Perspectives and Derived Extensions of Dialogue Acts

Anton Benz*

Abstract

Wir betrachten die Rolle, die die *Perspektiven* der Dialogteilnehmer bei der Erweiterung des Gebrauchs von Dialog–Akten spielen. Wir konzentrieren uns auf die Beispiele des referentiellen Gebrauchs von Kennzeichnungen und dem von Deklarativsätzen. Wenn wir eine Klasse von Dialogsituationen bestimmt haben, die für den jeweiligen Gebrauch prototypisch ist, dann erlaubt es die beschränkte Information — oder Desinformation — der Teilnehmer den Gebrauch systematisch auf neue Situationen zu übertragen, wobei der erweiterte Gebrauch allerdings zu anderen Interpretationen führen kann.

We consider the role of *perspectives* of dialogue participants for the extension of dialogue acts to new situations. We concentrate on the examples of the referential use of definite descriptions, and on the utterance of sentences in declarative mood. If we have characterised a class of prototypical situations where such a specific act can be performed, then the limited information — or misinformation — of the participants allow to derive systematically new situations where the same act can be performed, but might be interpreted differently.

1 Introduction

We investigate how the different perspectives of dialogue participants give rise to derived uses of already given dialogue acts. We understand by *perspective of a participant* the information the participants has about the dialogue situation, including e.g. the situation talked about and the beliefs of other participants. Let's consider the following prototypical example of an assertion *I hold an ace.*

(1) S and H sit at a table playing with cards. A notices that he has an ace, and says: "I hold an ace."

The assertion is, of course, a true assertion. Now, consider the slightly different situation:

(2) Assume now that S has no ace but mistakes a joker for an ace. Again he says: "I hold an ace."

*This work is part of the DFG–Project LA 633/5–1 on *Dialogue Semantics* located at the Humboldt–Universität Berlin, http://www2.rz.hu-berlin.de/asg/blutner/dialog/index.html. I want to thank the organisers of the *Communicating Agents* workshop for the opportunity to discuss and publish these results.

This is still an assertion but no more a true assertion. S is justified to make this assertion because *out of his perspective* it is still a true assertion.

(3) Now assume S knows that he does not hold an ace but knows that H can't know this. He says: "I hold an ace."

Now, it is not only a false assertion but a lie. S must believe that *out of H's perspective* it is still a true assertion.

(4) Assume that in situations (2) and (3) H can see through a mirror the playing cards of S.

In this case, H can recognise that S was lying to her. It is now important for us to notice that she can only understand S's utterance as a lie because she knows that he must believe that she can't know that it is one. In the following example it is common knowledge that *I hold an ace* is obviously not true:

(5) Both players have only one card. A puts his card face up on the table. It is a king. Hence, both can clearly recognise that they mutually must know that it is a king. Then A says: "I hold an ace."

Without special context, the sense of the utterance of S must be unintelligible for H. It can't be a false assertion, nor a lie. This shows that the knowledge of the participants about their beliefs and their mutual knowledge plays a central role in the interpretation of sentences.

A similar phenomenon can be found in the case of the referential use of definite descriptions. In (Benz, 1999) we could show how the systematic exploitation of perspectives provides extended referential uses and interpretations of definites for defective states, i.e. states where prima facie the conditions for a successful use are violated. We started with a felicity condition for *basic* dialogue situations, i.e. situations where both participants have only true beliefs, and where this is mutual knowledge:

(6) There are two playing cards c_1 and c_2 lying face down, side by side on a table. A and B can see them both and that they mutually see them. Then a supervisor turns the first card, c_1, around, so that both can see that it is an ace. And this will be, of course, common knowledge. Now, A says to B: "Please, point to *the ace*."

Then we extended this use and the interpretation to defective situations by use of four principles connected to the following types of situations:

- The speaker can believe that the felicity condition holds.
- The hearer can believe that the felicity condition might hold.
- The speaker can believe that the hearer believes that the felicity condition might hold.
- The hearer can believe that the speaker might believe that the felicity condition holds.

We could explain why in the following example the use of the definite description *the man with the red walking stick* is successful:

(7) There is a couple, A and B, sitting in the park. In some distance there are two men walking. One of them has a red umbrella. A thinks that he can see that it is a red walking stick. He believes that B would not be able to say what exactly the man carries with him, because she is somewhat short-sighted. Then he remembers that he knows the man.

A: Look there, *the man with the red walking stick.* Yesterday I had a game of chess with him.

B: Oh, really. I know him too. We talked together just before you met me. I saw that he does not have a walking stick but a very slim red umbrella.

Now, we can show that the same mechanism can be applied to assertions. This gives us an outline of a theory for how in general perspectives give rise to derived extensions of already given dialogue acts. This mechanism is especially interesting in dialogue theory as it allows us to explain the felicity and effects of uses in defective situations.

In Section 2 we will consider related approaches which try to explain the relation between the different uses of declarative sentences as in the examples **(1)**–**(4)** within a theory of rational behaviour. In Section 4 we outline the basic ideas of our approach, and define four operations which allow us to derive extended uses of dialogue acts. Then, in Section 5, we apply our theory to examples. In an appendix we provide a short overview of the basics of the so-called possibility approach, which we adopt for our representations of dialogue situations.

2 Rationally Based Speech Act Theories with Default Rules

The problems around the dependency of closely related speech acts, e.g. *assertions, lies, unsuccessful lies etc.*, have been discussed e.g. in the papers (Cohen & Levesque, 1985, 1990a,b; Perrault, 1990). Their principal idea is as follows: They postulate a correlation between the uttering of a sentence with a certain syntactic feature (the locutionary act, e.g. imperative, declarative) in a certain context (specified by *gating conditions*), and a complex propositional attitude expressing the speaker's mental state. The speaker's uttering of a sentence under those gating conditions results in the hearer's beliefs that the speaker has the corresponding attitudes. Then, general principles governing mental states allow to derive other consequences of the speaker's having the expressed state.

> If *gating–condition* holds and locutionary act X is performed, then *consequence–condition* holds.

The general axioms which allow to derive further consequences describe general properties of co-operative agents (like *sincerity*, *helpfulness*), and of some propositional attitudes (*believe*, *intend*, *mutual belief*).

C. Raymond Perrault (Perrault, 1990) presents as a point of departure for his theory a possible version of such an approach[1]. It characterises speech acts by axioms of the form described above. E.g. for sentences uttered in *declarative*

[1] He attributes this version to Cohen & Levesque. But there are significant simplifications in this picture, at least according to our reading of (Cohen & Levesque, 1990a,b).

mood with propositional content p there is an axiom which postulates that the following consequence condition will hold:

$$BMB_{H,S}G_SB_HG_SB_HB_Sp,$$

i.e. it will be the case that it is mutual belief ($BMB_{H,S}$) between hearer H and speaker S that the speaker has the goal (G_S) that the hearer believes (B_H) that the speaker wants that the hearer believes that the speaker believes that p. All stronger consequences, e.g. the standard case that the hearer believes after he has heard the utterance that the speaker believes p, or that he believes that p really holds, have to be derived by additional axioms characterising special circumstances. Of course, this condition is weak enough to cover all cases where a declarative sentence is used in a proper assertion, a lie, or the case where the hearer recognises the lie. As Perrault points out, such a theory can't handle *ironic* uses of declaratives, e.g. in case somebody says *This is the best meal I ever had* where it is obvious for speaker and hearer that the meal tasted quite bad. We will see that we can find even non–ironic uses of declaratives where this axiom is not weak enough.

Hence, the overall strategy of the approach is to formulate an axiom with very weak consequence condition, the strongest which capture *all* uses of a sentence with a certain syntactic feature. This is reflected in the fact that the *gating* condition for such axioms are very weak. In the above example they state no more than that S is the speaker in an utterance–event with hearer H and a sentence with content p. Special constraints, like *sincerity, competence, helpfulness*, have to be added to derive stronger consequences. But such a strategy meets the problem that it is difficult to find a consequent condition which is really weak enough to capture *all* uses of a certain class of sentences. We will adopt the converse strategy, i.e. we start with very strong conditions on the utterance situations with strong consequence conditions. They are intended to describe the most prototypical, or *basic*, cases for the use of a sentence with the relevant syntactic feature. Then we apply operations which reflect the influence of perspectives of participants on the dialogue situation to derive extensions. We can iterate this process, and in this way the conditions on the utterance situation and consequent situation would become weaker and weaker.

The approaches by Cohen & Levesque and Perrault handle the problem by *default mechanisms.* Perrault explicitly develops his theory as an application of default logic (Perrault, 1990). Cohen & Levesque use a classical monotonic framework but add an axiom (Cohen & Levesque, 1990a, Def. 5, p. 236), (Cohen & Levesque, 1990b, Def. 6) which works as a kind of default axiom. Hence, they too have a mechanism which allows to start with strong conditions. E.g. in Perrault's system we can prove the following default rules.

$$\text{(DR1)} \quad \frac{B_{H,t}B_{S,t}p : B_{H,t}p}{B_{H,t}p} \qquad\qquad \text{(DR2)} \quad \frac{B_{H,t+1}DO_{S,t}(p.) : B_{H,t+1}B_{S,t}p}{B_{H,t+1}B_{S,t}p}$$

(DR1) says: If the hearer believes at time t that the speaker believes at time t that p, and if it is not inconsistent to assume that H believes p, then he will believe that p. (DR2) says: If the hearer believes at time $t+1$ that the speaker uttered a sentence with propositional content p at time t, and if it is not inconsistent that he believes at this time that the speaker believed at time t that p, then H will believe at $t+1$ that the speaker really believed p. If the hearer knows that the speaker does not believe in p, then the default rules don't apply, and we arrive only on weaker consequence conditions. There are also default rules for intentions.

We can't go into a detailed discussion of the approaches. Cohen & Levesque and Perrault place their theories into a general theory of rational interaction. Especially, they explicitly describe the role of *intentions*. Of course, a full theory needs to deal with intentions and updates, but this needs more space than is available in this paper. So we can treat them only on an informal level. We claim that the fact that actions are performed relative to the perspectives of participants offer a real explanation for why extended uses of dialogue acts are possible. Hence, the theory of perspectives which we present in this paper can be seen as an empirical justification for such systems of default mechanisms.

3 The General Framework

We think of a dialogue situation as an example for what is known as a *multi–agent system*. In the following we take a dialogue situation to contain at a certain point of time a number of *dialogue participants* and an *outer situation*. We confine our considerations to the case where there are only two participants S and H. The outer situation may contain information about the immediate environment but most importantly it provides information about a situation talked about. The performance of a dialogue act like an assertion, question, use of a definite description etc. leads to changes in the state which describes the situation. An assertion e.g. aims at a change of the information state of the hearer. A question ultimately aims at a change of the information state of the speaker. We introduce this framework because it provides us with a conceptual basis for our theory of perspectives and dialogue acts.

In fact, we will not be interested in the actual representations an agent uses in his local state but only in the information which this representations contain about the state of the environment. Hence, we may identify the local states of the participants with the set of all states of the environment which are correlated to his actual representations, i.e. with the set of all *outer* situations which belong to a global dialogue situation where his local state is identical with the actual one. This set represents the *knowledge* of X about the world relative the overall system. This construction allows us to use a classical *possible worlds* approach to model the *information* of participants. Of course, this model would only contain the information the agent has about the environment. It is clear, that we also need to know what a dialogue participant knows about the knowledge of the other participant. Therefore, we represent dialogue situations as *possibilities* $w = \langle s_w, w(S), w(H)\rangle$, where s_w is the outer situation, and $w(S)$ and $w(H)$ are again sets of possibilities, so–called *information states*. This construction seems at first sight not to be well defined. In fact, it can't be defined in classical set theory. But it can be developed in (AFA) set theory (Aczel, 1988). The approach is known as *possibility* approach (Gerbrandy & Groeneveld, 1997). We will present our results in this framework. We introduce it in more detail in an appendix.

We assume that there are some minimal restrictions on the class of possibilities. First, that all information states of the participants are non–contradictory, that *full introspection* holds, i.e. that all participants know what they believe, and that this is common knowledge. We denote the class of all these possibilities by $\dot{\mathcal{I}}$. Hence, if $w = \langle s_w, w(S), w(H)\rangle \in \dot{\mathcal{I}}$, then for all participants X

$$(1)\ w(X) \neq \emptyset,\ (2)\ \forall v \in w(X)\, v(X) = w(X),\ \text{and}\ (3)\ w(X) \subseteq \dot{\mathcal{I}}.$$

Then, our basic dialogue situations will always be situations where both dialogue participants have only *true* beliefs, and where this is common knowledge. We denote the class of all such situations by $\mathcal{T}$. Hence, if $w \in \mathcal{T} \subseteq \dot{\mathcal{I}}$, then

$$(1)\ w \in w(S) \cap w(H),\ \text{and}\ (2)\ w(S) \cup w(H) \subseteq \mathcal{T}.$$

The intended applications include speech acts like assertions, lies and questions, but also examples like the (referential) use of a definite description. Therefore, the notion of *dialogue act* has a wider meaning to us than the concept of *speech act*. Roughly, we can characterise our use of *dialogue act* as applying to any utterance of certain type of phrase which aims at a change in the information states of the participants. All considered examples of dialogue acts have in common that they depend on and affect only the information states of speaker and hearer. If the question whether a dialogue act can be performed or not is dependent on facts related to the outer situation, then it is clear that we can not expect that an extended use should be possible just because speaker and hearer *believe* it to be possible.

4 Dialogue Acts and Perspectives

4.1 The Basic Consideration

Let us look again to the example of the assertion *I hold an ace*. We abbreviate it by ψ. As a statement it would be felicitous just in case the sentence ψ is true. But, of course, the speaker S can relay for what he says only on what he believes to be true.

More generally, let $M \subseteq \dot{\mathcal{I}}$ be any class which represents a property of possibilities, e.g. a class where a dialogue act can be performed successfully. We explicate the fact that this property obtains *under the perspective* of a dialogue participant X in world w as $w(X) \subseteq M$. This means that *all* of X's epistemic possibilities are elements of M.

For assertions this means that $w(S) \models \psi$ must hold, if the speaker should be justified to make them. If M specifies a class where some definite description $d = \operatorname{def} x.\varphi(x)$ can be used felicitously, then it means that speaker S seems to be justified to utter d if and only if all his epistemic possibilities are elements of M.

Now, we turn attention to the hearer H. If he hears an utterance, is it necessary for him that *all* his epistemic alternatives are elements of M in order to make sense of what he has heared?

(8) There are two playing cards on a table. The left one is an ace the right one a queen. S and H can see it and each other see it. Then they leave the room. An hour later they come back, the cards still there but face down. H has forgotten whether the first card is an ace or a king. He still knows that the second is a queen. S, who hasn't noticed this, says to H: "Give me the ace, then we leave and I invite you for a coffee."

Here, H can guess that S means the first card. The possibility that there is a king and a queen is ruled out by S's use of "the ace." Therefore, we arrive at a weaker condition on the perspective of H, namely that there should be at least one possibility v in his set of epistemic alternatives $w(H)$ that is in M.

We can find the same phenomenon in the case of assertions. If H is convinced that the speaker does not hold an ace, then he would not accept S' assertion *I hold an ace.* Again, we have as a minimal requirement for success that there is at least one epistemic alternative in H's set of possibilities where the uttered sentence is true. Moreover, it is essential for assertions that the addressee does not know that the uttered sentence is true, i.e. that there is at least one epistemic alternative where ψ is false.

If M specifies a property of dialogue situations, then the actual situation w *may* have this property *under the perspective of* H iff $w(H) \cap M \neq \emptyset$. And, if M specifies the felicity conditions of some dialogue act, then the respective act can be understood as *reasonable* by H just in case he thinks that the actual situation might be an element of M.

We define two operators on subclasses of possibilities for each dialogue participant X. They are closely related to the modal operators $\Box_X$ and $\Diamond_X$, so that we denote them by the same symbols:

$$\begin{aligned} \Box_X M &:= \{w \in \dot{\mathcal{I}} \mid w(X) \subseteq M\} \\ \Diamond_X M &:= \{w \in \dot{\mathcal{I}} \mid w(X) \cap M \neq \emptyset\} \end{aligned}$$

With these operators at hand we can reformulate our observations as: S is convinced that the actual world w belongs to M iff $w \in \Box_S M$; H can accept that w belongs to M iff $w \in \Diamond_H M$.

4.2 The Derivation of Dialogue Acts

We first show that there are four operations due to the perspective of one participant which extend the classes where some dialogue act can be performed. As an example we use again the assertion

(9) S**:** I hold an ace. (ψ)

Assume that S is convinced of the truth of ψ. Even if it is in fact false, he will think to be justified to make the assertion. If, furthermore, the hearer H trusts into the beliefs of S, they should both accept this assertion as if S had told the truth. Then, S can use the assertion ψ to mislead the hearer H, if H is in a situation where he accepts the utterance. This is, of course, no more an assertion but a lie. We can see here, how the limited perspective of one dialogue participant can give rise to an extension of a dialogue act. To explain the precise connection of lies and assertions it is, of course, necessary to introduce *goals* of participants into the model. In this paper, we concentrate only on the restrictions for extensions of dialogue acts which are due to epistemic perspectives. The acceptability of the utterance for S and H forms a necessary condition for a derived use. Hence, the class of situations where this condition holds forms a class of *candidates* for a possible extension. Additional constraints may enter to determine the real extensions.

Hence, if M is a class of possibilities where some dialogue act can be performed, then this act might be extended to the class where S is convinced that

H accepts the act. It is the class $\Box_S(M \cup \Diamond_H M)$. If $M \subseteq \Diamond_H M$, the definition of the extension can be simplified to $\Box_S \Diamond_H M$.

Now we can again turn to the perspective of H. Assume that the speaker S is convinced of the truth of ψ but H knows it to be false. H can make sense of the utterance, if it is possible for him that S might believe ψ. *Make sense* means here: he can understand it as an attempt of an assertion. Let us look again to a situation where this is *not* possible.

(10) Both players have only one card. A puts his card face up on the table. It is a king. Hence, both can clearly recognise that they mutually must know that it is a king. Then A says: "I hold an ace."

Probably, H will be quite puzzled about this utterance. It seems to be impossible to make sense out of it. In contrast to the following scenario:

(11) Now assume A knows that he does not hold an ace but believes that B can't know this. He says: "I hold an ace." Assume that in this situation B could see through a mirror the playing cards of A.

Of course, A's utterance was a lie, and H can not accept it as a true assertion. But he can make sense of it as he can recognise it as a lie. He can then react with a rejection, or he might *accept* it as assertion, thereby misleading the speaker.

If we look to referential uses of definite descriptions, we can find quite clear examples for this reasoning of the hearer.

(12) There are two playing cards on a table. They lay face down side by side. S gets told that the left one is an ace and the right one a queen. In fact, they are both joker. Then H is brought to the table. He has seen the cards before, so he knows that they are joker. A supervisor tells them that he has just informed S that the left card is an ace. Then S says to H: "Give me the ace!"

Of course, H should be able to identify the card, which S wants to refer to with *the ace*, although he knows that there are no aces. But he knows that out of the perspective of S there is one card in the common ground which is an ace, and that there is only one.

So if M is a class of possibilities where some dialogue act can be performed, then this act might be extended to the class where H thinks that it is possible that S is convinced that he can perform this act. We get as extension the class $\Diamond_H \Box_S M$.

We find in this way four operations which give us new classes where some dialogue act can be performed due to the perspective of one participant. Let M be given. Then we arrive at the following classification:

	direct	indirect
speaker	$\Box_S M$	$\Box_S \Diamond_H M$
hearer	$\Diamond_H M$	$\Diamond_H \Box_S M$

We have to mention that the simple form $\Box_S \Diamond_H M$ for the indirect operation for the speaker is sufficient only, if we can show that $M \subseteq \Diamond_H M$. Else, it should have the form $\Box_S(M \cup \Diamond_H M)$! If M characterises a dialogue act where $M \subseteq \mathcal{T}$, then we trivially have $M \subseteq \Diamond_H M$. In general, we can't expect this.

To get a real possible extension it is necessary that S is convinced that he can perform the act, and H must be able to make sense of this. It is not enough, if only one participant thinks that the act can be performed. Therefore, we have to build intersections of the derived classes. We get the following four groups:

	direct H	indirect H
direct S	$\Box_S M \cap \Diamond_H M$	$\Box_S M \cap \Diamond_H \Box_S M$
indirect S	$\Box_S \Diamond_H M \cap \Diamond_H M$	$\Box_S \Diamond_H M \cap \Diamond_H \Box_S M$

5 Applications

In this section we apply our theory to two examples: assertions and the referential use of definite descriptions. In the latter case we can refer to (Benz, 1999) for a more extended treatment. According to our strategy outlined in the last sections, we start in each case with a class of dialogue situations which characterises the most basic, or prototypical use. Then we show whether the examples belong to the basic or the derived new situations.

For the basic cases we assume that the participants mutually know that they have only *true* beliefs, i.e. for such a situation $w = \langle s_w, w(S), w(H) \rangle$ we assume that $w \in w(S) \cap w(H)$, and that the same holds for all $v \in w(S) \cup w(H)$. We introduced the class of all such situations in Section 3 as $\mathcal{T}$.

We assume that in the basic case where a participant S can utter a declarative sentence with propositional content ψ felicitously (1) ψ should in fact hold, (2) the speaker should be convinced of ψ, and (3) the hearer should not know whether ψ. If $[\psi] := \{w \in \mathcal{T} | w \models \psi\}$ denotes the class of possibilities in $\mathcal{T}$ where ψ holds, then the basic cases form the class $M := [\psi] \cap \Box_S[\psi] \cap \Diamond_H[\psi] \cap \Diamond_H[\neg\psi]$. In addition, we make an informal assumption about the intentions of the speaker. In the basic case he should always be sincere, i.e. if he says that ψ, he should really believe that ψ, and this should be mutually known. As already mentioned, we don't include this in the formal representation of a dialogue situation.

If we call an utterance of a sentence with content ψ *possible*, then we mean in the following that the utterance is *reasonable* for the speaker, and that the hearer can *make sense* of it. This was part of the general ideas motivating the operators introduced in the last section. Hence, we will call an utterance of ψ which is a lie, and which the hearer can recognise as a lie *possible* although it is unsuccessful, if we consider it as *speech act*.

We reconsider the examples given in the introduction. Hence, let ψ be the proposition "*S holds an ace.*" As we are only interested in this proposition, we allow only for *two* outer situations, one where ψ holds, and one where $\neg\psi$ holds. We can therefore denote dialogue situations by triples $\langle \psi/\neg\psi, w(S), w(H) \rangle$.

Example **(1)** describes the following situation w_1:

$$\begin{aligned} w_1 &= \langle \psi, \{w_1\}, \{w_1, v\} \rangle \\ v &= \langle \neg\psi, \{v\}, \{w_1, v\} \rangle . \end{aligned}$$

It is the situation where ψ holds, where the speaker knows all about the real situation, i.e. the only epistemically possible situation for him is w_1 itself, and where the hearer does not know whether ψ, but he knows that the speaker knows it, i.e. there are two possibilities for him, w_1 and another one which is identical to w_1 but where we replaced $\neg\psi$ for ψ. Of course, this interpretation is not fully justified by the way how the example was stated:

(13) A and B sit at a table playing with cards. A notices that he has an ace, and says: "I hold an ace."

There are a lot of dialogue situations where this can be part of the description. It seems to be quite natural to give it a very strong interpretation. But we don't want to explain how we arrive at such a strong reading, but only, given the reading, why the utterance of "I hold an ace" is felicitous.

w_1 is clearly an element of $[\psi]$, i.e. ψ holds and $w_1 \in \mathcal{T}$. As $w_1 \models \psi$ it follows that $w_1 \in \Box_S[\psi]$, and as $v \models \neg\psi$ it follows that $w_1 \in \Diamond_H[\neg\psi]$. This proves that $w_1 \in M = [\psi] \cap \Box_S[\psi] \cap \Diamond_H[\psi] \cap \Diamond_H[\neg\psi]$, and that therefore the utterance of "I hold an ace" is felicitous.

The situation in Example **(2)** is described by

$$w_2 \quad = \quad \langle \neg\psi, \{w_1\}, \{w_1, v\} \rangle\,,$$

i.e. both participants have the same convictions as in the first example but this time ψ does not hold. Hence, $w_1 \neq w_2 \notin w_2(S)$, therefore $w_2 \notin \mathcal{T}$. But $w_2 \in \Box_S M \cap \Diamond_H M =: M_1$. Hence, the utterance is possible although it is not true this time. It is not uninteresting to write out the definition of M_1. After some simplifications we arrive at $\Box_S([\psi] \cap \Diamond_H[\neg\psi]) \cap \Diamond_H([\psi] \cap \Box_S[\psi]) \cap \Diamond_H[\psi]$. We can see in the second conjunct that it is necessary that the hearer believes it to be possible that the speaker can know that ψ. It would not be enough, if there are just any two basic possibilities in his belief–state, one where ψ holds and one where $\neg\psi$ holds.

In Example **(3)** the beliefs of the hearer are the same as in Example **(2)**, but the speaker knows now that $\neg\psi$ holds.

$$w_3 \quad = \quad \langle \neg\psi, \{w_3\}, \{w_1, v\} \rangle\,.$$

This is an element of $\Box_S \Diamond_H M \cap \Diamond_H M =: M_2$. Hence, it can be derive from the basic case, and our theory predicts that it can be reasonable for the speaker to say that ψ. Of course, he must have the intention to mislead the hearer, i.e. his utterance must be a lie. More informally, his reasoning can be like this: *I know that I don't hold an ace. H does not know whether I hold it or not, but he knows that I know it, and believes that I am trustworthy. Hence, if I say to him that ψ, then he will believe me, and I want him to believe it.* But this reasoning makes essential use of the different perspectives.

In Example **(4)** the hearer recognised that the speaker believes to be in situation **(3)**. The new situation is an element of $M_3 := \Box_S \Diamond_H M \cap \Diamond_H \Box_S M_2 = \Box_S \Diamond_H M \cap \Diamond_H \Box_S (\Box_S \Diamond_H M \cap \Diamond_H M) = \Box_S \Diamond_H M \cap \Diamond_H \Box_S \Diamond_H M$.

$$w_4 \quad = \quad \langle \neg\psi, \{w_3\}, \{w_4\} \rangle\,.$$

The hearer can guess that the speaker reasoned in the way stated above, and can interpret his utterance as a lie, therefore, it makes sense for him. Again this is only possible because he can recognise that the utterance is reasonable out of the perspective of S.

We can see now how to construct further, more derived examples. Assume that the hearer believes himself to be in a situation as in Example **(4)** but the speaker knows this and has the intention to make the hearer believe that he is lying to him. Then it is reasonable for him to utter ψ. The underlying belief–states can be derived by a further application of the operator $\Box_S \Diamond_H$. In principle, we can repeat these constructions again and again, and get always new situations where an utterance of ψ is possible. We can find especially a situation where it is no more the case that it is mutually known that *the speaker has the goal (G_S) that the hearer believes that the speaker has the goal that the hearer believes that the speaker believes that* ψ, i.e. where it is no more mutually known that

$$(\Theta) \quad G_S \Box_H G_S \Box_H \Box_S \psi.$$

In this case it would be necessary that the speaker himself believes that Θ. But this means that he wants that $\Box_H G_S \Box_H \Box_S \psi$. If we iterate our construction three times, then we get

$$\begin{aligned} w_5 &= \langle \neg\psi, \{w_5\}, \{w_4\}\rangle \\ w_6 &= \langle \neg\psi, \{w_5\}, \{w_6\}\rangle \\ w_7 &= \langle \neg\psi, \{w_7\}, \{w_6\}\rangle, \end{aligned}$$

which can represent the information of S and H in a situation where the speaker wants that the hearer believes that S wants him to believe that he is lying to him. In this case, after the utterance, the hearer will of course not believe that the speaker wants him to believe that he believes ψ. Hence, this is a non–ironic use which is a counterexample to the claim that Θ is mutually known in *all* possible uses of a declarative sentence. Of course, this example is already quite artificial, but this has to be expected from a counterexample.

Until now we have only considered examples where the use of a declarative sentence was possible. We now want to examine some examples where the dialogue situation does not belong to one of the derived classes.

The most simple case is represented in Example **(5)**. It is represented by

$$v_1 = \langle \neg\psi, \{v_1\}, \{v_1\}\rangle .$$

This is, of course, an element of $\mathcal{T}$ but not of $[\psi]$, hence, not of M. It is provable that v_1 is not an element of any of the derived classes[2].

(14) H is brought into a room where a supervisor puts some cards on a table. He can see all of them before they are turned around. The first card is an ace. Then S, who waited outside where he couldn't see the cards, enters and says: "The first card is an ace."

If ψ denotes the proposition *the first card is an ace*, then we can represent the utterance situation by v_2:

$$\begin{aligned} v_2 &= \langle \psi, \{v_2, v_3, v_4, v_5\}, \{v_2\}\rangle, \\ v_3 &= \langle \neg\psi, \{v_2, v_3, v_4, v_5\}, \{v_3\}\rangle, \\ v_4 &= \langle \psi, \{v_2, v_3, v_4, v_5\}, \{v_4, v_5\}\rangle, \\ v_5 &= \langle \neg\psi, \{v_2, v_3, v_4, v_5\}, \{v_4, v_5\}\rangle, \end{aligned}$$

None of the possibilities v_2 to v_5 belongs to M because S is never convinced that ψ holds. Hence, his utterance can't be a basic use, and it is again provable that

[2]If we denote by $T(w)$ the *transitive hull* of $\{w\}$, i.e. the smallest set σ, such that $w \in \sigma$ and $\forall v \in \sigma\, v(S) \cup v(H) \subseteq \sigma$, then it holds that: If w is in one of the derived classes, then $T(w) \cap M \neq \emptyset$. We can see that $T(v_1) = \{v_1\}$. The same criterion helps also in the next examples.

the situation v_2 does not belong to any of the derived classes. Intuitively, it is not a reasonable sincere assertion as the speaker lacks the necessary knowledge, and H knows this, and it is even common knowledge that he lacks this information. Therefore, H may not only reject this utterance, it should also be difficult for him to make any sense of it.

Finally, we want to consider the referential use of a definite description. For the basic case we assume that a speaker can refer with $\operatorname{def} x.\varphi(x)$ to an object a iff it is common knowledge that $\varphi(a)$[3], and that a is the only object where this is commonly known. An example is **(6)**. We now reconsider Example **(8)**. There are two cards, c_1 and c_2. c_1 is an ace and lies to the left of c_2, a queen. The speaker wants to pick out c_1 by use of *the ace*. The utterance situation introduced in (**8**) can be described by w:

$$\begin{array}{rcl} w & = & \langle [A(c_1), Q(c_2)], \{u\}, \{w, v\} \rangle \\ v & = & \langle [K(c_1), Q(c_2)], \{u'\}, \{w, v\} \rangle \\ u & = & \langle [A(c_1), Q(c_2)], \{u\}, \{u\} \rangle \\ u' & = & \langle [K(c_1), Q(c_2)], \{u'\}, \{u'\} \rangle \end{array}$$

We denote the class of basic cases again by M. We see that only u is an element of M, and therefore we can show that $w \in \Box_S M \cap \Diamond_H M$. The theory predicts that the use of "the ace" should be successful.

6 Conclusion

I have argued that perspectives play an essential role in the derivation of extended uses of dialogue acts. The essential idea was to start with a class of basic, or prototypical situations, and then to extend the use from this basic class by application of four operators which reflect the four ways how the partiality of information can give rise to uses in situations which do not belong to the basic class.

- The speaker can believe that the performance is possible.
- The hearer can believe that the performance might be possible.
- The speaker can believe that the hearer believes that the performance might be possible.
- The hearer can believe that the speaker might believe that the performance is possible.

We concentrated on the referential use of a definite description, and especially the use of a sentence in declarative mood. In the last case, our theory could show how a speech act like lying can be derived from the act of asserting.

References

P. Aczel: *Non–Well–Founded Sets*; CSLI–Lecture Note 14, Stanford 1988

[3] We provide a precise definition of *common knowledge* in the appendix.

A. Baltag: *A Logic of Epistemic Actions*; Proceedings of the ESSLLI99 Workshop on Foundations and Applications of Collective Agent Based Systems, Utrecht 1999

A. Baltag, L. Moss, S. Solecki: *The Logic of Public Announcements, Common Knowledge and Private Suspicion*; to be published 1998

J. Barwise, L. Moss: *Vicious Circles*; Stanford, CSLI, 1996

A. Benz: *Perspectives and the Referential Use of Definite Descriptions in Dialogue*; ms, Berlin 1999

P.R. Cohen, H.J. Levesque: *Speech Acts and Rationality*; In *Proceedings of the Twenty–third Annual Meeting*, Association for Computational Linguistics, Chicago, IL 1985

P.R. Cohen, H.J. Levesque: *Rational Interaction as the Basis for Communication*; in: P. Cohen, J. Morgan, M. Pollack (ed.): *Intentions in Communication*, Cambridge, Massachusetts 1990

P.R. Cohen, H.J. Levesque: *Performatives in a Rationally Based Speech Act Theory*; In *Proceedings of the Twenty–eighth Annual Meeting*, Association for Computational Linguistics, Pittsburgh, Pennsylvania 1990

R. Fagin, J.Y. Halpern, Y. Moses, M.Y. Vardi: *Reasoning About Knowledge*; Cambridge, Massachusetts 1995

J. Gerbrandy: *Changing the Common Ground*; in: A. Benz & G. Jäger (ed.), *Proceedings of the Munich Workshop on Formal Semantics and Pragmatics of Dialogue MunDial'97*, pp.40–58 München 1997

J. Gerbrandy: *Bisimulations on Planet Kripke*; ILLC Dissertation Series, Institute for Logic, Language and Computation, Universiteit van Amsterdam 1998

J. Gerbrandy, W. Groeneveld: *Reasoning about information change*; Journal of Logic, Language and Information 1997

W. Groeneveld: *Logical Investigations into Dynamic Semantics*; PhD thesis, ILLC dissertation series 1995–18, University of Amsterdam 1995

J. Jaspars: *Calculi for Constructive Communication. A Study of the Dynamics of Partial Sets*; PhD thesis, ITK, Katholieke Universiteit Brabant 1994

C.R. Perrault: *An Application of Default Logic to Speech Act Theory*; in: P. Cohen, J. Morgan, M. Pollack (ed.): *Intentions in Communication*, Cambridge, Massachusetts 1990

A The Possibility Approach and the Representation of Beliefs of Dialogue Participants

The possibility approach is essentially a possible worlds approach, i.e. it identifies the beliefs of an individual with the set of all worlds which are *possible* according to those beliefs. We denote the set of participants by $\mathrm{DP} = \{S, H\}$, where S will denote the *speaker*, H the *hearer*. A *possibility* consists of a model for the *outer* situation, and *information states* for each participant, where those states are again sets of possibilities. The *outer* situation describes the *non–modal* part of the dialogue context. In case of e.g. assertions, this outer situation will be identified with the situation *talked about*. The possibility approach was first developed by *J. Gerbrandy* and *W. Groeneveld* in (Gerbrandy & Groeneveld, 1997). It makes use of an extension of classical set theory, the theory of *Non–Well–Founded Sets* developed by *P. Aczel* (Aczel, 1988)[4]. The original problem motivating the development of the possibility approach was to define suitable *update operations* for dialogue. The approach proved here to be especially useful[5]. For a proper understanding of the details the reader may need to have some familiarity with the underlying set theory. We hope that he can get an intuitive understanding without it.

Let $\mathcal{S}$ be a class of models for the possible outer situations. For simplicity we assume that all models in $\mathcal{S}$ have the same set of individuals. We define *possibilities* and *information states* in the following way:

- A *possibility* w is a triple $\langle s_w, w(S), w(H)\rangle$ where $s_w \in \mathcal{S}$ and $w(S)$ and $w(H)$ are information states.
- An *information state* σ is a set of possibilities.

s_w describes an outer situation, $w(S)$ and $w(H)$ the set of worlds S and H believe to be possible. We denote the class of all possibilities with $\mathcal{W}$. The theory of non–well–founded sets allows for sets containing themselves, so it is possible that there exist possibilities w with $w \in w(X)$, $X \in \mathrm{DP}$.

The definition may seem to be circular, and therefore ill–defined. It is, of course, not a recursive definition. For an explanation we need some machinery from (AFA)–set theory. Generally, we can define the class $\mathcal{W}$ as the *largest fixed point* of a set–continuous operator Φ with $\Phi(x) := \{\langle s, i, j\rangle \mid s \in \mathcal{S}, i, j \subseteq x\}$. It is provable that this fixed point is also the class of all *solutions* for a certain class of *systems of equations*. The definition above directly translates into a definition of this *certain* class: We assume that there are two classes of *urelements*, P and I. A system of equations over P and I belongs to the desired class if it contains only equations of the form:

- $w = \langle s_w, i, j\rangle$ for $w \in P$, where s_w belongs to $\mathcal{S}$ and i, j to I.
- $i = \sigma$ for $i \in I$, where σ is a subset of P.

[4]For more information about (AFA) Set Theory we can also refer to *Barwise* & *Moss* (Barwise & Moss, 1996). For a more thorough discussion of the possibility approach we can refer to the thesis of Gerbrandy (Gerbrandy, 1998).

[5]There is some literature concerning the proper definition of updates in dialogue: (Jaspars, 1994), (Groeneveld, 1995), (Gerbrandy & Groeneveld, 1997), (Gerbrandy, 1998), (Baltag, Moss, Solecki, 1998), (Baltag, 1999).

In this way, it is really a proper definition[6].

We introduce a formal Language $\mathcal{L}^M$. Let $\mathcal{L}$ be a language of predicate logic for the class $\mathcal{S}$. We assume that $\mathcal{L}$ contains all the predicates the dialogue participants can use to talk about an outer situation. Then $\mathcal{L}^M$ should be the smallest language containing $\mathcal{L}$ and the following sentences for $\varphi, \psi \in \mathcal{L}^M$ and $X \in DP$:

$$\neg\varphi, \varphi \wedge \psi, \Box_X \varphi, \Diamond_X \varphi, E\varphi, C\varphi$$

Let $w = \langle s_w, w(S), w(H) \rangle$ be a possibility. We define truth conditions for $\varphi, \psi \in \mathcal{L}^M$:

1. $w \models \varphi$ *iff* $s_w \models \varphi$, φ a sentence in $\mathcal{L}$.
2. $w \models \neg\varphi$ *iff* $w \not\models \varphi$.
3. $w \models \varphi \wedge \psi$ *iff* $w \models \varphi \;\&\; w \models \psi$.
4. $w \models \Box_X \varphi$ *iff* $\forall w' \in w(X)(w' \models \varphi)$.
5. $w \models \Diamond_X \varphi$ *iff* $\exists w' \in w(X)(w' \models \varphi)$.
6. $w \models E\varphi$ *iff* $w \models \Box_S \varphi \wedge \Box_H \varphi$. Let $E^0\varphi := E\varphi$, $E^{n+1}\varphi := E(E^n\varphi)$.
7. $w \models C\varphi$ *iff* $\forall n \in \mathbf{N}\; w \models E^n\varphi$.

For a dialogue participant X a possibility w is *epistemically possible* in v iff $w \in v(X)$. X *believes* that φ in w iff φ holds in all his epistemic alternatives in w, i.e. iff $w \models \Box_X \varphi$. $w \models E\varphi$ means that *everybody* believes φ in w. φ is *common belief* in w iff $w \models C\varphi$. For information states we define

$$\sigma \models \varphi \textit{ iff } \forall w \in \sigma\; w \models \varphi.$$

Until now, we did not restrict the properties of possibilities. A subclass $M \subseteq \mathcal{W}$ is called *transitive*, iff

$$\forall w \in M\; \forall X \in DP\; w(X) \subseteq M.$$

Let $\mathcal{I} \subseteq \mathcal{W}$ be the largest transitive subclass with

$$\forall w \in \mathcal{I}\; \forall X \in DP\; \forall v \in w(X):\; w(X) = v(X).$$

This property is called *introspectivity*. It means: (1) If a dialogue participants believes φ, then he knows that he believes it; (2) if he does not believe that φ, then he knows that he does not believe φ; and (3) it means that (1) and (2) are common knowledge. We will always assume that introspectivity holds.

Let $\mathcal{T} \subseteq \mathcal{I}$ be the largest transitive subclass with

$$\forall w \in \mathcal{T}\; \forall X \in \mathrm{DP}\; w \in w(X).$$

If $w \in \mathcal{T}$, then w is for both participants an element of their sets of epistemic alternatives. Hence, if a participant believes that φ, then φ must in fact hold. Therefore, $\mathcal{T}$ denotes the class of possibilities where (1) the dialogue participants can only have *true beliefs*, i.e. *knowledge*, and (2) where this fact is common

[6]For the set–theoretic machinery we must refer to (Barwise & Moss, 1996). A very readable account can be found in (Gerbrandy, 1998).

knowledge. The *Anti–Foundation–Axiom* (AFA) of the underlying set theory guaranties that $\mathcal{T}$ is not empty. We can easily see that the $S5$–Axioms hold for $\mathcal{T}$: If $w \in \mathcal{T}$, $X \in \mathrm{DP}$, we have (1) $w \models \Box_X \varphi \Rightarrow w \models \varphi$, (2) $w \models \Box_X \varphi \Rightarrow w \models \Box_X \Box_X \varphi$, and (3) $w \models \varphi \Rightarrow w \models \Box_X \Diamond_X \varphi$.

We are only interested in non–contradicting information states of participants. This means that the set containing all their epistemic possibilities should contain at least one element. Let $\dot{\mathcal{W}}$ denote the largest transitive subclass of $\mathcal{W}$ with

$$w \in \dot{\mathcal{W}} \Rightarrow w(S) \neq \emptyset \neq w(H)$$

If M is any class of possibilities, then we denote by $\dot{M}$ the intersection of M and $\dot{\mathcal{W}}$. Note that $\dot{\mathcal{T}} = \mathcal{T}$.