Perspectives and the Referential Use of Definite Descriptions in Dialogue

Anton Benz

Abstract

This is an investigation into the pragmatics of the referential use of definite descriptions. We examine situations where a speaker uses a description to induce a hearer to pick out a certain object from a set of mutually given objects, in order to state some proposition about this object. We ask for a criterion which provides the conditions for a successful use. It depends on the belief states of the participants. Here we allow for situations with arbitrarily complex information states. The idea is to extend the successful use from a class of *basic dialogue situations* to a wider range using principles reflecting the influence of the epistemic perspectives of the participants.

1 Introduction

This paper is an investigation into the *pragmatics* of definite descriptions. We concentrate on a special use of definites, the so–called *referential* use, and it's dependence on the epistemic perspectives of a speaker and a hearer. Most work has been done on the relation between the *semantic* content of a definite and the *pragmatic principles* which together establish the different uses. We don't want to contribute to these discussions. We just start with the observation that there is such a use.

Our question is: What is the criterion for a successful referential use of a definite description? Where we consider a use to be *referential*, if we are in a *dialogue situation* where the speaker wants to induce the hearer to pick out a certain object, and where he wants to communicate some sentence about this object. We examine the relation between this use and one aspect of dialogue contexts, namely the *beliefs of participants about a described situation, and their beliefs about each other*. Dialogue is of chief interest here due to the complex relations with respect to what one dialogue participant believes about what an other participant believes.

(1) 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.

Here, the hearer thinks that it is possible that the speaker believes that they mutually see one and only one man with a red walking stick. The speaker, who knows this, exploits that fact, and the reference is successful. But, there is of course no object for which they both believe that it carries a red walking stick. Basically, we adopt the view that a speaker can refer with a definite def $x.\varphi(x)$ to an object a iff it is in some way the *unique* object for which $\varphi(a)$ holds. For us it is a central question how this depends on the beliefs of the dialogue participants. Here, the relation to the *common ground* has been extensively studied (Clark & Marshall, 1981). The common ground is normally explicated as the set of all sentences φ for which the speaker believes that φ holds, the hearer believes that φ holds, the speaker believes that the hearer believes that φ holds, etc. Or, in short, for which it is *mutual* or *common* belief that φ holds. Clark & Marshall claimed that mutual knowledge of $\varphi(a)$ is necessary, if a description should refer successfully to an object a. But in the example above there seems to be, at least at first sight, no connection to the so defined common ground. The relevant fact, that there is a man carrying a red walking stick, is not mutually believed. Nevertheless, the intended reference is possible. We want to show that there must be a connection, and that this connection is established through the perspectives of the dialogue participants.

Our central idea is: There are *basic* dialogue situations where the referential use is primarily defined. We get extensions of this use to other dialogue situations by operations which reflect the influence of *perspectives* and *exploitation* of perspectives.

First we reconsider the examples of Clark & Marshall which show that common belief is in fact of crucial importance. In Section 3 we clarify which use and which contexts exactly we investigate. There we also present the formal framework. In Section 4 we develop our own approach. There we first motivate the introduction of perspectives and basic dialogue situations by a series of examples with more and more complex belief structures. We apply our theory in detail to a representative set of examples in the last section.

The idea that a dialogue act can be extended from basic uses to new contexts due to the perspectives of participants should be applicable not only to the referential use of a definite. The principles which govern the extension are stated in a general form, so that we can hope that they open a way to a general treatment of perspectives in pragmatics.

Such an investigation has been facilitated through the development of suitable structures for the representation of knowledge states. Basically, we will use a possible worlds approach but a version that is much easier to use for the purposes of dialogue description. It is known as the *possibility* approach, see especially (Gerbrandy & Groeneveld, 1997). Our results are presented in a variant of this framework.

2 The Relevance of Common Knowledge for Definite Descriptions

H.H. Clark & *C.R. Marshall*'s paper (Clark & Marshall, 1981) contains a thorough discussion of the relation between mutual knowledge and different uses of definite descriptions. We want to confine our attention, as already mentioned, to the so–called referential use. A major point was to show that, indeed, mutual knowledge — or common knowledge — of a fact $\varphi(x)$ is necessary to refer with *the* x *that* φ to a given object a. Mutual knowledge of a fact φ is thereby identified with a conjunction of all sentences of the form:

X_1 knows that X_2 knows that X_3 knows that ... that X_n knows that φ,

where the X_n's are dialogue participants and n ranges over all natural numbers. To this end Clark & Marshall construct a series of examples, where in every step these sentences hold for a formula $\varphi(a)$ up to a certain number, and then fail for all larger ones. In all cases we find a referent for the definite description but one different from the object a for which we want to have reference. We want to emphasise this fact, a point not mentioned by Clark & Marshall. In principle, it is always possible to continue these constructions inductively, showing that we really need common knowledge. Due to their importance for our own discussion, we reconsider Clark & Marshall's examples in full length.

Version 1 On Wednesday morning Ann reads the early edition of the newspaper which says that *Monkey Business* is playing that night. Later she sees Bob and asks, "Have you ever seen the movie showing at the Roxy tonight?"

Clark & Marshall now ask for the conditions which must obtain for the knowledge states of Ann and Bob so that Ann can use the definite description *the movie showing at the Roxy tonight* to refer to *Monkey Business*. First, Ann must know that *Monkey Business* is showing at the Roxy tonight. But speaker's knowledge is not enough:

Version 2 On Wednesday morning Ann and Bob read the early edition of the newspaper and discuss the fact that it says that *A Day at the Races* is showing that night at the Roxy. Later, after Bob has left, Ann gets the late edition, which prints a correction, which is that it is *Monkey Business* that is actually showing that night. Later, Ann sees Bob and asks, "Have you ever seen the movie showing at the Roxy tonight?"

Bob must still believe that Ann refers to *A Day at the Races*. So, at least, Bob should also know that *Monkey Business* is showing. But this still does not secure the intended reference:

Version 3 On Wednesday morning Ann and Bob read the early edition of the newspaper and they discuss the fact that it says that *A Day at the Races* is showing that night at the Roxy. When the late edition arrives, Bob reads the movie section, notes that the film has been corrected to *Monkey Business*, and circles it with his red pen. Later, Ann picks up the late edition, notes the correction and recognizes Bob's circle around it. She also realizes that Bob has no way of knowing that she has seen the late edition. Later that day Ann sees Bob and asks, "Have you ever seen the movie showing at the Roxy tonight?"

Here, Bob knows that *Monkey Business* is showing but he thinks that Ann still believes that it is *A Day at the Races*. Therefore, he will think Ann is refering to *A Day at the Races*.

Version 4 On Wednesday morning Ann and Bob read the early edition of the newspaper and discuss the fact that it says that *A Day at the Races* is playing that night at the Roxy. Later, Ann sees the late edition, notes that the movie has been corrected to *Monkey Business*, and marks it with her blue pencil. Still later, as Ann watches without Bob knowing it, he picks up the late edition and sees Ann's pencil mark. That afternoon, Ann sees Bob and asks, "Have you ever seen the movie showing at the Roxy tonight?"

Here, Bob must reason as follows: Ann knows that *Monkey Business* is playing tonight. But she thinks I believe that we both are mutually convinced that *A Day at the Races* is showing. So she must think that I think she refers to *A Day at the Races.*
Going on, we can find more and more complicated examples, which show that any finite sequence of sentences *Ann knows that Bob knows that ... that* Monkey Business *is showing that night* is not enough to ensure the reference to *Monkey Business.*

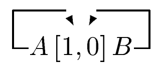
Version 5 On Wednesday morning Ann and Bob read the early edition of the newspaper and discuss the fact that it says that *A Day at the Races* is playing that night at the Roxy. Later, Bob sees the late edition, notices the correction of the movie to *Monkey Business*, and circles it with his red pen. Later, Ann picks up the newspaper, sees the correction, and recognizes Bob's red pen mark. Bob happens to see her notice the correction and his red pen mark. In the mirror Ann sees Bob watch all this, but realizes that Bob hasn't seen that she has noticed him. Later that day, Ann sees Bob and asks, "Have you ever seen the movie showing at the Roxy tonight?"

Here, Bob thinks that Ann thinks that he thinks that she believes that it is mutual knowledge that *A Day at the Races* is showing. So he must think that Ann must think that, if she uses the definite description *the movie showing at the Roxy tonight*, that Bob must believe she refers to *A Day at the Races.* Hence, she can use it only if she wants to induce him to pick out *A Day at the Races* as the referent. And therefore, he must calculate that, indeed, Ann wants to know whether he has seen *A Day at the Races.*
Clark & Marshall arrive at the conclusion that both participants need to know that all sentences of the form

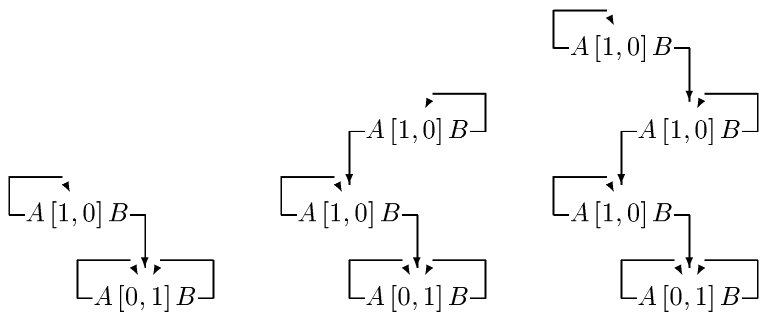
> X_1 knows that X_2 knows that X_3 knows that ... that X_n knows that *Monkey Business* is showing tonight,

where the X_i's are dialogue participants, are true for any natural number n. But this is equivalent to: It must be *common knowledge* that *Monkey Business* is showing tonight.
To make the whole procedure even more perspicuous, we give some graphical representations for the Clark & Marshall examples. To get away from their special design, we choose a more simplified basic situation. We assume there to be two objects which have or have not a certain property φ. We write the values 1 and 0 in square brackets to indicate which object has the property and which does not. So, e.g. $[1, 0]$ means that the first object has the property φ, and the second doesn't. We write the capitals A and B for the dialogue

participants (DP's) on the left and right of the square brackets: $A\,[0,1]\,B$. We represent the epistemic possibilities of a DP by the set of all states of affairs which are compatible with his beliefs. If a state of affairs v is possible for a DP in a situation w we indicate this by an arrow starting from the DP in the representation of w to the representation of v. It should be emphasised that the set of accessible states in a dialogue situation w is assumed to be part of w. So we get the following representation for the basic situation of the Clark & Marshall examples:

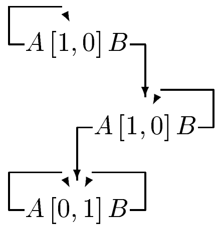


The two objects are *Monkey Business* and *A Day at the Races*, where the first is depicted at the left, the second on the right side. 1 means: it is showing at the Roxy tonight. 0 means: it is not showing. Below we will give precise meaning to these representations. They depict Kripke structures which are derived from *possibilities*. For Version 2 and 4 we get the following structures:

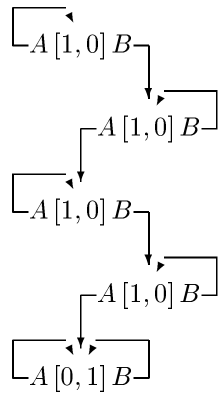


Version 2 No Example Version 4

In the structure for Version 2 we see at the bottom the representation of the situation where Ann and Bob mutually believe that *A Day at the Races* is playing. When Ann reads the correction, she knows that *Monkey Business* is showing, and that Bob's information state has not changed. That Ann knows everything is depicted by the arrow going from A in the topmost situation back to the situation itself, which means that the real situation is the only epistemically possible situation for Ann. That Bob's information is still the old one is represented by the arrow going from B to the structure at the bottom which depicts exactly this old situation. If Bob learns about this and the correct film without Ann noticing it, her information doesn't change, and Bob knows it. This is depicted by the structure in the middle: The arrow from A in the top situation goes to the situation which is the top of the representation of Version 2. From there, a further step — Ann learning about Bob's situation without Bob noticing this — leads to the representation of Version 4.
In the same way we can understand the following representations for Version 3 and 5.



Version 3



Version 5

The graphical representations clarify how the basic situation gets more and more embedded. In the basic situation we get a referent, namely *A Day at the Races*. This referent is *handed over* to the situations higher up in the structure. It is this picture we want to exploit later on. Notice that this observation is in variance with the analysis given by Clark & Marshall themselves. It is here *not* common knowledge that *A Day at the Races* is playing, yet, it would be possible for Ann to establish reference to this movie in Versions 2, 3, 4 and 5. It is common knowledge only in the basic situation.

3 Intended Applications and Framework of the Approach

Before we start with our investigation, we first want to say more about which uses we exactly want to describe. In the second part, we define the framework we adopt to model the relevant aspects of dialogue situations.

3.1 The Epistemic Contexts and the Referential Use of Definites

The discussion of the Clark & Marshall examples suggest the following heuristic description of the situations where we have to explain the successful use of a definite description:

1. There is a fixed set of mutually known objects (in the case of the Clark & Marshall examples they are the movies made by the *Marx Brothers*).
2. The speaker intends to communicate some sentence $\psi(a)$[1], where a is one of the given objects (the movie *Monkey Business*).
3. He fills the argument position for a by a definite description $\mathrm{def}\, x.\varphi(x)$ (*the movie showing at the Roxy tonight*).
4. The hearer can infer that the speaker wanted to communicate $\psi(a)$ by his use of $\mathrm{def}\, x.\varphi(x)$.

There are two points in this description which clearly put the phenomenon in the realm of pragmatics: there is an *intention* of the speaker to *induce* the

hearer to do something, and an *inferential act* of the hearer, which involves a *recognition* of a speaker goal. This remains an underlying assumption in the whole investigation. It already suggest that the beliefs of speaker and hearer about each other should play a central role.
In such a scenario, the use of a description $\text{def}\, x.\varphi(x)$ fits the definition for the *referential use* as it was introduced by *K.S. Donellan*. According to Donellan (Donellan, 1966, § III)

> A speaker who uses a definite description referentially in an assertion [...] uses the description to enable his audience to pick out whom or what he is talking about and states something about that person or thing.

We don't want to argue whether a referentially used description semantically has a referent or not. In examples like

(2) A couple, A and B, sit on a park bench. In some distance there are two men walking. One of them has a red umbrella, the other has a brown one. A who thinks that the umbrellas are walking sticks says

A: The man with the red walking stick visited me yesterday in my office.

But unfortunately B can't discriminate red from brown and asks:

B: Do you mean the man with grey trousers or with the black?

One can argue that the description *The man with the red walking stick* has a meaning, independent of whether B understands it or not. But we are only interested in the referential use in as far as it is used by the speaker to induce the hearer to pick out a certain object. In the example, this aim is not achieved, and so we classify it as not successful. There is a rich philosophical literature on the semantics of definite descriptions, and especially on the question of whether they are basically quantifiers or referring expressions. *Stephen Neale* (Neale, 1990) provides a thorough and up–to–date discussion of this tradition. He defends a decidedly Russellian position. In order to compare it to a referentialist position he works out three theses about referential uses of definites (Neale, 1990, Sect. 3.5):

(A1) If a speaker S uses a definite description 'the F' referentially in an utterance u of 'the F is G', then 'the F' functions as a referring expression and the proposition expressed by u is *object–dependent* (rather than descriptive).

(A2) A speaker S uses a definite description 'the F' referentially in an utterance u of 'the F is G' iff there is some object b such that S means by u that b is the F and that b is G.

(A3) If a speaker S uses a definite description 'the F' referentially in an utterance u of 'the F is G', 'the F' still functions as a quantifier and the proposition expressed by u is the object–independent proposition given by [the x: Fx](Gx)[2].

Neale claims that a *referentialist* endorses (A1) and (A2), whereas a Russellian endorses (A2) and (A3). Hence, both positions meet in (A2). Therefore, both

admit for the phenomenon described in 1.–4. But, of course, (A2) does not contain a criterion for success, therefore it considers only the speaker's perspective. In the following we never need to recur explicitly to the *semantics* of a definite $\mathrm{def}\, x.\varphi(x)$, therefore we prefer to remain neutral in this debate. What we need — and this will turn out only in the course of the following investigation — is a straightforward criterion for the successful use in a class of basic dialogue situations. For all more complicated situations we can explain the success by systematic application of pragmatic principles.
Clark & Marshall themselves provide the following tentative formulation:

> *The direct definite reference convention.* In making a direct definite reference with term t sincerely, the speaker intends to refer to
>
> 1. the totality of objects or mass within a set of objects in one possible world, which set of objects is such that
> 2. the speaker has good reason to believe
> 3. that on this occasion the listener can readily infer
> 4. uniquely
> 5. mutual knowledge of the identity of that set
> 6. such that the intended objects or mass in the set fit the descriptive predicates in t , or, if t is a rigid designator, are designated by t.

What is most missing in our first formulation is an equivalent of what is implicit in conditions 4., 5. and 6., namely a condition that states how the beliefs of the participants, the properties of the given objects and the descriptive content of the description $\mathrm{def}\, x.\varphi(x)$ must be related. But this is exactly what we are about to investigate in this paper. The discussion of the previous examples has shown that the condition stated by Clark & Marshall is sufficient only for basic cases.
Other differences are due to the wider coverage of Clark's and Marshall's formulation. They consider e.g. *plural* definite noun phrases, whereas we will concentrate on the case of *singulars*. But more importantly, they try to cover the referential use in a wide range of cognitive contexts (Clark & Marshall, 1981, p. 22f). But such an investigation in a formal framework is far beyond what can be done in this paper.
E.g. we want to exclude a purely *anaphoric* use of a definite which rest upon *speaker's meaning*, and which is sometimes, e.g. by Donellan (Donellan, 1966, 1978) himself, classified as referential. An example (Donellan, 1978, p. 58) is

> A man came to the office today. The man tried to sell me an encyclopedia.

Such a sentence can begin an anecdote told to friends who know little or nothing about what goes on in the speaker's office. But nevertheless, as Donellan points out, the definite description *the man* can't be understood as implying that there is one and only one person who fits the description, even if we extend the description to *the man who came to the office today.* This constitutes an example of a referential use.

We want to exclude such contexts, because they add problems having to do with anaphoric relations. This would only obscure our main point. This complex can be better studied in connection within a dynamic theory of indefinites, see e.g. (Heim, 1982; Kamp & Reyle, 1993), and especially *P. Dekker* (Dekker, 1997) for a recent approach in the context of dialogue.
Further, we want to exclude *salience*, *indirect anaphoricity* or *community co-membership* as a motive for the established reference, see (Clark & Marshall, 1981).
As might be clear from the discussion of the Clark & Marshall examples, the exact epistemic context is of crucial importance to us. We want to specify our context as one where the dialogue participants can anchor all discourse objects in a fixed set of mutually given objects. But they may have only partial beliefs about the properties of these objects, and about the beliefs of each other. Such a situation may arise most naturally in a *visual–situation–use*–context (Clark & Marshall, 1981, p. 22f), e.g. in a situation where a number of covered or uncovered playing cards are shown to two dialogue participants. But the set of object may equally well be given by *general knowledge* (Clark & Marshall, 1981, p. 22f), e.g. the set of movies by the Marx Brothers.
Below we will choose playing card examples to test various conditions which may describe the referential use. I.e. we will always have a situation where some playing cards are presented to dialogue participants A and B. They lie side by side covered or uncovered on a table, so that they can be discriminated from each other. We don't want to have them in a heap. This is to make sure that we really have a case where the definite is used to select an object out of an *openly* given set of objects.
The use of a definite description "the ace" in a sentence like "Point to *the ace*" refers successfully to a card a if the hearer is able to point unambiguously to the card the speaker wanted him to point at. It is not enough that out of the speaker's perspective, or the hearer's, the use *should* be successful.
Playing card examples are very clear, allow for explicit descriptions of knowledge states, and can be constructed in a perspicuous way. This simplicity allows for consideration of very complex information states. We want to allow for dialogue situations where the participants may have been misinformed about the properties of the playing cards. We will look e.g. at situations where one dialogue participant gets different information than the other. In some examples they know about the information the other one gets, get informed together, or separate from each other. They may even have openly contradicting opinions about the properties of the cards. We want to know in which contexts it is still possible for speaker A to establish successfully the reference to a certain object with a definite description.

3.2 Representing Beliefs in Dialogue: The Possibility Approach

We here take a semantic approach to the representation of the beliefs of dialogue participants, i.e. we identify what a DP A believes with the set of all worlds which are possible according to those beliefs. This idea has been extensively discussed by *J. Hintikka* (Hintikka, 1962). In the light of the discussion above it is obvious that we need to represent the beliefs of dialogue participants about the beliefs of others. We want to restrict our model to the case of two dialogue participants, so that we have real *dialogue*. It would be possible to extend our treatment to the case of n participants, but first this would lead to technically complicated descriptions, and second it is not clear whether this extension would be really empirically significant. Let $DP = \{A, B\}$ denote the *set of dialogue participants*.

A possible world w should contain three things: a description of the situation talked about, a set of possible worlds representing the beliefs of A, and a set of possible worlds representing the beliefs of B. This leads straightforwardly to the so–called *possibility approach* developed by *J. Gerbrandy* and *W. Groeneveld* (Gerbrandy & Groeneveld, 1997). It has already been adopted in a number of formal papers relating to dialogue. The approach presupposes an extension of classical set theory, a theory with an *Anti–Foundation–Axiom* (AFA) introduced by *P. Aczel* (Aczel, 1988). We don't want to go into the technical details of this set theory and refer to (Barwise & Moss, 1996), and for more information about possibilities and Kripke structures to (Gerbrandy, 1998). Historically, this approach developed out of theories on knowledge in multi–agent systems (Fagin et al., 1995). The possibility approach proved to be advantageous especially when it comes to defining updates on information states. We concentrate in our presentation of a possibility approach on the material which is absolutely necessary to formulate the following theory of referentially used definite descriptions in dialogue.

Let S be a class of models for the possible situations the dialogue participants can talk about. For simplicity we assume that all models in S have the same set of individuals. Then we can define *possibilities* and *information states* in the following way[3]:

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

s_w describes the outer situation talked about, $w(A)$ and $w(B)$ the sets of worlds A and B believe to be possible. We call s_w an *outer* situation because we don't want it to contain information about beliefs of dialogue participants. We denote the class of all possibilities with $\mathcal{W}$. Due to the background set theory it is possible that there exist possibilities w with $w \in w(X)$.

We further introduce a formal Language $\mathcal{L}^M$. Let $\mathcal{L}$ be a language of predicate logic for the class 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 $X \in DP$:

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

If $w = \langle s_w, w(A), w(B)\rangle$ is a possibility, then we can 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* $w \models \Box_A \varphi \wedge \Box_B \varphi$. 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*[4] believes φ in w. φ is *common belief* in w iff $w \models C\varphi$.
For information states we can define

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

Until now, we did not impose any conditions on the class of possibilities. Therefore, we want to introduce some relevant subclasses of $\mathcal{W}$.
A subclass $M \subseteq \mathcal{W}$ is called *transitive*, iff

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

This means: If $w \in M$, then all possibilities which belong to the epistemic alternatives of one dialogue participant X belong also to M. Hence, transitivity denotes a *closure* property. If M characterises some property of possibilities, then the transitivity of M means that for $w \in M$ all epistemic alternatives of the participants have also this property, all possibilities for which one of the participants thinks that there is a participant who believes them to be possible have the property, etc.
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.* We want to assume it throughout the rest of the paper. It means that every dialogue participant knows about himself that he thinks that exactly those states of affairs are possible he actually believes to be possible. And it means that this fact is mutual knowledge for the dialogue participants. It follows from introspectivity that in case a participant believes φ he also knows that he believes φ, and if he does not know whether φ, he also knows that he does not know it.
Let $\mathcal{T} \subseteq \mathcal{I}$ be the largest transitive subclass with

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

Hence, if $w \in \mathcal{T}$, then every DP believes w to be possible. Therefore, if he believes that φ then φ must in fact hold. Hence, $\mathcal{T}$ denotes the class of possibilities where the dialogue participants can have only *true beliefs*, i.e. *knowledge*, and where this fact is common knowledge.
If $w \in \mathcal{T}$, $X \in \mathrm{DP}$, we have

1. $w \models \Box_X \varphi \Rightarrow w \models \varphi$;

2. $w \models \Box_X \varphi \Rightarrow w \models \Box_X \Box_X \varphi$;

3. $w \models \varphi \Rightarrow w \models \Box_X \Diamond_X \varphi$.

It should be mentioned here that without the (AFA)–Axiom the class $\mathcal{T}$ would be the empty set.
We want to add a semantic definition of the *Common Ground*:

- For a possibility $w \in \mathcal{W}$ we denote by $CG(w)$ the smallest transitive subclass containing $w(A)$ and $w(B)$.

This definition is justified by

$$\forall \varphi \in \mathcal{L}^M \; CG(w) \models \varphi \text{ } \textit{iff} \text{ } w \models C\varphi.$$

For information states σ we take $CG(\sigma)$ to be the smallest transitive set containing every $CG(w)$ for $w \in \sigma$.
We should add some words on the relation of possibilities to *Kripke*–structures. Let σ be any transitive information state. Then we can construct a Kripke–structure $K = \langle \sigma, \rightarrow_A, \rightarrow_B \rangle$ with two accessibility relations out of σ in the following way: Take as possible worlds just the possibilities in σ. A world v is accessible from world w for dialogue participant X, $w \rightarrow_X v$, iff $v \in w(X)$. If we define truth for unboxed sentences in the same way as for possibilities, and define $(K, w) \models \varphi$ in the usual way for boxed sentences and take over the definition for the operators E and C, we get for all $w \in \sigma$

$$w \models \varphi \text{ } \textit{iff} \text{ } (K, w) \models \varphi.$$

There is also an inverse result which proves that we can find for all Kripke–structures possibilities which validate the same formulas, see (Gerbrandy, 1998, Prop. 3.7). This shows that the possibility approach is essentially a *possible worlds* approach, and that, *in principle*, it would be possible to provide corresponding structures in a standard framework for multi–agent systems like (Fagin et al., 1995).
The graphical representations we used to depict the Clark & Marshall examples can be seen as pictures of Kripke-structures derived from a possibility w, where we take as the transitive information state just the smallest transitive information state containing $\{w\}$. We take Version 2, p. 3 and p. 5, as an example. By $[1, 0]$ we denote again a model where it holds that *Monkey Business* is showing. $[0, 1]$ means that *A Day at the Races* is playing. Ann knows that *Monkey Business* is showing, and that Bob believes that they mutually believe that *A Day at the Races* is playing. Thus, the real situation is of the form $w = \langle [1, 0], w(A), w(B) \rangle$, where $w(A) = \{w\}$ and $w(B) = \{v\}$, with $v = \langle [0, 1], v(A), v(B) \rangle$ and $v(A) = v(B) = \{v\}$. In short:

$$\begin{aligned} w &= \langle [1, 0], \{w\}, \{v\} \rangle \\ v &= \langle [1, 0], \{v\}, \{v\} \rangle \end{aligned}$$

The smallest transitive information state which contains $\{w\}$ is $\{w, v\}$. As accessibility relations we find $w \rightarrow_A w$, $w \rightarrow_B v$, $v \rightarrow_A v$ and $v \rightarrow_B v$. We see that the accessibility relations correspond to the arrows in the representation of Version 2 on page 5.
In the following 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 possibility. Let $\dot{\mathcal{W}}$ denote the largest transitive subclass of $\mathcal{W}$ with

$$w \in \dot{\mathcal{W}} \Rightarrow w(A) \neq \emptyset \neq w(B)$$

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

4 The Conditions for Successful Use of a Definite Description

We use the following conventions: If a definite description is given then $\varphi(x)$ denotes the corresponding formula (with one free variable) for the property characterising an object. We use the capitals A and B as names for the dialogue participants. In all examples, the relevant definite description will be used by participant A. We denote by a the object to which A wants to refer with $\operatorname{def} x.\varphi(x)$.

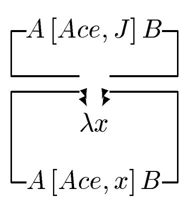
4.1 Some Trials

The examples of *Clark* & *Marshall* have shown that it is necessary to consider the influence of the mutual information states of the dialogue participants if one wants to give conditions for the successful use of a definite description. We identify the *common beliefs* with the *common ground*. If the use of a definite description $\operatorname{def} x.\varphi(x)$ should refer to an object a, the object should be given in *some* unique way in the common ground. But in which way exactly? We are not only interested in dialogue situations with *true* beliefs or *non–conflicting* beliefs but also in situations with erroneous and conflicting views.
To motivate our own solution, we first go through a series of possible formulations of the criterion of the successful referential use of a definite description. We will use playing card examples to show their defectiveness. In all examples we assume that A wants to refer with “the ace” to a *first* or *leftmost* playing card in a row.
At first we want to try a formulation that postulates that *uniqueness* is *common belief*: A speaker A can refer successfully with $\operatorname{def} x.\varphi(x)$ to an object a in a world w, iff

$$CG(w) \models [\varphi(a) \wedge \forall x \neq a\, \neg\varphi(x)].$$

The following example shows that this condition is too strong:

(3) There are two playing cards a and b 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, a, 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*.”



Example **(3)**

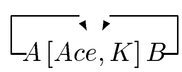
We use λx to indicate that x can vary over all other possible playing cards. If an arrow comes from above, it means that, in fact, we would have arrows to all alternatives we can get, if we replaced x by its possible values. We use J (Joker) where we need a specific playing card to define a concrete outer situation but where this card could have been chosen arbitrarily.

Here it holds that $CG \models \varphi(a)$, i.e. "$ace(a)$", but both dialogue participants know that it is possible that also "$ace(b)$" holds — in fact, it is common knowledge that this is possible. So the above condition is violated, but nevertheless A will refer successfully to a.
It seems necessary to weaken the condition:

$$(\mathbf{C})\ \ CG(w) \models \varphi(a) \ \text{ and } \ \forall b \neq a\, CG(w) \not\models \varphi(b).$$

Now it is not necessary that A and B are both convinced that object a is the *unique* object that satisfies the predicate φ. There must only be no other object which they mutually believe to have this property. Here, *common knowledge* is *unique*.
This criterion gives correct results for examples where both dialogue participants can have only *true* beliefs, and where this is common knowledge, i.e. for situations w with $w \in \mathcal{T}$:

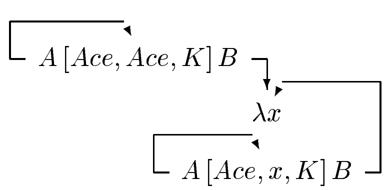
(4) Two cards lie face up, side by side on a table. The first one is an ace, the second a king. A and B can see each other and the cards. Here A can obviously use "the ace" to refer to the first card.



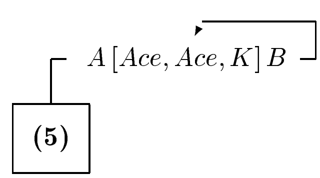
Example **(4)**

In the next example, the speaker has more knowledge than the hearer. Especially, he knows that there is more than one ace.

(5) There are three playing cards on a table. The first and the second one are aces, the third one is a king. A is brought to the table and all three cards are shown to him. Then all cards are turned over again, and B is brought to the table in the presence of A. He is told that A knows all cards. Then a supervisor shows the first and third card to A and B. Here again, A can refer to the first card in the utterance: "Please, point to the ace."



Example **(5)**



Example **(6)**

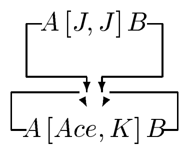
For λx see Example **(3)**. The box in the right structure has to be replaced by the structure for Example **(5)**.

The criterion **(C)** gives the correct prediction even in cases where the beliefs of A and B are only non–conflicting with respect to the outer situation but not necessarily true.

(6) Suppose, B has seen everything in **(5)** but A did not notice this. Still, it is possible for A to refer successfully to the first card in "Please, point to the ace."

This holds, of course, because B knows that A thinks to be in situation **(5)**. There, reference goes to the first card.
In the next example, both dialogue participants get misinformed. But both mutually believe this misinformation to be true:

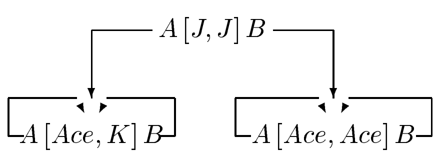
(7) There are two cards on a table. They are lying covered side by side. In the presence of both A and B, a supervisor tells them that the first one is an ace and the second one a king. In fact, there are two jokers on the table. Still, it is possible for A to refer successfully to the first card in "Please, point to the ace."



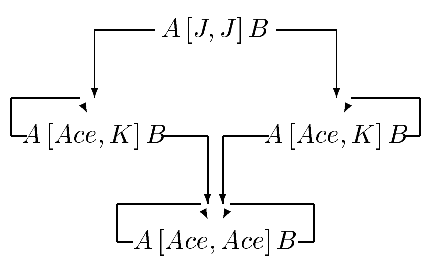
Example **(7)**

If we drop the condition that the beliefs of the dialogue participants are non–contradicting, then we soon find examples that don't fit to the criterion.

(8) There are two playing cards on a table. A thinks that they both are mutually convinced that the first card is an ace and the second a king. B believes that they are both mutually convinced that both cards are aces. Here, it is not possible for A to refer successfully to the first card with "the ace."



Example **(8)**



Example **(9)**

In this example it obviously holds

$$CG(w) \models ace(a) \text{ and } \forall b \neq a\, CG(w) \not\models ace(b),$$

if a denotes the first card and b the second. It is only from the *perspective* of A that he can refer successfully with the definite description to a but not from the perspective of B. Hence, it seems reasonable to introduce the concept of *perspectives* into the model. If $w(X)$ represents the beliefs of an agent X

in world w, then φ holds *under the perspective* of X iff $w(X) \models \varphi$. So the common ground under the perspective of X can be identified with $CG(w(X))$. The criterion for a successful use of a definite description $\mathrm{def}\, x.\varphi(x)$ can now be reformulated as:

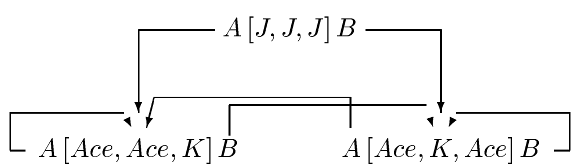
$$\forall X \in \mathrm{DP}\, [CG(w(X)) \models \varphi(a)\ \&\ \forall b \neq a\, CG(w(X)) \not\models \varphi(b)].$$

This criterion means that successful reference should be possible, if the criterion **(C)** applied separately to the two perspectives gives the same object. It can be rephrased as: a is a card such that everyone believes that it is the only card for which it is commonly known that it is an ace. But there are still counter examples.

(9) There are two playing cards side by side on a table. A thinks that the first is an ace and the second a king and that B thinks that both are mutually convinced that they are aces. Assume that B thinks the same, if we interchange A and B. Here, it is also not possible for A to refer to the first card with "the ace."

The problem here is that the criterion gives a referent for both perspectives but, as soon as we change perspectives a second time, it fails.
As in the case of the Clark & Marshall examples it is possible to embed the problematic situation deeper and deeper into the perspectives of the dialogue participants. So these examples seem to recommend a criterion like the following: It must be possible to reach after *some* changes of perspectives only perspectives of dialogue participants such that then after *every* change of perspectives there exists only one object a in the associated common ground for which $\varphi(a)$ holds. The Clark & Marshall versions can be handled in this way. But even here, there is a simple counter example:

(10) There are three cards on the table. Assume that A gets informed by *his* supervisor that the first card and the second card are aces and that the third one is a king. B gets informed by *his* supervisor that the first and the third playing card are aces, and the second a king. A and B take for granted that their respective supervisor always tells the truth. Then, A says: "The first card and the second card are aces, the third a king." B says: "The first and the third card are aces, the second a king." Here, it seems not to be possible for A to refer to the first card with "the ace."



Example **(10)**

We may change perspectives as often as we like, we always get using our criterion that the reference should be possible. But it is not!
If we only try to combine criterion **(C)** with some principles for changes of perspectives, we fail to get the correct predictions for *all* possible epistemic contexts. In the solution proposed in the next section, we determine in addition classes

of reasonable candidates for the application of criterion **(C)**. These classes are generated out of *basic* dialogue situations by operations reflecting changes of perspectives.

4.2 Formulation of a Criterion under Systematic Exploitation of Basic Dialogue Situations through the Perspectives of the Dialogue Participants

If we have a look back to the Clark & Marshall examples, we can find there, in the graphical representations, a very simple possibility at the bottom of the graph.

$$A\,[0,1]\,B$$

We can capture this simplicity precisely as being an element of $\mathcal{T}$, the class with real *knowledge* and common knowledge of knowledge. For this class our discussion could show that criterion **(C)** gives the correct predictions. We can also see that we should be able to handle the Clark & Marshall examples by some additional principle for exploitation of perspectives. Then, if we look again to our sample cases without reference, i.e. **(8)**, **(9)** and **(10)**, we find that there is no simple dialogue situation embedded at the *bottom* of the structure, or, if it is, then criterion **(C)** fails for this situation. This leads to the following picture:

> The interlocutors need to exploit a basic dialogue situation where reference is possible, to get reference in more complicated epistemic contexts. The exploitation works through the use of the different perspectives of the participants.

In the following we first give a definition of basic dialogue situations. Then we define how we can get new successful referential uses of definite descriptions when we take into account the perspectives. Unfortunately, it turns out that at this point we need an additional condition of *non–ambiguity*.

4.2.1 Referential Uses of Definite Descriptions in Basic Dialogue Situations

We take the class of *basic* dialogue situations to be identical with $\mathcal{T}$, i.e. where all participants have only *true* beliefs, i.e. knowledge. For these situations we already saw that the criterion

$$(\mathbf{C})\quad CG(w) \models \varphi(a) \;\text{ and }\; \forall b \neq a\, CG(w) \not\models \varphi(b)$$

leads to the correct predictions. Examples **(3)**, **(4)** and **(5)** belong to this class. We summarise this first result in the following definitions

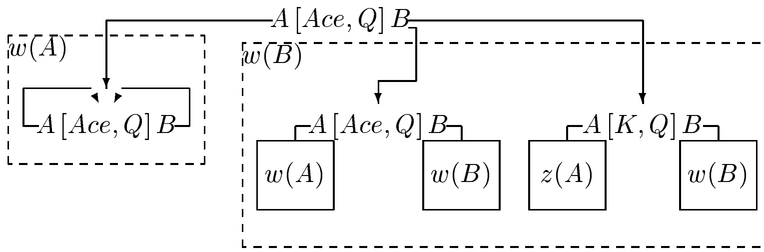
$$\begin{array}{lcl} \mathbf{C}_0(w,a) & :\Leftrightarrow & CG(w) \models \varphi(a)\;\&\;\forall b \neq a\, CG(w) \not\models \varphi(b) \\ \mathit{Succ}_0(a) & = & \{w \in \mathcal{T} \mid \mathbf{C}_0(w,a)\} \end{array}$$

Notice that $\mathit{Succ}_0(a)$ is transitive. It follows from $(\mathbf{C})$ and: $w \in \mathit{Succ}_0(a)\;\&\;v \in w(X) \Rightarrow w \in w(X) = v(X)$, which implies $CG(w) = CG(v)$.

4.2.2 Principles for Deriving New Uses by Changing Perspectives

Let $M \subseteq \dot{\mathcal{I}}$ be a class which represents some property of possibilities, e.g. being a basic dialogue situation. We already specified what it means that this property obtains *under the perspective* of a dialogue participant X in world w, namely as $w(X) \subseteq M$. This means that *all* his epistemic possibilities are elements of M.
If we think of M as specifying a class where some definite description $d = \mathrm{def}\, x.\varphi(x)$ can be used felicitously, then speaker A seems to be justified to utter d if and only if all his epistemic possibilities are elements of M. But looking at the hearer B, if he hears an utterance of d, it seems that the requirement that he really believes that he is in a situation in M is too strong.

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



Example **(11)**

To get the full representation, the boxes have to be replaced by the structures described with the dashed lines. $z(A)$ has the same structure as $w(A)$ with K instead of Ace.

Here, B can guess that A means the first card. The possibility that there is a king and a queen is ruled out by A's use of "the ace." So we come to a weaker condition on the perspective of B, namely that there should be a possibility v in his set of epistemic alternatives $w(B)$ that is in 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 $\Diamond_X$, therefore 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: A is convinced that the actual world w belongs to M iff $w \in \Box_A M$; B can accept that w belongs to M iff $w \in \Diamond_B M$.
Now we can specify what it means that speaker A is convinced that the hearer B thinks that his utterance of d is acceptable. It is just the case, if the actual world is an element of $\Box_A \Diamond_B M$, i.e. in all his epistemic alternatives it is the case that B thinks that there is a possibility where d can be used.

On the other hand, B thinks that it is possible that A is convinced that he can utter d iff the actual world w is an element of $\Diamond_B \Box_A M$.
So we get four basic extensions which give us new classes where some utterances are possible due to the perspectives of the dialogue participants. Let M be given. Then we classify the perspectivally derived uses in the following way:

	direct	indirect
strong	$\Box_X M$	$\Box_X \Diamond_Y M$
weak	$\Diamond_X M$	$\Diamond_X \Box_Y M$

We need to mention the following facts:

Fact 4.1 *Let* $M \subseteq \dot{\mathcal{I}}$ *be a class,* $X \in \mathrm{DP}$.

- $\Diamond_X \Diamond_X M = \Diamond_X M$ *and* $\Diamond_X \Box_X M = \Box_X M$.
- $\Box_X \Box_X M = \Box_X M$ *and* $\Box_X \Diamond_X M = \Diamond_X M$.

and

Fact 4.2 *Let* $M \subseteq \dot{\mathcal{I}}$ *be a transitive class,* $X \in \mathrm{DP}$. *Then we have*

1. $\Diamond_X M = \Box_X M$.
2. $M \subseteq \Box_X M$

Proof:

1. Let $w \in \Diamond_X M$. Then there is a $v \in w(X)$ with $v \in M$. M is transitive, therefore, $v(X) \subseteq M$. Due to introspectivity $v(X) = w(X)$. Therefore, $w(X) \subseteq M$, so $w \in \Box_X M$.
2. Let $w \in M$, then trivially $w(X) \subseteq M$.

$\Box$
The first fact is due to introspectivity. It shows that we don't get anything new if we iterate the operators, i.e. we can't derive new cases where we can use d, if we look to the same perspective again.
The second fact shows that a transitive class is contained in its image, and that the $\Diamond$ and $\Box$ operators are identical for those classes. Transitivity is of interest to us because our class of basic dialogue situations $\mathcal{T}$ is transitive.

4.2.3 The First Perspectival Expansions of Basic Uses of Definite Descriptions

We have already defined the class of basic dialogue situations, $\mathcal{T}$, and described the subclass of basic situations where A can successfully use a definite description $d = \operatorname{def} x.\varphi(x)$ to refer to an object a. It is exactly the subclass of those situations where our criterion **(C)** holds.
We can substitute $\mathcal{T}$ for the class M of the last section, Section 4.2.2. We first want to define a new class of the reasonable candidates for a successful use of d, where we apply the direct operations to get them:

$$Der_1 := \Box_A \mathcal{T} \cap \Diamond_B \mathcal{T}$$

This is the class of dialogue situations where A thinks he is in a basic situation and B thinks it is at least possible that he is in such a situation too. We call the elements of Der_1 *perspectivally derived* candidates.
Accordingly, we get a criterion for these perspectivally derived candidates:

$$\mathbf{C}_1(w,a) :\Leftrightarrow w \in \Box_A Succ_0(a) \cap \Diamond_B Succ_0(a)$$

Fact 4.2 shows that this is equivalent to

$$Der_1 = \{w \in \dot{\mathcal{I}} \mid w(A) \subseteq \mathcal{T} \,\&\, w(B) \subseteq \mathcal{T}\}$$

and

$$\mathbf{C}_1(w,a) \Leftrightarrow w(A) \subseteq Succ_0(a) \,\&\, w(B) \subseteq Succ_0(a)$$

$\mathbf{C}_1(w,a)$ holds exactly, if both participants think that the definite description $\mathrm{def}\, x.\varphi(x)$ must refer to a. This is in fact equivalent to the criterion stated already in Section 4.1

$$\forall X \in \mathrm{DP}\,[CG(w(X)) \models \varphi(a) \,\&\, \forall b \neq a\, CG(w(X)) \not\models \varphi(b)].$$

But now, we only allow it to be applied to the class Der_1! In this way we get the new class of derived successful uses of the definite description $d = \mathrm{def}\, x.\varphi(x)$:

$$Succ_1(a) := \{w \in Der_1 \mid \mathbf{C}_1(w,a)\}$$

Example **(7)** belongs to this class. We can easily see that Der_1 and $Succ_1(a)$ are again transitive classes.
Until now, we have only considered the *direct* operations. We now can go on and consider the next possibility, the *indirect* operations, to derive new referential uses of the definite description d. Again we need to have both participants believe that d should successfully refer. In this way we again get an intersection of two sets which define the situations where the two participants individually think that the expression refers. We get as new candidates

$$Der_2 := \Box_A \Diamond_B Der_1 \cap \Diamond_B \Box_A Der_1$$

This class contains all situations where A thinks that B believes that it is possible that they both believe to be in a basic situation, and where B thinks it is possible that A is convinced that they both believe to be in a basic situation. Due to Fact 4.2 this is equivalent to

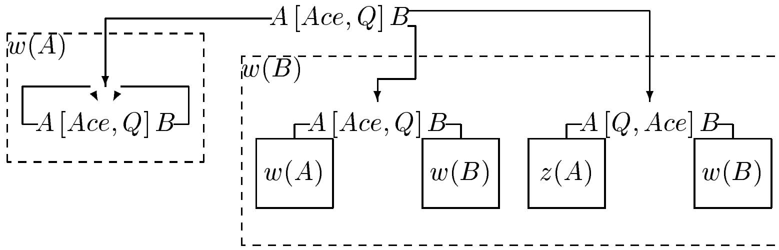
$$Der_2 = \Box_A \Box_B Der_1 \cap \Diamond_B \Box_A Der_1$$

We can't replace $\Diamond_B$ in the second half, because $\Box_A Der_1$ is not transitive. We can see this if we think of a situation w where A's information state $w(A)$ is the same as in Example **(4)**, and B's is like that of Example **(10)**. Then $w(A) \subseteq \mathcal{T} \subseteq \Box_A \mathcal{T} = Der_1$. Hence, $w \in \Box_A Der_1$. But $w(B) \not\subseteq \Box_A Der_1$. We will see in the next section that the $\Diamond$–operator causes new problems.

4.2.4 Adding a Condition of Non–Ambiguity

The following example shows where the trouble comes from:

(12) There are two playing cards on a table. The left one is an ace and the right one a queen. A and B can see them and each other seeing them. Then they leave the room. An hour later they come back, the cards are still there but face down. B has forgotten whether the first card is an ace or the second. He still knows that the other one is a queen. He knows that A has not noticed this. Then A says to him: “Give me the ace, then we leave and I will invite you for a coffee.”



Example **(12)**

$z(A)$ is like $w(A)$, if we interchange Ace and Q in $w(A)$, i.e. $s_z = [Q, Ace]$, $z(A) = \{\langle s_z, z(A), z(A)\rangle\}$.

Here there is a possibility in $w(B)$ where A should be convinced that he can refer to the first card. But there is also a second possibility where A should be convinced that he can refer to the second card. Hence, B can't decide which card is meant, and the reference fails. We can derive the following condition: If there is a possibility in B's epistemic alternatives where A is convinced that he can successfully refer to an object a with the definite description d, then there should be no other object b and epistemic alternative v such that A should think in v that he can refer to b with the *same* definite description d.
Hence, if the use should be successful in w, then first w has to belong to $\Box_A \Diamond_B Succ_1(a)$. Then, it must also belong to $\Diamond_B \Box_A Succ_1(a)$. But it is not allowed to belong to any $\Diamond_B \Box_A Succ_1(b)$ for $b \neq a$. We collect the three observations in a condition $\mathbf{C}_2(w, a)$, which holds exactly, iff

$$w \in \Box_A \Diamond_B Succ_1(a) \cap \Diamond_B \Box_A Succ_1(a) \setminus Riv(a)$$

$Riv(a)$ is intended to include all dialogue situations where there is a *rival* situation in the set of B's epistemic alternatives where reference goes to an object $b \neq a$. As a preliminary approximation we can set $Riv(a) := \bigcup\{\Diamond_B \Box_A Succ_1(b) \mid b \neq a\}$. This can be only preliminary because not all successful references to a $b \neq a$ are contained in $\Box_A Succ_1(b)$. Therefore, as we derive more cases, we add more possibilities to $Riv(a)$ which need to be eliminated from $\Box_A \Diamond_B Succ_1(a) \cap \Diamond_B \Box_A Succ_1(a)$. With $\mathbf{C}_2$ we would get a new class of cases where a successful referential use of d is possible:

$$Succ_2(a) := \{w \in Der_2 \mid \mathbf{C}_2(w, a)\}$$

Examples **(6)** and **(11)** belong to this class, as well as Version 2 of the Clark & Marshall examples, if we take a to be *A day at the races*. As our construction starts with $\mathcal{T}$, we want to define $Der_0 := \mathcal{T}$. It turns out that we will not offer a simple formulation of $\mathbf{C}_2$. So the classes Der_0 and Der_1 get an exceptional status in this respect.

4.2.5 Iterating the Construction

Now, we have all elements at hand to iterate the derivation of referential uses of a definite description $d = \operatorname{def} x.\varphi(x)$.
We start again with the classes $Der_0 = \mathcal{T}$, $Succ_0(a) = \{w \in \dot{\mathcal{I}} \,|\, \mathbf{C}_0(w,a)\}$, where $\mathbf{C}_0(w,a) :\Leftrightarrow CG(w) \models \varphi(a) \,\&\, \forall b \neq a \, CG(w) \not\models \varphi(b)$ and a is the object referred to.
We proceed by first defining for each dialogue participant X separate classes of situations M^X where *he* thinks that the use of a definite d refers to a. We use the four principles of perspectival derivation and the non–ambiguity principle. In the end we have to form intersections of those classes to get the cases where reference is in fact possible.
To make the construction clearer we collect for each participant the classes M^X in a class S^X. We do the same for the classes of candidates. In this way we can easily form all combinatorially possible intersections, and therefore get a better overview of which combinations are there. We define the classes $Succ_\alpha(b)$ simultaneously for all objects b.
Let $\mathrm{D}_0^{(X)} := \{N \mid N \subseteq Der_0\}$, $\mathrm{S}_0^{(X)}(a) := \{N \mid N \subseteq Succ_0(a)\}$. Assume that Der_α and $Succ_\alpha(a)$ are defined.

$$\mathrm{D}^A_{\alpha+1} := \{\Box_A \Diamond_B M \mid M \subseteq Der_\alpha\} \cup \{\Box_A M \mid M \subseteq Der_\alpha\} \cup \mathrm{D}^A_0$$

$$\mathrm{D}^B_{\alpha+1} := \{\Diamond_B \Box_A M \mid M \subseteq Der_\alpha\} \cup \{\Diamond_B M \mid M \subseteq Der_\alpha\} \cup \mathrm{D}^B_0$$

The classes D^X_α collect the classes where due to the perspectives of the individual dialogue participants a basic dialogue situation is embedded in such a way that it makes sense to ask whether reference is possible.
If we are now going to the subclasses where in fact the referential act is successful from the individual perspectives, we have to exclude cases where in the perspective of B there are other objects $b \neq a$ such that reference to those other objects can occur. We collect those *rival* situations into the classes $Riv^{d/i}(a)$, where d and i are used to indicate whether the *direct* or *indirect* operations are at work to derive new cases. As indicated, $Riv(a)$ has to be defined recursively.

$$Riv^d_{\alpha+1}(a) := \bigcup \{\Diamond_B N \mid N \subseteq Succ_\alpha(b), b \neq a\}$$

$$Riv^i_{\alpha+1}(a) := \bigcup \{\Diamond_B \Box_A N \mid N \subseteq Succ_\alpha(b), b \neq a\}$$

We follow our ideas developed in the last section and define the class of classes where the referential act is successful out of the perspective of A as:

$$\begin{aligned} \mathrm{S}^A_{\alpha+1}(a) \quad := \quad & \{\Box_A((\Diamond_B M) \setminus Riv^d_{\alpha+1}(a)) \mid M \subseteq Succ_\alpha(a)\} \cup \\ & \{\Box_A M \mid M \subseteq Succ_\alpha(a)\} \cup \mathrm{S}_0(a) \end{aligned}$$

The first class can be understood to contain all classes of possibilities where A is convinced ($\Box_A$) that (1) B thinks it to be possible ($\Diamond_B$) that he is in a situation in M where reference goes to a ($M \subseteq Succ_\alpha(a)$), but (2) B can't be ($\setminus$) in a situation ($Riv^d_{\alpha+1}$) where he thinks reference could possibly go to a b different from a.
For B we define $\mathrm{S}^B_{\alpha+1}(a)$ accordingly as:

$$\begin{aligned} \mathrm{S}^B_{\alpha+1}(a) \quad := \quad & \{(\Diamond_B \Box_A M) \setminus Riv^i_{\alpha+1}(a) \mid M \subseteq Succ_\alpha(a)\} \cup \\ & \{(\Diamond_B M) \setminus Riv^d_{\alpha+1}(a) \mid M \subseteq Succ_\alpha(a)\} \cup \mathrm{S}_0(a) \end{aligned}$$

If reference should be successful, both participants should be convinced that it is in fact possible and must go to the same object. Therefore, we form in the next step all possible intersections:

$$\mathrm{D}_{\alpha+1} := \{M \cap N \mid M \in \mathrm{D}^A_{\alpha+1}, N \in \mathrm{D}^B_{\alpha+1}\}$$

$$\mathrm{S}_{\alpha+1}(a) := \{M \cap N \mid M \in \mathrm{S}^A_{\alpha+1}(a), N \in \mathrm{S}^B_{\alpha+1}(a)\}$$

Now we can get $Der_{\alpha+1}$ and $Succ_{\alpha+1}(a)$ by forming the unions:

$$Der_{\alpha+1} := \bigcup \mathrm{D}_{\alpha+1},$$

and

$$Succ_{\alpha+1}(a) := \bigcup \mathrm{S}_{\alpha+1}(a).$$

We can continue this construction up into the transfinite case. For Der_λ we can take the union of the classes already constructed. For $Succ_\lambda$ we must acknowledge the fact that in the way of construction some possibilities may be added or removed from the $Succ_\alpha(a)$'s. So we keep only the situations which remain constantly in the classes:

$$Succ_\lambda(a) := \{w \in \dot{\mathcal{I}} \mid \exists \beta < \lambda \, \forall \alpha \geq \beta \, w \in Succ_\alpha(a)\}$$

Finally we can define the class of all candidates Der as the union of all Der_α's, where α is an ordinal. For $Succ(a)$ we get in a similar way as for limit ordinals:

$$Succ(a) = \{w \in \dot{\mathcal{I}} \mid \exists \beta \, \forall \alpha \geq \beta \, w \in Succ_\alpha(a)\}.$$

We already mentioned that the classes Der_0 and Der_1 have a somewhat exceptional status. To get the class of all successful uses it would not have been necessary to define in parallel the classes Der_α of derived candidates, as we can't offer a simple criterion $\mathbf{C}_\alpha$. We want to add now a further observation that makes the first class of directly derived situations even more exceptional. If $N \in \mathrm{S}^B_{\alpha+1}(a)$ and $M \in \mathrm{S}^B_\alpha(a)$, then we have

$$\Diamond_B(\Box_A M \cap N) \subseteq \Diamond_B N = N.$$

Therefore, $\Diamond_B(\Box_A M \cap N) \subseteq \Diamond_B \Box_A M \cap N$. For $N \in \mathrm{S}^A_{\alpha+1}(a)$ and $M \in \mathrm{S}^A_\alpha(a)$ we have

$$\Box_A(N \cap \Diamond_B M) = \Box_A N \cap \Box_A \Diamond_B M \subseteq \Box_A N = N.$$

Hence, $\Box_A(N \cap \Diamond_B M) \subseteq N \cap \Box_A \Diamond_B M$. This shows that we don't get new cases if we apply the *direct* operations, if we already applied the *indirect* ones. Hence, it is only of interest to consider the direct operations in the *base* case. Therefore, we define a separate class for them only there. This leads to the classes Der_i, $Succ_i(a)$ for $i = 0, 1$ as introduced in the last sections. For all higher classes it would have been enough to use only the indirect operations, so that the definitions for e.g. the $\mathrm{S}^X_\alpha(a)$'s simplify to:

$$\mathrm{S}^A_{\alpha+1}(a) := \{\Box_A((\Diamond_B M) \setminus Riv^d_{\alpha+1}(a)) \mid M \subseteq Succ_\alpha(a)\} \cup \mathrm{S}_1(a)$$

$$\mathrm{S}^B_{\alpha+1}(a) := \{(\Diamond_B \Box_A M) \setminus Riv^i_{\alpha+1}(a) \mid M \subseteq Succ_\alpha(a)\} \cup \mathrm{S}_1(a)$$

In the next section we will prove in detail that Example **(5)** belongs to Der_0 and Example **(6)** to Der_2. Version 3 of the Clark & Marshall examples is an element of Der_3, if we take $\mathrm{def}\, x.\varphi(x)$ to be "The movie showing at the Roxy tonight" and a to be *A day at the Races*. Version 4 is an example for Der_4 and Version 5 for Der_5. This shows that we in principle can find examples for all natural numbers n, if we follow the Clark & Marshall constructions. Of course, for higher n the examples get more and more unnatural. On the other hand, there is no good reason to stop the construction for any specific n.

It may be noted at this point that in the construction of Der there is nothing specific about the referential use of a definite description. We could have started with any class $\mathcal{B}$ of *basic* situations, and then extended this class by application of our operators $\Box_A$, $\Diamond_B$, $\Box_A\Diamond_B$ and $\Diamond_B\Box_A$. We can derive in this way an equivalent of the classes Der_α for $\mathcal{B}$. If we think about $\mathcal{B}$ as specifying any class of situations where some dialogue act is primarily defined, then the so constructed hierarchy defines systematically all extensions where this dialogue act can occur due to the influence of the perspectives of the participants. The operations reflect two principles: A participant can be convinced that he is in a basic situation where the dialogue act is justified (direct case), or he can think that the other one believes that the act works (indirect case). This opens a way to treat generally the impact of perspectives in pragmatics. Therefore, we can see that the results of the last section receive a general importance. They are not restricted to the particular problem of the referential use of a definite. Of course, this is not the place to follow this idea further.

To get the class of successful uses we needed to consider an additional constraint of non–ambiguity. It can be seen as a special form of a Gricean maxim. Therefore, we can see the above hierarchy of the $Succ_\alpha$'s as the result of a recursive application of three general principles.

On the other hand this general perspective gives us an idea concerning possible shortcomings of our approach: If there are further cases where a successful referential use of a definite description is possible, then these cases must be derived by different principles. Hence, if we could find such principles, we should be able to systematically construct examples which are not covered by our theory. Until now, we don't know of such principles or examples.

5 Some Worked Examples

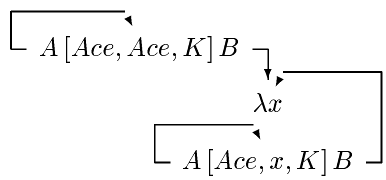
If we have a dialogue situation $w \in \dot{\mathcal{I}}$, when will it be successful to use a definite description $\mathrm{def}\, x.\varphi(x)$ to refer to an openly given object a? The answer we offer is: The situation must be an element in $Succ(a)$.

We want to look again at some of our examples and give a detailed treatment. They are **(5)**, **(6)**, **(10)** and **(12)**. **(5)** is an example of a *basic* dialogue situation where reference is possible. In **(6)** we find an example where the hearer can pick out the right referent due to his knowledge of the perspective of the speaker. In Example **(10)** we find a case with openly contradicting beliefs. We will show that the reference is not possible. In the last example, Example **(12)**, we have a case where the reference fails because the hearer has two alternative interpretations for the definite. We start with two useful observations. Let us denote by $T(w)$ the smallest transitive superset of $\{w\}$, i.e. $T(w) = \{w\} \cup CG(w)$.

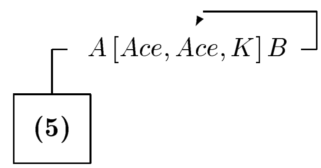
Fact 5.1 • *For all objects* a, b*:* $a \neq b \Rightarrow Succ_0(a) \cap Succ_0(b) = \emptyset$.

- *If* $w \in Succ(a)$*, then* $T(w) \cap Succ_0(a) \neq \emptyset$.

The first observation follows directly from the definition, the second is proved by induction.
We begin with Examples **(5)** and **(6)**.



Example **(5)**



Example **(6)**

As already stated, **(5)** is an element of $\mathcal{T}$ and our criterion applies. Now we want to prove this. A possibility $w = \langle s_w, w(A), w(B)\rangle$ consists of three parts, a model s_w for the situation talked about, an information state for A and an information state for B.
For the models of the outer situation we want to simplify things as their exact representation plays no role in our theory. All our models should have a common range of objects. As we have only playing cards on a table we assume this range to consist of two or three cards. We should have (one–place) predicates for *left*, *middle* and *right*, and for *Ace*, K, Q, etc. As the details are of no importance, we continue to write e.g. $[Ace, Ace, K]$, if there are three cards in the model, one card in the extension of *left* and *ace*, one in *middle* and *ace*, and one card in the extension of *right* and K. It seems not necessary here to be more specific. We denote the set of the predicates *Ace*, K, Q etc. as *Cards*.
In Example **(5)** we can describe the world by a possibility w with

$$\begin{array}{rcl} w & = & \langle [Ace, Ace, K], w(A), w(B)\rangle \\ w(A) & = & \{w\} \\ w(B) & = & \{v_p \mid p \in Cards\} \\ v_p & = & \langle [Ace, p, K], v_p(A), v_p(B)\rangle \\ v_p(A) & = & \{v_p\} \\ v_p(B) & = & w(B) \end{array}$$

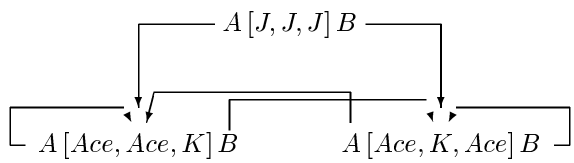
$T(w)$ denotes the smallest transitive superset of $\{w\}$, hence $T(w) = \{w\} \cup \{v_p \mid p \in Cards\}$.
We have to prove first that $w \in \mathcal{T}$: We observe that for all $u \in T(w)$ we find for all dialogue participants $X \in \mathrm{DP}$ that $\forall z \in u(X)\, u(X) = z(X)$. Then we have $\forall X \in \mathrm{DP}\, \forall u \in T(w)\, u \in u(X)$. The first observation proves that $w \in \mathcal{I}$, and together with the second that $w \in \mathcal{T}$.
Next we show that our criterion **(C)** applies. $CG(w)$ is the smallest superset of $w(A)$ and $w(B)$. As $w \in w(A)$, we have $CG(w) = T(w) = \{w\} \cup \{v_p | p \in Cards\}$.
We easily see that for all $u \in CG(w)$ the left card, we denote it by a, is an ace. Hence, $CG(w) \models ace(a)$. For the other cards on the table we find models in $CG(w)$ where they are not aces. Hence, $CG(w) \not\models ace(b)$ for $b \neq a$. This shows that the criterion **(C)** applies, hence $w \in Succ_0(a) \subset Succ(a)$.

Next, we look at Example **(6)**. We can describe it with a possibility w' with

$$\begin{array}{rcl} w' & = & \langle [Ace, Ace, K], w'(A), w'(B) \rangle \\ w'(A) & = & w(A) \\ w'(B) & = & \{w'\}. \end{array}$$

For w we have to add the equations of Example **(5)**. We immediately see that $w'(A) \subseteq Succ_0(a)$, as $w'(A) = \{w\}$. Hence, $w' \in \Box_A Succ_0(a)$. We also see, as in the last example, that $w' \in \dot{\mathcal{I}}$.
w' is not an element of $\mathcal{T}$: If so, w' has to be an element of $w'(A) = w(A) = \{w\}$. Hence, it would follow that $w = w'$. But $w'(B)$ is a singleton set, and $w(B)$ is not. Hence, w' can't be identical with w, and therefore $w' \notin \mathcal{T}$. In addition, this shows $w' \notin Succ_0(a) \subseteq \mathcal{T}$.
With $w' \in w'(B)$, it follows that $w' \notin \Box_B Succ_0(a) = \Diamond_B Succ_0(a)$. Hence, $w' \notin Succ_1(a)$.
Again with $w'(B) = \{w'\}$ we see that $w' \in \Box_B \Box_A Succ_0(a) \subseteq \Diamond_B \Box_A Succ_0(a)$. Hence, $w' \in \Box_A Succ_0(a) \cap \Diamond_B \Box_A Succ_0(a)$.
Finally, we have to show that w' is not an element of any class $Succ(b)$ for $b \neq a$. Assume that $w' \in Succ(b)$ with $b \neq a$. Then, due to Fact 5.1, $T(w') \cap Succ_0(b) \neq \emptyset$. As $T(w') \cap \mathcal{T} = T(w)$ and $Succ_0(b) \subseteq \mathcal{T}$ it follows that $T(w) \cap Succ_0(b) \neq \emptyset$. But $w \in Succ_0(a)$, and as $Succ_0(a)$ is transitive, it follows that $T(w) \subseteq Succ_0(a)$. Hence, $Succ_0(a) \cap Succ_0(b) \neq \emptyset$, which contradicts Fact 5.1.
Hence, we have proved that $w' \in Succ_2(a)$, and reference should be successful.

Now we want to look at an example where the reference fails. We choose Example **(10)**.



Example **(10)**

We can describe the dialogue situation by the possibility w with

$$\begin{array}{rcl} w & = & \langle [J, J, J], \{v\}, \{u\} \rangle \\ v & = & \langle [Ace, Ace, K], \{v\}, \{u\} \rangle \\ u & = & \langle [Ace, K, Ace], \{v\}, \{u\} \rangle \end{array}$$

Here, $T(w) = \{w, u, v\}$. We observe that $w \notin w(A) \cup w(B)$, $v \notin v(B)$ and that $u \notin u(A)$. Hence, $T(w) \cap \mathcal{T} = \emptyset$. With Fact 5.1 it follows that $w \notin Succ(a)$ for any a. Therefore, no reference should be possible.
Finally, we want to look at Example **(12)**. We describe the situation w by the system of equations:

$$\begin{array}{rcl} w & = & \langle [Ace, Q], \{u\}, \{w, v\} \rangle \\ u & = & \langle [Ace, Q], \{u\}, \{u\} \rangle \\ v & = & \langle [Q, Ace], \{z\}, \{w, v\} \rangle \\ z & = & \langle [Q, Ace], \{z\}, \{z\} \rangle \end{array}$$

We see that $w \notin w(A)$, hence $w \notin \mathcal{T}$. As $w \in w(B)$, it follows that $w \notin Der_1$ either. Hence, w can't be an element of $Succ_1(a)$ or $Succ_1(b)$, if a denotes the left and b the right playing card. As in the previous examples we see that $u \in Succ_0(a)$ and $z \in Succ_0(b)$. From this we can follow that $w \in \Box_A Succ_0(a) \cap \Diamond_B \Box_A Succ_0(a)$, and $w \in \Diamond_B \Box_A Succ_0(b)$. But then it follows that w belongs for all α's to $Riv_\alpha(a)$ and $Riv_\alpha(b)$. Therefore, w can't be in any $Succ_\alpha(a)$ or $Succ_\alpha(b)$.

6 Conclusion

We asked for the conditions for successful referential use of definites in dialogue situations which we characterised by the following conditions:

1. There is a fixed set of mutually known objects.
2. The speaker intends to communicate some sentence $\psi(a)$, where a is one of the given objects.
3. He fills the argument position for a by a definite description $\mathrm{def}\, x.\varphi(x)$.
4. The hearer can infer that the speaker wanted to communicate $\psi(a)$ by his use of $\mathrm{def}\, x.\varphi(x)$.

Dialogue is of special interest due to the complex information states of the participants which may arise. We first investigated the claim that the condition for a successful use of $\mathrm{def}\, x.\varphi(x)$ is connected to the fact that there is only one object a where $\varphi(a)$ is mutually known by the interlocutors. Later, we developed out of this claim a first condition (**C**), but we found out that it can be applied only to a class $\mathcal{T}$ of basic dialogue situations where everybody has true beliefs, and where this is common knowledge.
Our central idea was: In non–basic dialogue situations the referential use of a definite description is justified because it can be extended from basic cases to this new situation. We systematically constructed these extensions. We did it by a recursive application of operations which reflect the fact that the participants can only rely on their own perspectives for the justification of their use of a definite. There were two principles which can lead to new cases where successful reference is possible: A participant can be convinced that he is in a situation where the use is successful (direct case), or he can think that the other one believes that the reference works (indirect case). The reference can be established if both participants individually believe that it is possible, and if they both pick out the same object. In addition to the influence of the individual perspectives we considered the impact of a non–ambiguity constraint.
It is essential that a description can be used only if for both speaker and hearer there is a connecting link to basic dialogue situations. Therefore the insights of Clark & Marshall concerning the role of mutual knowledge remain valid in this way.
We indicated how our approach can be used to explicate the way how the fact that the participants have restricted perspectives can lead to extensions of arbitrary dialogue acts.

Notes

[1]For simplicity we may assume here that ψ represents the content of an assertion. Generally, we can take it to represent any linguistic context where some definite description def$x.\varphi(x)$ occurs, and where the speaker intends to induce the hearer to interpret it as the object a.

[2]Where [the x: $Fx](Gx)$ is true iff 1. all F's are G's, and 2. there is exactly one F.

[3]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 $\Phi(x) := \{\langle s, i, j\rangle \mid s \in S, i, j \subseteq x\}$. This operator is *uniform*, therefore it can be *represented* as the class of all *solutions* for a certain class of *systems of equations*. The definition given in the text 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 S and i, j to I.
- $i = \sigma$ for $i \in I$, where σ is a subset of P.

In this way, the definition in the text is really a proper definition.

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

[4]We introduce the E–operator just for convenience. We can use the set $CG(w)$, which will be defined below, to provide a direct *semantic* definition of $w \models C\varphi$. But the use of the E–operator is more intuitive.

References

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

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

H.H. Clark, C.R. Marshall (1981): *Definite Reference and Mutual Knowledge*; in: A.K. Joshi, B. Webber, J. Sag (ed.): *Elements of Discourse Understanding*.

P. Dekker (1997): *On First Order Information Exchange*; in: A. Benz & G. Jäger (ed.), *Proceedings of the Munich Workshop on Formal Semantics and Pragmatics of Dialogue MunDial'97*, pp.21–39 München.

K.S. Donellan (1966): *Reference and Definite Descriptions*; Philosophical Review 60, pp.281–304.

K.S. Donellan (1978): *Speaker Reference, Descriptions and Anaphora*; in: P. Cole (ed.): *Syntax and Semantics: Pragmatics*, Vol.9, p.47–68, San Diego.

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

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

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

I. Heim (1982): *The Semantics of Definite and Indefinite Noun Phrases*; Ph.D. dissertation. University of Massachusetts.

J. Hintikka (1962): *Knowledge and Belief*; Ithaka.

H. Kamp, U. Reyle (1993): *From Discourse to Logic*; Dordrecht.

S. Neale (1990): *Descriptions*; Cambridge, Massachusetts.