeathed to the subjective conditional probability

functions for the agent. That is, in Jeffrey's theory the preference ranking does not determine which one of the infinite probs would be used in every instance of applying *Bayes theorem*. This would result in an enormous indeterminacy concerning interpretation of sentences containing theoretical terms and sentences less directly keyed to observation.

Davidson (1990, fn pp. 323-324) comments that the feature of a Jeffrey-style radical decision theory of not yielding unique subjective probability representations from preferences is acceptable as another sort of indeterminacy that already exists and is expected in radical interpretation theory. But diminishing indeterminacy is a desirable property of interpretation theory. If so, an alternative radical decision theory that would result in unique probability representations of doxastic attitudes would be preferable to a radical decision theory that like a Jeffrey-style radical decision theory results in an infinite number of doxastic representations.

7.3 The Savage-style radical decision theory

Let us now consider the possibility of an integration of Savage's decision theory with radical interpretation theory. In a way similar to the way we considered a Jeffrey-style radical decision theory, i.e. as Jeffrey's decision theory in which the identity of propositions is not granted, a Savage-style radical decision theory will be Savage's decision theory in which the identity of acts is not granted. The identity of acts, which in Savage's theory are seen as functions from states/events to consequences is not in the data of the theory. The options among which the agent shows preference or indifference are represented by matrices in which the entries are uninterpreted sentences.

I start my discussion of the Savage-style radical decision theory by considering Davidson's and Jeffrey's objections. Davidson formulates his objection against Ramsey's theory and the essential use of gambles in that theory. Gambles are complex entities and they are difficult for the interpreter to identify. A gamble, though it can be represented by a sentence/proposition, involves a relation between sentences/propositions. The objection has it that the relation between sentences/propositions involved in gambles is causal. This is an aspect of gambles to which both Davidson (1984, 1990) and Jeffrey (1965, 1983) object. A gamble specifies causal connections between a condition (the coin lands head up) and more

or less desirable outcomes (you win £100, you lose £100). Since, in addition, Ramsey's theory requires the existence of arbitrary gambles, the theory would require attributions of arbitrary causal connections between the attitudes of agents. For instance, take the case that among the consequences are that there will be a thermonuclear war next week and that there will be fine weather next week. And among the conditions is that a coin lands head up. In this case, the agent needs to have preference for the gamble that there will be a thermonuclear war next week if the coin lands head up and there will be fine weather next week if not. In order for the agent to entertain this gamble he needs, the objection goes, to consider causal relations between the condition and the consequences. This is weird, because it seems to suggest that the agent needs to revise his beliefs about the causes of a war (Jeffrey, 1983, p. 157).

Another objection to Ramsey's theory, coming from Davidson, is that use of gambles in Ramsey's theory requires the existence of ethically neutral conditions; that is conditions whose obtaining or not has no value in itself. For an incorporation of Ramsey's theory into a radical perspective identification of ethically neutral conditions would be required.

Here we will discuss the objections by seeing how they can be reformulated so that they apply to Savage's theory. Since acts in Savage's theory are the structural equivalent of gambles in Ramsey's theory, the reformulation would go along the following line. Acts, as functions from states/events to consequences are complex entities embodying causal relations. Acts can be represented by complex sentences that specify causal relations between events and consequences. Events have to be identified as ethically neutral; the bearers of utility, acts/consequences and the bearers of subjective probability constitute distinct classes.

Let us first discuss the causal connection worry. The conditions that determine the notion of gamble in Ramsey's theory and the notion of act in Savage's theory do not need to involve causal connectives⁸⁰. Acts in Savage's theory are seen as functions between events and consequences so that the two following conditions hold:

⁸⁰ For a non causal formulation of Ramsey's gambles see Levi I. (1966).

1. Act-consequence determinism: where C_i is consequence A is act and E_i an event such that $A(E_i)=C_i$,

$$P(C_i|A \cap E_i)=1$$

2. Act independence of events: where A is act and E is event,

$$P(E|A)=P(E)$$

Conditions 1) and 2) do not require causal connectives. So, the part of the Jeffrey-Davidson objection against gambles as notions that involve causal relations does not hold⁸¹.

Let us now discuss the objection against use of gambles/acts based on their complexity. The point I think we should consider in relation to this objection is that the complexity of the notion of gamble in Ramsey's theory and of the notion of act in Savage's theory mirrors the structural complexity of decision situations. The motivation for introducing structural conditions to the notion of act is that the structural complexity of the options corresponds to the structure of decision situations. The structural conditions inherent in the notion of act intend to grasp (more or less effectively) a distinction within the parameters of deliberation. The doxastic parameters that are outside the influence of the actions in question and the affective parameters that are absolutely dependent on the actions. The suggestion here is that a distinction between those two parameters of the notion of action mirrors a structural distinction in the notion of deliberation. If so, the complexity of Ramsey's gambles and of Savage's acts is not unmotivated.

The last thing we need to show in order to defend our proposal for a Savage-style radical decision theory is how Savage's theory can accommodate a radical perspective. In presenting Davidson's suggestion for a Jeffrey-style radical decision theory we saw how a Jeffrey-style radical decision theory is possible. This was shown by delivering the logical connectives of the language (Sheffer stroke) from the relation of preference or indifference between (uninterpreted) sentences.

There is a comment that needs to be made at this point. The possibility for Jeffrey's theory of accommodating a radical perspective was not shown on the basis of actual choices only. Davidson shows how the logical connectives of the language

⁸¹ The observation that gambles/acts can be accounted without use of causal connectives does not clash with a comment made in the previous section, namely that Savage's theory can accommodate a causal reading.

can be determined in a radical way by showing how the Sheffer stroke can be determined on the basis of preference or indifference between uninterpreted sentences (fn 76). The pattern of preferences between uninterpreted sentences that would determine that the sentences are composed by the Sheffer stroke involves both actual and counterfactual choices of the agent. The point of the suggestion is to show that from the relation of preference or indifference between uninterpreted sentences (the empirical primitive) the structural condition of Jeffrey's decision theory -objects of belief, of desire and preference form an atom free complete Boolean algebra- can be approached. The analogous argument required for showing the possibility of Savage's theory to accommodate a radical perspective would have to show that the structural conditions of Savage's decision theory can be tackled on the basis of the relation of preference or indifference between uninterpreted acts. The objects of preferences will, in the context of a Savage-style radical decision theory, be represented by matrices in which the entries are uninterpreted sentences.

The act-consequence determinism condition of Savage's theory can be addressed by representing acts as functions between sentences. The other structural condition of Savage's theory is the event independence of acts. In order to address from preference or indifference the independent condition we need to show that from preference or indifference we can determine that one sentence is irrelevant to another. That is, what is required for showing that independence structural conditions of Savage's theory holds is to be able to tell from a relation of preference or indifference only that the condition $P(E|A) = P(E)$ holds for two sentences A, E. Since a subjective probability function is not available in advance of the theory we will have to show that the irrelevance condition can be approached by the relation of qualitative probability. We need to be able to show from a relation of preference or indifference that the events (sentences figuring as arguments of act functions) E given A and E are equivalent with respect to qualitative probability.

The general task of showing two events (sentences) to be independent is not straightforward from the resources of Savage's theory. The reason is that it is not straightforward how to determine a personal probability relation 'not more probable than' between (B given C) and (G given H) for any four events B, C, G, H. The relation of preference between acts f, g given an event B that determines the relation 'not more probable than' between event A given B and event C given B is defined

only for the same event B⁸².

In order to deal with this problem of a Savage-style radical decision theory we will appeal to the notion of conditional act and to Axiom 2' that introduces an extended notion of preference between conditional acts. In chapter 6 we saw how a notion of qualitative independence can be defined from the notion of preference between conditional acts. Since this condition is granted, it is possible to determine a notion of independence between (uninterpreted) sentences from the notion of preference between (uninterpreted) conditional acts.

The conditions to which we appeal for addressing the independence structural condition of Savage's decision theory requires the identification of logical connectives between sentences. For this problem I suggest we use a modification of Davidson's suggestion for determining the logical connectives of the Boolean algebra of propositions/sentences in the context of a Jeffrey-style radical decision theory. The modification of Davidson's suggestion about how to determine the Sheffer stroke (not both) from a relation of preference or indifference that needs to be made here would take as objects of preference or indifference functions between composed sentences ($A|B$, '|' the Sheffer stroke here) and appropriate consequences. These considerations intend to show the possibility for Savage's decision theory for accommodating a radical perspective. From the relation of preference or indifference between uninterpreted acts the structural conditions of Savage's decision theory can be approached.

7.4 A comparison between Jeffrey's and Savage's radical decision theories

In the two previous sections of this chapter the possibility of a Jeffrey-style radical decision theory and the possibility of a Savage-style radical decision theory were presented and discussed. Here I present a comparison between the two suggestions.

In section 7.1 of the present chapter we discussed the kind of contribution decision theoretic considerations are required to make to a radical interpretation theory. One respect in which decision theory is relevant to interpretation theory is

⁸² The problem of qualitative probability with independence has been approached by B. O. Koopman (1940), R.D. Luce (1968),

that decision theoretic considerations can be used to elucidate pre-theoretical concepts like belief, desire and preference that are employed in understanding actions of speakers/agents. On this ground it has been suggested that the degree of affinity of the introducing notions of a decision theory to their pre-theoretic counterparts is relevant to the assessment of a decision theory. The assessment concerns the incorporation of a theory of decision to a theory of understanding. In discussing a Jeffrey-style radical decision theory we saw that the theory disregards some pre-theoretic intuitions about the notion of deliberation. These pre-theoretic intuitions constitute a condition that was named 'the Levi-Spohn embargo'. That is that credences of acts should not figure as conditions for deliberation. In contrast to Jeffrey's decision theory, Savage's decision theory respects the condition. The credences that constitute parameters for deliberation are independent of the acts to be chosen. Savage's theory respects the particular condition and in the relevant respect its introducing notion of deliberation - along with its belief, desire and preference components- has a greater degree of affinity than Jeffrey's theory to its pre-theoretic counterpart.

Another respect in which decision theoretic considerations are relevant to interpretation theory is that decision theories introduce probability and utility functions that can be used in order to cope with certain problems within interpretation theory. These problems are the determination of relations of evidential support between sentences required for interpretation of some sentences and a classification of the doxastic and affective attitudes of speakers/agents required for the applicability of the principle of charity and for the rational embedding of the attitude attributions. With respect to the first problem we saw that a Jeffrey-style radical decision theory introduces an enormous indeterminacy concerning the evidential relations between sentences for a subject that has a certain pattern of preferences. The subjective probability function that would determine the relations of evidential support between sentences is not uniquely determined in Jeffrey's theory of decision. On the other hand Savage's decision theory introduces a unique subjective probability function. The theory yields for an agent who has preference relations that conform to the rationality constraints of the theory between options that conform to the structural constraints of the theory a unique subjective probability function. Such an agent, Savage's theory suggests, has a unique pattern of evidential

relations that characterises his doxastic system of attitudes.

Another point of comparison between a Jeffrey-style radical decision theory and a Savage-style radical decision theory is the common content assumption. Jeffrey's theory requires the common content assumption. The structural condition of Jeffrey's decision theory is that belief, desire and preference are attitudes towards propositions. The domain of the objects of belief, desire and preference in Jeffrey's theory constitutes a complete, atom-free Boolean algebra (minus the null element). On the other hand, Savage's theory is liberal concerning the assumption. In Savage's theory the assumption for the domain of events is that it constitutes a Boolean algebra, while the assumption for the domain of consequences is that it is a finite ordered set. Subjects to doxastic-probabilistic considerations are events while subjects to value considerations indicated by the preference ordering of utilities are consequences. In addition, acts, which are functions from events to consequences, are subjects to expected value considerations indicated by expected utility preference ordering. The expected utility considerations that characterise the domain of act-functions embody both doxastic-probabilistic considerations –inherited from the doxastic component of acts which are events- and utility considerations –inherited from the value component of acts which are consequences.

Now in Chapters 2 and 5 of the dissertation it was pointed out that the common content assumption between doxastic and affective speech acts (chapter 2) and between doxastic and affective attitudes/states (chapter 5) could be challenged. We considered some conditions for identifying content of utterance/propositional attitude and it was pointed out that those conditions do not grant the common content assumption. Needless to say, the common content assumption is a useful theoretical tool for theories of understanding language and mind of agents that results in elegant theories of interpretation. The attempt in the dissertation has been to point out than an assumption considered uncontroversial and made almost uniformly by propositional attitude and speech act theorist might not be properly established. Taking those considerations into account suggests that Jeffrey's decision theory would not be compelling on the ground of the common content assumption solely. Thus, the aspect of Jeffrey's theory of taking the objects of belief, desire and preference to constitute a single domain does not, on its own right, present us with a compelling reason in favour of the suggestion of a Jeffrey-style radical decision

theory.

In discussing the logical analysis of attitude attribution sentences (chapter 4, 4.3) we suggested that we let a substantial theory about attitude attributions determine the common content condition across doxastic and affective attributions. The substantial theories of attitude attributions considered in the dissertation are elaborations of decision theoretic considerations. We considered Jeffrey's and Savage's theories of decision for an account of belief and desire attributions. In this chapter we presented some reasons for which Savage's decision theory is preferable to Jeffrey's theory as a candidate decision theory to be incorporated with radical interpretation. The reasons for the particular proposal are that Savage's theory does not violate the 'Levi-Spohn embargo' and it results in a unique subjective probability function for the agent/speaker. Now, Savage's theory does not require the common content assumption. Furthermore, concerning probability and utility representations, it suggests that the domain of doxastic-probabilistic considerations (events) and the domain of value-utility considerations (consequences) are not isomorphic. And we pointed out that there are certain structural differences between the domain of events and the domain of conditional acts. However these considerations do not constitute a proof that in Savage's theory the common content assumption is false. It might be possible, while keeping the spirit of Savage's theory (act-independence of events and act-consequence determinism) to modify Savage's theory so that it implies the common content assumption. In this case, given that Savage's theory is preferable to the alternative Jeffrey's theory, we will be offered with more comprehensive reasons for accepting the common content assumption. A substantial theory of attitudes would imply it. In case it is shown that such a modification of Savage's theory cannot be made and the theory implies that the common content assumption is false, we will have to distinguish between two kinds of content. In this case Savage's decision theory would ground the project interpretation theory in which there would be a distinction between two kinds of theoretical entities corresponding to attitude attributions. A substantial theory of attitudes would suggest that in a theory of interpretation two kinds of content are introduced. In case that, for reasons independent of the issue of the common content assumption in Savage's theory, a proof of the assumption is offered, then acceptance or rejection of Savage's theory will be decided on whether it can accommodate the assumption or not.

7.5 Some closing remarks

The background suggestion adopted in the dissertation is that a theory of interpretation results in redescriptions of utterances and doings of speaker/agents. The utterance of a speaker “Gib mir wasser” is redescribed as a certain saying. It is redescribed as a saying to the hearer to be given water. A speaker utters the sentence “Es schneit”. The utterance is redescribed as a saying that it is snowing. Redescriptions of this kind satisfying certain conditions enable understanding of the utterances of speakers. A condition that redescriptions need to satisfy is that the attitudes ascribed to the speaker by those redescriptions show the speaker to be a rational. They are attitudes that could be part of a system of attitudes that would show the preferences of the speaker/agent to satisfy certain constraints of rationality. The rationality constraints considered here on preferences are introduced by decision theoretic considerations.

The general problem that pertains the dissertation is the possibility of a distinction between truth-evaluable and non-truth-evaluable utterances in the context of radical interpretation theory. It is pointed out that the possibility of a distinction between truth-evaluable and non-truth-evaluable utterances is a substantial problem for a theory of radical interpretation. One projection of this problem is the issue of the applicability of the principle of charity. Another projection of the problem is the rational embedding of utterances. The suggestion made in this dissertation is that the required distinction could be given on the basis of a classification within the attitudes attributed to an agent. The classification within the attitude attributions is one between doxastic and affective attitudes. Utterances related to doxastic attitudes – expressing or ascribing a doxastic attitude- are classified as truth-evaluable utterances. Utterances related to affective attitudes – expressing or ascribing an affective attitude- are classified as non-truth-evaluable utterances.

A theory of interpretation that results in a redescription of the utterance “Gib mir wasser” that would ascribe an affective attitude to the speaker related to the utterance would classify the utterance “Gib mir wasser” as one not susceptible of truth evaluation. This utterance is one beyond the domain of applicability of the principle of charity. A theory of interpretation that suggests a redescription of the utterance “Es Schneit” ascribing a doxastic attitude to the speaker in making the

utterance would classify the utterance as one susceptible of truth-evaluation. The theory that introduces doxastic classifications considered here introduces degrees of doxastic attitudes. In case the redescription ascribes a doxastic attitude with degree 1 (or close enough to 1) the utterance is classified as one within the domain of applicability of the principle of charity.

Needless to say there are many problems that are left unanswered in this dissertation. The detailed workings of the principle of charity were not discussed here. Moreover the workings of what was called *the extended principle of charity*, were left completely out of consideration. The question of whether a modification of Tarskian mechanisms could be possible for the value-evaluable (non-truth-evaluable) utterances was left out of consideration. And though the difference between doxastic and affective content were stressed, the systematic connections between them were not discussed. Adequate answers to those questions are required for a complete assessment of the views presented in the dissertation. In particular, answers to those questions are required for showing the adequacy of a Savage-style radical decision theory as it was introduced and interpreted in the context of this dissertation. The attempt that has been made here is to show the possibility of an alternative radical decision theory that promises good answers to some problems facing a theory of radical interpretation.

REFERENCES

- ADAMS, E. & ROSENKRANTZ, R.D. 1980. "Applying the Jeffery decision model to rational betting and information acquisition", *Theory and Decision*, 12, 1-20.
- ANSCOMBE, E. 1957. *Intention*. Oxford: Blackwell.
- ARMON-JONES, C. 1991. *Varieties of Affect*. Harvester Wheatsheaf.
- AUSTIN, J. L. 1962. *How to do things with words*. Oxford University Press.
- BACCHARAH, M. and HURLEY, S. (eds.). 1991. *Foundations of Decision Theory*. Basil Blackwell.
- BALCH, M / McFADDEN, D. L./ WU, S. Y. (eds.). 1974. *Essays on Economic Behaviour under Uncertainty*. North-Holland| American Elsevier.
- BALCH, M. and FISHBURN P. 1974. "Subjective Expected Utility for Conditional Primitives". In BALCH, M/ McFADDEN, D. L./ WU, S.Y. eds. 1974. pp. 57-69.
- BLACK, M. 1970. *Margins of Precision, Essays in Logic and Language*, Cornell University Press, Ithaca and London.
- BLACKBURN K., W. 1987. "Davidson on Force and Convention", *Analysis* 47, pp. 72-74.
- BLACKBURN, S. 1975. "The Identity of Propositions", in Blackburn, S., 1975. pp.182-205.
- BLACKBURN, S., (ed.). 1975. *Meaning, Reference and Necessity*, Cambridge, New York: Cambridge University Press.
- BLOCK, N., (ed.). 1980. *Readings in Philosophy of Psychology*, vol.1 Cambridge, Mass.: Harvard University Press.
- BOLKER, E. D. 1965. *Functions Resembling Quotients of Measures*, Ph.D. Dissertation, Harvard University.
- BOLKER, E. D. 1967. "A simultaneous axiomatisation of utility and subjective probability. *Philosophy of Science*, 34, 333-340.
- BROOME, J. 1991. *Weighing Goods*, Basil Blackwell.

CARNAP, R. 1947 and 1956. *Meaning and Necessity, A study in Semantics and modal logic*, The University of Chicago Press, Phoenix Edition.

CARNAP, R. 1951. *Logical Foundation of Probability*. London: Routledge and Kegan.

COLLINS COBUILD. 1990. *English Grammar*. Collins Publishers, The University of Birmingham.

CHURCH, A. 1950. "On Carnap's Analysis of Statements of Assertion and Belief", *Analysis* 10, pp. 97-99.

DAVIDSON, D. 1968. "On Saying That", in Davidson 1984, pp. 93-108.

DAVIDSON, D. 1970. "Mental Events", in Davidson 1980, pp.207-227.

DAVIDSON, D. and HARMAN, G. (eds.). 1972. *Semantics of Natural Language*, Holland. D. Reidel Publishing Company/Dordrecht.

DAVIDSON, D., 1973a. "Radical Interpretation", in Davidson 1984, pp.125-154.

DAVIDSON, D., 1973b. "In Defence of Convention T", in Davidson 1984, pp. 65-75.

DAVIDSON, D.1974a. "Belief and the Basis of Meaning", in Davidson 1984, pp.141-154.

DAVIDSON, D., 1974b. "On the Very Idea of a Conceptual Scheme", in Davidson 1984, pp.183-198.

DAVIDSON, D. 1975. "Thought and Talk", in Davidson 1984.

DAVIDSON, D. 1976. "Reply to Foster", in Davidson 1984.

DAVIDSON, D. 1979. "Moods and Performances", in Davidson 1984, pp. 109-121.

DAVIDSON, D. 1980. *Essays on Actions and Events*, Oxford Clarendon Press, Oxford.

DAVIDSON, D. 1982a. "Communication and Convention", in Davidson 1984, pp.265-280.

DAVIDSON, D. 1982 b. "Paradoxes of Irrationality" in Hopkins, J. and Wollheim, R. (eds.). 1982.

DAVIDSON, D. 1982c. "Expressing Evaluations", The E. H. Lindley

Memorial Lecturship Fund.

DAVIDSON, D. 1984. *Inquires into Truth and Interpretation*, Clarendon Press, Oxford.

DAVIDSON, D. 1985. "A new basis for decision theory", *Theory and Decision* 18. pp. 87-98.

DAVIDSON, D. 1990. "The Structure and Content of Truth". *The Journal of Philosophy*, Vol. LXXXVII, No 6, June 1990, pp.279-328.

DAVIDSON, D. 1991. "Three Varieties of Knowledge". In Griffiths A.P. (ed.) 1991, pp.153-166.

DE FINETTI, B. 1937. *Foresight: Its Logical Laws, Its Subjective Sources*, in Kyburg and Smokler (eds.) 1964. pp. 93-158.

DE MORGAN, A. 1847. *Formal Logic*. London: Taylor and Walton.

DENNETT, D. 1981. "True Believers" in Dennett 1987.

DENNETT, D. 1987. *The Intentional Stance*. Cambridge Massachusetts, London, England. Bradford book, MIT Press.

DENNETT, D. 1975. "Brain writing and mind reading", *Minnesota studies in the Philosophy of Science*, 7.

DUMMETT, M. 1976. "What is a Theory of Meaning? (II)", in *Truth and Meaning, Essays in Semantics*, Evans and Mc Dowell (eds.) 1976.

DUMMETT, M. 1981 (2nd Edition). *FREGE Philosophy of Language*, Duckworth.

EELLS, E. 1982. *Rational Decision and Causality*, Cambridge University Press.

EVANS, G. and McDOWELL, J. (eds.). 1976. *Truth and Meaning, Essays in Semantics*, Oxford: Clarendon Press.

EVANS, G. 1981. "Semantic Theory and Tacit Knowledge" in *Wittgenstein: to Follow a Rule*. Holtzman S. and Leich C. (eds). 1981. pp. 118-137.

FIELD, H. 1977. "Logic, Meaning and Conceptual Role". *Journal of Philosophy* 74, pp. 375-406.

- FIELD, H. 1978. "Mental Representation", *Erkenntnis* 13, pp. 9-61.
- FISHBURN, P. 1970. *Utility Theory for Decision Making*. John Wiley & Sons, Inc. New York
- FISHER, R. 1959. *Smoking: The cancer controversy*. Oliver and Boyd. Edinburgh and London.
- FODOR, J. 1985. "Fodor's Guide to Mental Representation" in Stich, S. (eds.) 1994.
- FODOR, J. 1987. *Psychosemantics*, The MIT Press.
- FREGE, G. 1892. "On Sense and Reference", in Geach & Black (eds.) 1970, pp. 56-78.
- FREGE, G. 1918. "Thoughts", in *Logical Investigations*, Oxford Basil Blackwell 1977, pp. 1-30.
- GAERDENFORS, P. and SAHLIN, N-E. (eds.). 1988. *Decision Probability and Utility*, Cambridge University Press.
- GARNER, R. 1972. "On Saying What is True", *Nous* 6, pp. 201-224.
- GEACH, P. 1965. "Assertion", *Philosophical Review* 74, pp. 449-465.
- GEACH, P. & BLACK, M. (eds.). 1970. *Frege Translations*, 2nd ed. Basil Blackwell, Oxford.
- GIBBARD, A. & HARPER, W. 1978. "Counterfactuals and two kinds of Epistemic Utility", in Gaerdenfors, P. and Sahlin, N-E, (eds.). pp. 341-77.
- GORDON, R. 1969. "Emotions and Knowledge", *The Journal of Philosophy* 66, pp. 408-413.
- GRANDY, R. 1973. "Reference, Meaning, and Belief", *The Journal of Philosophy* 70, pp. 439-452.
- GRIFFITHS, A. P. (ed.) 1991. *A.J. Ayer Memorial Essays*. The Royal Institute of Philosophy 1991. Cambridge: Cambridge University Press.
- HAACK, R., 1971. "On Davidson's Paratactic Theory in Oblique Contexts". *Nous* 5, N71, 351-361.
- HARMAN, G. 1973. *Thought*. Princeton. New Jersey: Princeton University Press.

- HOLTZMAN, S and LEIGH, C. (eds.). 1981. *Wittgenstein: To Follow a Rule* Routledge & Kegan Paul. London.
- HOOK, S. (ed.). 1960. *Dimensions of Mind*, New York.
- HOPKINS, J. and WOLLHEIM, R. (eds.). 1982. *Philosophical Essays on Freud*. Cambridge: Cambridge University Press.
- HUMBERSTONE, L. 1992. "Direction of Fit". *Mind* 101.pp. 59-83.
- HURLEY, S. 1989. *Natural Reasons*. Oxford: oxford University Press.
- JACOBS, R. A. and ROSENBAUM, P.S. (eds.). 1970. *Readings in Transformational Grammar*. Ginn and Company.
- JEFFREY, R. C. 1965. *The logic of decision*. McGraw-Hill Book Company.
- JEFFREY, R. 1976. "Savage's Omelet". In *PSA 1976*, Vol. 2. pp. 361-371.
- JEFFREY, R. 1983. *The Logic of Decision*. 2nd ed. Chicago IL. University of Chicago Press.
- KOOPMAN, B. O. 1940. "The axioms and algebra of intuitive probability", *Annals of Mathematics*, 41, pp. 269-292.
- KRANTZ, D., LUCE, D., SUPPES, P., TVERSKY, A., 1971. *Foundations of Measurement*. Vol.1. Academic Press. New York
- KYBURG, H. and SMOKLER, H. (eds.). 1964. *Studies in subjective probability*. New York, London, Sydney: John Wiley & Sons.
- LAKOFF, G. 1972. "Linguistics and Natural Logic", in Davidson, D., and Harman, G. (eds.) 1972.
- LEPORE, E. and VAN GULICK, R. 1991. *John Searle and his critics*. Oxford: Blackwell.
- LEWIS, D. 1966. "An Argument for the Identity Theory" in Lewis 1983.
- LEWIS, D. 1970. "How to Define Theoretical Terms", *Journal of Philosophy* 67, no.13, pp. 427-444.
- LEWIS, D. 1970. "General Semantics" in *Philosophical Papers* Vol.1, New York Oxford, Oxford University Press, 1983, pp.189-229.
- LEWIS, D. 1972. "Psychophysical and Theoretical Identifications", *Australian*

Journal of Philosophy, 50. pp. 249-258.

LEWIS, D. 1974. "Radical Interpretation" in D. Lewis 1983, pp. 108-121.

LEWIS, D. 1978. "Mad Pain and Martian Pain" in Lewis 1983, pp.122-132.

LEWIS, D. 1980. "A Subjectivist's Guide to Objective Chance" in Lewis 1986 pp. 83-132.

LEWIS, D. 1981. "Causal Decision Theory". *Australasian Journal of Philosophy*, 59. pp. 1-30.

LEWIS, D. 1983. *Philosophical Papers* Vol.1, New York, Oxford: Oxford University Press.

LEWIS, D. 1986. *Philosophical Papers* Vol. 2, New York, Oxford: Oxford University Press.

LEVI, I. 1966. "Review of Jeffrey's Logic and Decision" *Synthese* 16, pp. 239-244.

LEVI, I. 1991. "Consequentialism and Sequential Choice", in Baccharach M., and Hurley S. (eds.). 1991.

LOAR, B., 1981. *Mind and Meaning*. Cambridge: Cambridge University Press.

LUCE, R.D., 1968. "On the numerical representation of qualitative conditional probability", *Annals of Mathematical Statistics*, 39, pp. 481-491.

McDOWELL, J. 1977. "On the Sense and Reference of a Proper Name", *Mind*, 86. pp. 159-185

MELLOR, D. H. 1971. *The Matter of Chance*. Cambridge: Cambridge University Press.

MELLOR, D. H. (ed.), 1990. *F. P. Ramsey, Philosophical Papers*. Cambridge: Cambridge University Press.

MILLER, S. 1991. "Davidson's paratactic analysis of mood", *Journal of Pragmatics* 15, pp. 1-10.

NOZICK, R. 1969. "Newcomb's problem and two principles of choice", in Recher N. (ed) 1969.

NOZICK, R. 1993. *The Nature of Rationality*. Princeton: Princeton University Press.

- PEACOCKE, C. 1986. *Thoughts: An Essay on Content*. Aristotelian Society Series. Volume 4. Basil Blackwell. Oxford.
- PLATTS, M. 1980. *Reference, Truth and Reality*. Routledge & Kegan Paul. London, Boston and Henley.
- POPPER, K. 1990. *A World of Propensities*. Bristol: Thoemes.
- PRICE, H. 1983. " 'Could a Question Be True?': Assent and the Basis of Meaning", *The Philosophical Quarterly*, vol. 33. pp. 354-364.
- PUTNAM, H., 1960. "Minds and Machines", in *Dimensions of Mind*, Sidney Hook, (ed.). 1960, pp. 148-179.
- PUTNAM, H., 1975. *Mind, Language and Reality*. Cambridge. Cambridge University Press.
- QUINE, W.V.O., 1960. *Word and Object*, The M.I.T. Press, Cambridge, Massachussets.
- RAMBERGH, B., 1989. *Donald Davidson's Philosophy of Language*. Basil Blackwell.
- RAMSEY, F. 1926. "Truth and Probability" in D.H. Mellor ed. 1990, pp. 52-109.
- REICHENBACH, H. 1949. *The Theory of Probability: an Inquiry into the Logical and Mathematical Foundation of the Calculus of Probability*. 2nd edition. Berkeley: University of California Press.
- RESCHER, N. (ed.). 1969. *Essays in Honour of Carl Hempel*. Dordrecht: Reidel.
- ROSS, J. R. 1970. "On Declarative Sentences" in *Readings in English Transformational Grammar*, Jacobs, R. A. and Rosenbaum, P. S. (eds.) 1970.
- SAVAGE, L. J. 1954. *The Foundations of Statistics*. John Wiley & Sons, inc., New York. Chapman & Hall, Limited, London.
- SCHICK, F. 1986. "Dutch Bookies and Money Pumps". *The Journal of Philosophy*. Vol. 83. pp. 112-119.
- SCHIFFER, S. 1987. *Remnants of Meaning*. A Bradford Book, The MIT Press, Cambridge Massachusetts, London, England.

SCOTT, D. and SUPPES, P. 1958. "Foundational Aspects of Theories of Measurement", *The Journal of Symbolic Logic*, Vol. 23, pp. 113-128.

SEARLE, J. 1969. *Speech Acts*. Cambridge: Cambridge University Press.

SEARLE, J. 1977. "Locutions and Illocutions". *Phil. Review*, 1977, p.405-424.

SEARLE, J. 1983. *Intentionality*. Cambridge University Press.

SEARLE, J, and VANDERVEKEN, D. 1985. *Foundations of Illocutionary Logic*. Cambridge University Press, Cambridge.

SKYRMS, B. 1982. "Causal Decision Theory". *Journal of Philosophy* 79. pp. 695-711.

SKYRMS, B. 1990. "Ratifiability and the Logic of Decision". *Midwest Studies in Philosophy*. Vol. XV. The University of Notre Dame Press.

SPOHN, W. 1977. "Where Luce and Kranz do really generalize Savage's decision model". *Erkenntnis* 11, pp. 113-34.

STALNAKER, R. 1984. *Inquiry*. Cambridge, Mass. London: MIT Press, A Bradford Book.

STENIUS, E. 1967. "Moods and Language-Game". *Synthese* 17, pp. 254-274.

STICH, S. and WARFIELD, T. (eds.). 1994. *Mental Representation, a reader*. Oxford, Blackwell.

STOCKER, M. 1996. *Valuing Emotions*. Cambridge. Cambridge University Press.

TARSKI, A. 1956. "Concept of Truth in Formalised Languages", in *Logic Semantics and Metamathematics*. Hackett Publishing Company, 1983.

VAN STRAATEN, Z. 1980. *Philosophical Subjects: Essays Presented to P.Strawson*. Oxford University Press.

VELLEMAN, D. 1998. "How Belief Aims at the Truth". Unpublished Paper.

VON MISES, R. 1946. *Notes on Mathematical Theory of Probability and Statistics*. Cambridge: Massachusets.

VON NEUMANN, J & MORGENSTERN, O. 1947. *Theory of games and economic behaviour*, Princeton, Princeton University Press.

WIGGINS, D. 1980. "What would be a Substantial Theory of Truth?", in *Philosophical Subjects: Essays Presented to P. Strawson* Z., Van Straaten, (ed.), pp.189-221.